THE UNIVERSITY OF CHICAGO

DETERMINING THE GENETIC BASIS OF GENE EXPRESSION RESPONSES TO *IN VITRO* MODELS OF OSTEOARTHRITIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM: GENETICS, GENOMICS,

AND SYSTEMS BIOLOGY

BY

ANTHONY HUNG

CHICAGO, ILLINOIS

DECEMBER 2022

To my friends and family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Working towards a PhD thesis is difficult for anyone, and it is only through the collective work and contributions of countless people that I have been able to achieve what you see here. This section is devoted to trying to scratch the surface of all those who have been instrumental in making this work possible and oftentimes enjoyable.

First, I would like to thank my mentors, who have supported me throughout the PhD and provided advice, served as examples to look up to, and allowed me to develop as a thinker, writer, and researcher. My advisor Yoav Gilad has shaped the way I approach science and fostered a nurturing lab environment and community that I have been fortunate enough to call home over the past years. Genevieve Housman, who was my first of many direct mentors in the lab and first trained me in cell culture, has shown unending patience and compassion throughout the PhD. Other members of the lab and its affiliates, including past and present post docs Michelle Ward, Kenneth Barr, Ben Umans, Abhishek Sarkar; science writer Natalia Gonzales, fellow graduate student Wenhe Lin; and laboratory technicians Emilie Briscoe and Olivia Allen have provided direct training, assistance, or discussions that have contributed greatly to this work. And these names only are small fraction of the long list of researchers in the lab and at the university who should be thanked if there were enough room and time to list them all.

I would also like to thank members of my thesis committee, including chair Tong-Chuan He, Xin He, and Marcus Clark. Their additions and perspectives on my work throughout the inception through its execution, as well as the guidance and advice they have offered regarding science and career have been very helpful, and the patience they have shown towards me has been extremely generous.

Thank you to administration staff in the MSTP, Genetic Medicine, human genetics, and GGSB. The wheels of the University would grind to a halt and all of our lives would be much more difficult without the thankless work that you do every day.

I would also like to thank my friends who I have met during my time at the University of Chicago. Especially members of my MSTP cohort, GGSB cohort, lab members, and Pritzker medical students have been sources of bright light these past few years. It has been a great pleasure to have the opportunity to spend time on this campus and in this city with you all, and I look forward to more of our time together.

Of course, I would never have even made it to the point of having the privilege of being in the place I am now without the support and upbringing from my family. I can always count on my parents, grandparents, and brothers to be there when I need them the most, and I have needed them many times. Thank you all to those who have helped me and taught me. I owe this work and all that I do to the influences you have had on me, and I am eternally grateful.

ABSTRACT

Genetic variation contributes greatly to differences between people, including differences in susceptibility to disease. A critical way that genetic variation influences traits is through impacts on gene regulation. In particular, context-specific gene regulation only visible during certain points in an organism's development or under specific conditions is likely an important and understudied contributor to trait and disease variation. Using induced pluripotent stem cell-derived chondrogenic cells and *in vitro* mechanical and cytokine treatments, I characterized a system to study context-specific gene regulation in the lab, with relevance to the joint disease osteoarthritis (OA). First, I used bulk and single cell transcriptional information from iPSC-derived chondrogenic cells (iPSC-chondrocytes) from three individuals subjected to control and mechanical stress conditions used to model OA. From this study, I found that patterns of gene expression that differ between conditions are relevant for gene sets related to joint health and OA. I also found examples of genes that exhibit inter-individual differences in responses to biomechanical strain, potentially representing examples of gene-by-environment interactions in response to perturbations. I expanded this system to survey transcription from iPSC-chondrocytes derived from 22 genotyped individuals under static, biomechanical strain, and inflammatory cytokine conditions. Through this study, I found shared and unique gene expression patterns that are activated in iPSC-chondrocytes in response to the two treatments. I further performed expression quantitative trait locus (eQTL) mapping to correlate genetic variation to gene regulation in each treatment condition. By comparing eQTLs between conditions, I determined a set of dynamic response eQTLs that are only visible in specific environmental contexts and represent the effects of gene-by-environment interactions on gene regulation. This work represents a study of

genetic control of gene regulation in a disease-relevant cell type in disease-relevant environmental contexts.

CHAPTER 1

INTRODUCTION

1.1 Human genetics and the genetic determinants of phenotypic diversity in humans

Humans share 99.5% sequence identity across a genome containing 3 billion nucleotides [1], and the wide range of phenotypic diversity observed between individuals results from variation in the remaining 0.5% as well as epigenetic differences and differences in environmental exposures. A primary goal of human genetics is to identify genetic contributions to observed phenotypic variation, including variation in disease phenotypes. Knowledge of how genetic variation contributes to human disease variation can aid in the development of prediction tools, allow for the establishment of preventative measures, and guide treatments for diseases in the clinic. Many advances in human genetics, including genome-wide association studies (GWAS), clarified somewhat the correlation between genetic variation and trait variation. GWAS take measurements of genetic variants at hundreds of thousands to millions of locations (loci) in the genomes of study subjects and test these variants for statistical associations with phenotypic measurements from the same individuals. GWAS have been performed for hundreds of human traits and diseases [2], revealing that many phenotypes, despite being highly heritable, often have polygenic genetic architectures. Such complex traits result from the actions of and interactions between many genes, in contrast to monogenic phenotypes that result from mutations in a single gene. One example of a highly heritable complex human phenotype is height, which varies considerably across human populations and results from the aggregate effect

of hundreds of genetic variants throughout the genome, each contributing small effects and interacting with each other and the environment [3].

In GWAS of complex traits, trait-associated genetic loci that fall within protein-coding regions of the genome can be biochemically predicted to directly impact the structure and function of proteins. However, most trait-associated loci fall within non-coding regions, which make up over 99% of the human genome [2]. While it is difficult to predict and test the functional consequences of variants in non-coding regions, these regions may impact phenotypes indirectly by regulating the expression of gene products rather than their structure. Determining which gene is impacted by a particular trait-associated variant is also a challenging task, as regulatory sequences may be far from the genes they impact [4,5]. Additionally, isolating the causal variant driving a trait association is further complicated by linkage disequilibrium, or the tendency of neighboring genetic variants to correlate with each other due to co-segregation during meiotic recombination [6]. Therefore, more broadly identifying which non-coding loci impact gene regulation is currently an active area of study in human genetics, which will help to clarify the mechanisms of action by which trait-associated non-coding genetic loci impact phenotypes [7,8].

1.2 Identifying the potential functional role of trait-associated variants

Regulatory quantitative trait locus (QTL) mapping is one method that is helping to elucidate the potential functional role of genetic variants in non-coding genomic loci. Much like in GWAS, QTL approaches seek to statistically associate variation at genomic loci with phenotypic variation. However, the phenotypes of interest for QTL analyses are typically intermediate regulatory molecular phenotypes such as RNA or protein expression levels [9–11].

Thus, regulatory QTLs can offer potential explanations as to how trait-associated genomic variants may mechanistically impact their associated traits. For example, suppose a genetic locus is associated with both a trait and the transcriptional expression levels of a gene, making it both a GWAS locus and an expression QTL (eQTL). In that case, it is plausible that this genetic locus impacts its associated trait by regulation of an intermediate gene expression phenotype. This potential mechanism becomes more credible when the locus acts as an eQTL in a relevant cell type (e.g., developing bones) and acts on relevant genes (e.g., genes regulating growth and development). Regulatory QTLs for gene expression, splicing, and chromatin accessibility are all enriched among trait-associated genetic variants [12,13], suggesting that genetic variants have the potential to mediate their effects on complex traits through many mechanisms of gene regulation. In addition to clarifying these mechanisms, regulatory QTLs can also help to reveal novel trait-associated loci. For instance, GWAS are often underpowered to detect the many genetic loci associated with complex traits because these loci contribute very small effects to the overall heritability of their associated traits [14]. However, the statistical power of this approach can be increased by incorporating prior information of which genetic loci have evidence for functional effects [15]. Thus, integrating regulatory QTL (especially eQTL) studies with GWAS has both the potential to enhance the discovery of complex trait-associated genetic loci and to uncover an understanding of the biology underlying those complex traits.

When designing regulatory QTL analyses to examine trait-associated loci, an important consideration is the cell and tissue type of interest. While the genetic sequences of cells throughout an organism's body are almost entirely identical, gene regulatory patterns can differ substantially from tissue to tissue and from cell type to cell type [16]. Depending on the complex trait of interest, functionally consequential regulatory changes may be isolated to specific cell

types or tissues, making it essential to study regulatory QTLs in many cell types. One coordinated effort to broadly study eQTLs in dozens of human tissues was the Genotype-Tissue Expression (GTEx) consortium [17], first established in 2010 and completed in 2020. The GTEx project collected RNA sequencing and genotype data from 54 adult tissues and cell lines across 838 donors. With these data, the project has mapped over 4.2 million eQTLs in at least one of the studied tissues, including at least one eQTL for 94.7% of all protein coding genes, representing a large step towards understanding the cell type specificity of gene regulatory programs in human cell types [17]. However, while the great wish of QTL analyses is to explain trait associations, applying even this massive collection of detected eQTLs to existing GWAS loci has been less than satisfying. Only 43% of GWAS disease-associated loci are also classified as GTEx eQTLs in any tissue [17]. While this low proportion may be attributed to a lack of power with current eQTL study sample sizes or to the contribution of other regulatory QTLs aside from eQTLs, there may be further explanations to clarify this gap [18].

1.3 Context-specificity of gene regulation and dynamic eQTLs

Context-specific gene regulation likely contributes to the gap in disease-association eQTLs. Much like different cell and tissue types, a cell's environmental context and state are also integral determinants of gene regulation. An expansive definition of cell state encompasses cell type, the stage of differentiation, and the cellular environment brought on by infection, local nutrient availability, signaling, or other perturbations. Genetic variants associated with complex traits may only impact gene regulation under specific conditions. Thus, it is possible that condition-specific eQTLs play essential roles in mediating the effects of trait-associated genetic loci. Prior eQTL research has predominantly focused on cells and tissues collected at steady-state

conditions (static eQTLs), which are often shared across tissues and are not disease- or context-specific [5,18]. However, as gene regulation is highly dynamic and context-specific, the potential of identifying eQTLs throughout continuous cellular differentiation and environmental perturbation states has not been fully realized.

Furthermore, non-steady state environmental contexts, including cellular stress, infection, or chemical stimuli, may represent selective pressures that have shaped inter-population and potentially inter-individual genetic variation [19]. Along this line, dynamic eQTLs, or eQTLs associated with gene expression in certain developmental or environmental states but not others, can better explain trait-associations than static eQTLs that have been identified in more powerful studies [16,20,21]. Therefore, it becomes important to consider the nature of gene regulation in disease-relevant cell types and within dynamic temporal and environmental contexts to more fully understand the genetic basis of complex human traits and diseases. As such, dynamic eQTL studies offer a potentially fruitful complementary approach to solving the eQTL-trait associated variant gap as opposed to increasing the size of static eQTL studies.

1.4 Genetic and environmental underpinnings of osteoarthritis susceptibility

One complex human disease where both genetic variation and environmental context is highly relevant is the joint disease osteoarthritis (OA). OA is characterized by progressive articular cartilage degeneration, altered bone structure resulting from changes to the biomechanics of movement at affected joints, pain, and low-grade chronic joint inflammation. OA affects 12% of adults in the United States between 25-74 years of age [22], and this proportion is expected to increase as the population ages and the prevalence of obesity increases. There are no disease-modifying treatments for this disorder, which often results in disability and

joint replacement. Although initially described as a disease of "wear and tear," OA is now known to result from perturbations of joint homeostasis. These perturbations trigger active disease processes, including excessive degradation of extracellular matrix [23], hypertrophy and mineralization of chondrocytes (the primary cells of cartilage) [24], and increased apoptosis of chondrocytes [25,26]. Furthermore, gene expression patterns differ between intact and degraded cartilage in OA patients, reflecting the activation of biological pathways associated with OA [27,28].

Genetic variation plays a meaningful role in OA risk. The heritability of OA, estimated through familial and twin studies, is between 40-70% depending on the affected joint [29,30]. Large-scale GWAS of OA have identified over 100 disease-associated variants at 95 independent loci throughout the genome, meaning OA is a highly polygenic disease [31,32]. As these known OA-associated loci still only account for around one-fifth of the measured heritability of OA, though, there remain many more loci to be discovered [33].

The effect sizes of individual OA-associated variants are typically small, and they usually reside in non-coding regions of the genome, with less than 10% of OA-associated variants found in protein-coding sequences [32]. However, many OA-implicated variants are located near genes that code for protein components of the extracellular matrix that encase chondrocytes in cartilage tissue, genes that are integral during cartilage and bone development, and genes involved in mechanosensation and mechanoadaptation [31]. OA-associated variants are also enriched in gene regulatory elements that impact chondrocyte development and joint shape formation [34,35], suggesting that genetic contributions to OA risk may mediate their effects at least partly during joint development rather than solely in later life.

While genetics seem to have a clear influence on OA risk and susceptibility,, observational and experimental evidence also points to the importance of environmental factors in determining joint disease risk. Two important factors include mechanical stress and inflammation, and it remains a question of active debate within the field as to whether OA is primarily a mechanical or inflammatory disease.

1.4.1 The influence of the joint mechanical environment on OA

Biomechanical factors are particularly interesting in OA as joint health is impacted negatively by excessive [36–39] or insufficient [39–41] mechanical loading. In animal models, surgically-induced joint instability leads to changes in biomechanical loading at joints that result in the development of skeletal changes that mimic early human OA [42–49]. Similarly, human joint developmental abnormalities, such as hip dysplasia that alters the load distribution at joints with an increased cumulative contact stress-time, are also associated with increased risk for OA [50–53]. Biomechanical stress caused by acute joint injuries can also lead to post-traumatic OA, which is responsible for 12% of the overall burden of OA [54]. Furthermore, even in uninjured human joints, observational studies of physically demanding occupations, ranging from miners to professional sports players, suggest that repetitive excessive joint loading is associated with early onset of OA [55–61]. Biomechanics can promote healthy homeostasis in normal loading conditions and lead to dysregulation and degeneration in non-physiologic loading conditions. This external force may mediate its effect on joints through gene regulation.

Studying molecular responses to mechanical stress in the context of OA development at an organismal level is difficult. Still, the impact of biomechanical factors on joint health has been modeled in the lab. For example, *in vitro* models have measured the response of joint cells to

mechanical stress treatments, including the effect on cultured primary joint cell viability, joint chondrocyte calcification, and developmental regulatory factor expression in chondrocytes [62–67]. Additionally, *in vitro* biomechanical stress models using cultured primary joint cells have been developed that use specific types and patterns of biomechanical strain to induce transcriptional and biochemical changes characteristic of human OA [68–71]. Altogether, these systems allow for further investigation of the mechanisms by which mechanical forces contribute to the pathogenesis of OA, particularly at load-bearing joints such as the hips and knees.

1.4.2 The influence of the joint inflammatory environment on OA

While the appreciation for mechanical forces in OA development and progression is longstanding, it has also been long known that low grade chronic inflammation is a common symptom of this disease and that it plays a role in the process of cartilage degeneration and repair, even in early stages [72–75]. The inflammation observed in OA is characterized by an innate immune response rather than an adaptive autoimmune response in which the body's immune cells attack its own joint tissues as is the case in rheumatoid arthritis.

Joint inflammation, also known as synovitis, can be triggered by factors like biomechanical stress, which stimulate immune and synovial cells to release early-stage inflammatory cytokines such as interleukin-1 beta (IL-1β) and downstream inflammatory molecules [76]. Subsequently, inflammation pathways initiated by immune cells induce the expression of catabolic enzymes in chondrocytes, including matrix metalloproteinases, that work to degrade the extracellular matrix surrounding chondrocytes in cartilage [77,78]. Inflammatory-stimulated chondrocytes also make specific pro-inflammatory molecules that help to initiate and

perpetuate low grade inflammation in the OA joint [79,80]. Inflammation can be triggered by age or conditions that create a pro-inflammatory environment such as obesity. For example, the metabolic disorders diabetes or obesity, which are known to increase inflammatory activity throughout the body, increase risk for OA in non-load-bearing joints like the hands [81]. Therefore, although increased biomechanical stress on load-bearing joints in obesity may directly or indirectly contribute to OA through mechanically-induced inflammatory responses, the close link between obesity and OA may also be due to the high levels of circulating pro-inflammatory cytokines common in this metabolic disease, which can lead to increased synovitis [82].

In animal models and *in vitro* models of OA, exposing joint cells to the inflammatory cytokines IL-1β and TNFα suppresses the synthesis of extracellular matrix components such as proteoglycans and type II collagen, induces chondrogenic apoptosis, and promotes the expression of matrix metalloproteinases by chondrocytes [77,78,83–87], all processes observed in human OA. Overall, inflammation mediated through the release of inflammatory cytokines by immune cells and joint cells contribute to the pathogenesis and progression of OA. Moreover, *in vitro* inflammatory cytokine treatment systems allow for further investigation of the mechanisms by which inflammatory forces contribute to the pathogenesis of OA.

1.5 iPSC-derived cell types for quantitative genetic studies of gene regulation

Given the complex combination of genetic and environmental factors that play a driving role in susceptibility to OA, this disease could benefit from additional studies of gene regulation and gene-by-environment interactions. Nevertheless, it is difficult to obtain adequate high-quality cell and tissue samples from many healthy human joints. Indeed, despite its expansive survey of tissues and cell lines, the GTEx project does not include any gene expression data from

bone and cartilage [17]. Conversely, *in vitro* differentiated cell types have great potential to expand the scope of eQTL studies, allowing access to cell types that are otherwise difficult to obtain due to obstacles surrounding their sample collection or their rarity in adult human subjects.

Such cell types can be differentiated from induced pluripotent stem cells (iPSCs), which are cells reprogrammed from adult somatic cells into a pluripotent state [88]. iPSCs can be generated ethically and noninvasively from many volunteer subjects by inducing the transient expression of a few transcription factors in adult cell types like skin fibroblasts or transformed B cells [89–91]. In addition to their ever-increasing accessibility, iPSCs can also self-replicate and be cryopreserved, meaning they can serve as a dependable renewable resource for functional genomics studies. Further, differentiating iPSCs into desired cell types using established protocols allows otherwise difficult to access cells to be studied in the lab [92]. Of note, iPSCs and iPSC-derived cell types reliably recapitulate the gene expression patterns of their primary cell counterparts and have been successfully used to identify several types of static and dynamic regulatory QTLs [16,20,93–96]. Finally, the existence of genetically diverse iPSC panels makes performing QTL studies in differentiated cells a realistic possibility, allowing for studies of the genetic basis of gene regulation and gene-by-environment interactions in many different cell types [96,97].

1.6 Dissertation overview

The work I will present in this dissertation examines the role of gene-by-environment interactions in determining inter-individual susceptibility for OA. In chapter two, I describe the establishment and characterization of a system for investigating gene regulation and gene-by-

environment interactions in iPSC-derived chondrogenic cells (iPSC-chondrocytes) and *in vitro* environmental treatments. I also describe the potential of this system for use in studies of gene regulation relevant to OA. In chapter three, I describe the use of this system to identify differential gene regulation responses between two established models of OA. I also use a quantitative genetics approach to identify genetic variation associated with gene expression variation in iPSC-chondrocytes in several different environmental contexts, as well as context-specific associations that represent the effects of gene-by-environment interactions on gene regulation. Overall, these analyses establish iPSC-derived joint cells and *in vitro* OA treatments as tractable and powerful models of context-specific gene regulation for improving our understanding of joint health and OA.

CHAPTER 2

CHARACTERIZING GENE EXPRESSION IN AN *IN VITRO* BIOMECHANICAL STRAIN MODEL OF JOINT HEALTH

2.1 Abstract[1]

Both genetic and environmental factors appear to contribute to joint health and disease. For example, pathological levels of biomechanical stress on joints play a notable role in initiation and progression of osteoarthritis (OA), a common chronic degenerative joint disease affecting articular cartilage and underlying bone. Population-level gene expression studies of cartilage cells experiencing biomechanical stress may uncover gene-by-environment interactions relevant to human joint health. To build a foundation for population-level gene expression studies in cartilage, we applied differentiation protocols to develop an *in vitro* system of chondrogenic cell lines (iPSC-chondrocytes). We characterized gene regulatory responses of three human iPSC-chondrocyte lines to cyclic tensile strain treatment. We measured the contribution of biological and technical factors to gene expression variation in this system. We identified patterns of gene regulation that differ between strain-treated and control iPSC-chondrocytes. Differentially expressed genes between strain and control conditions are enriched for gene sets relevant to joint health and OA. Furthermore, even in this small sample, we found several genes that exhibit inter-individual expression differences in response to mechanical strain, including genes previously implicated in OA. Expanding this system to include iPSC-chondrocytes from a larger number of

[1] This chapter is reproduced, with some modification, from Hung A, Housman G, Briscoe EA *et al*. Characterizing gene expression in an *in vitro* biomechanical strain model of joint health [version 1; peer review: 1 approved with reservations.] F1000Research 2022, 11:296 (https://doi.org/10.12688/f1000research.109602.1)

individuals will allow us to characterize and better understand gene-by-environment interactions related to joint health.

2.2 Introduction

Disorders of the joints can often lead to pain and disability and have far-reaching impacts on quality of life. For example, osteoarthritis (OA) is a chronic degenerative joint disease characterized by defects in articular cartilage integrity and alterations to underlying bone structure [98]. OA is a major cause of disability in older adults and impacts approximately 300 million people worldwide [99]. There are no disease-modifying treatments for this painful disorder, and its specific pathogenic mechanisms are still under investigation.

Genome-wide association studies (GWAS) have identified over 86 genetic loci associated with OA risk [33]. Most of these loci fall within non-coding regions of the genome and have eluded functional characterization. Therefore, it remains unclear how associated genetic factors modulate OA onset and progression. One possibility is that regulatory changes in key structural and metabolic genes may modulate OA-related outcomes. Regulatory changes occurring in response to relevant environmental factors, including biomechanical stress, may be particularly important. Indeed, gene expression studies have identified broad patterns of gene expression that differ markedly between healthy and osteoarthritic human cartilage [27,28]. These gene expression differences reflect activation of biological pathways associated with joint disease, suggesting that studies of gene regulation in cartilage and other skeletal tissues are valuable for understanding normal joint health and joint disease and pathogenesis in joint conditions like OA.

However, few studies have measured gene regulatory phenotypes in human skeletal tissues or cells. Even the Genotype-Tissue Expression (GTEx) Project, one of the largest efforts

to examine gene expression variation across human tissues and cell types, does not include samples from cartilage [17]. This lack of tissues is partially due to the practical limitations and ethical issues associated with collecting healthy, high-quality cartilage samples from human donors. Nevertheless, protocols to differentiate induced pluripotent stem cells (iPSCs) into cells relevant to joint health and disease, such as chondrocytes (the primary cells of cartilage), exist [92,100], and these methods can circumvent some of the challenges associated with inaccessible primary tissues.

iPSC-derived cells also allow for the study of dynamic cellular responses to specific environmental conditions. It has become increasingly evident that studying gene regulation in disease-relevant states is crucial for understanding the genetic basis of disease [18]. Thus, numerous studies have begun identifying dynamic regulatory expression quantitative trait loci (eQTLs) in various cell types and contexts, including drug-induced cardiotoxicity [21], cardiomyocyte differentiation [20], vitamin D exposure [101], and response to infection [19,102–105]. These studies highlight the merits of exploring gene regulation beyond steady-state conditions.

In human joints, biomechanical stress is a particularly relevant environmental condition. Joint health deteriorates in response to excessive or insufficient amounts of mechanical loading [36,37,39–41]. Further, biomechanical factors may impact gene expression regulation in joint tissues and may interact with genetic factors to impact risk for joint diseases [106]. Such interactions are difficult to examine *in vivo*. However, iPSC-derived chondrogenic cells (iPSC-chondrocytes) provide an alternative system in which to study the effects of cyclic tensile strain (CTS), a type of controlled biomechanical stress regimen designed to induce joint disease-like phenotypes [68–71]. Thus, iPSC-chondrocytes offer an opportunity to study gene expression

responses to joint disease-relevant states. Studies of iPSC-chondrocytes may also help uncover mechanisms through which OA-associated genetic loci modulate OA outcomes.

Both chondrocyte differentiation protocols and methods for inducing CTS *in vitro* existed prior to this study. Still, little is known about the suitability of this system for studies of gene regulatory dynamics. Therefore, we designed a study using human iPSC-chondrocytes to examine the combined effects of genetic variation and biomechanical stress on gene regulation during CTS. Through this study, we evaluated whether the process of chondrocyte differentiation is robust to individual differences. We also ascertained whether iPSC-chondrocytes exhibit a robust gene expression response to CTS. Finally, we determined whether expanding the sample size of this experimental system might further improve our understanding of gene-by-environment interactions within joint health.

2.3 Results

We designed this study to determine whether iPSC-chondrocytes are a useful system for studying gene-by-environment regulatory interactions relevant to joint health. First, we asked whether the efficiency of chondrocyte differentiation is similar in different individuals. Next, we evaluated the effects of CTS on gene regulation in iPSC-chondrocytes to determine whether this system is suitable for studying gene regulatory effects on joint health. Finally, we estimated the contribution of sample and batch effects to variation in gene expression response to CTS, to assess the suitability of our iPSC-chondrocyte system for response eQTL mapping studies.

2.3.1 Study design and data collection in the iPSC-based *in vitro* system

We used three human iPSC lines that were previously established and characterized as part of a panel of iPSCs derived from Yoruba individuals [96]. We differentiated the iPSCs along the chondrogenic lineage with an intermediate differentiation step into mesenchymal stem cells (MSCs; Figure 2.1a) using previously established protocols [92]. iPSC-derived MSCs (iPSC-MSCs) exhibited phenotypes and cell surface marker expression patterns characteristic of primary MSCs [107] (Supplementary Figure 2.1). iPSC-chondrocytes showed a modest increase in collagenous extracellular matrix (ECM) production as compared to matched iPSC-MSCs (Figure 2.1b; Supplementary Figure 2.2a). Additionally, immunostaining for COL2A1 of iPSC-chondrocytes from one individual demonstrated increased expression compared to matched iPSC-MSCs (Supplementary Figure 2.2b).

We treated iPSC-chondrocytes from each individual with 24 hours of CTS, which is known to induce a hypertrophic phenotype in cartilage [68–71]; see Methods). We simultaneously kept a second, matched set of untreated iPSC-chondrocytes in the same incubator for the same period as a control. We performed three technical replicates of this experiment, starting with MSCs from the same cryopreservation batch and carrying out an independent differentiation of the MSCs to iPSC-chondrocytes in each replicate. We extracted bulk RNA from all biological and technical iPSC-chondrocyte treated and untreated replicates (n=9). We also collected single cell RNA sequencing (scRNA-seq) data from one technical replicate (n=3) using the 10X Genomics platform.

We could not detect differential expression between treatment conditions for markers of chondrocyte degeneration and hypertrophy in the bulk RNA sequencing data. This is because these genes are expressed at a low level across all samples and are filtered out in pre-processing steps. To evaluate the changes in the expression of these gene markers between treatment

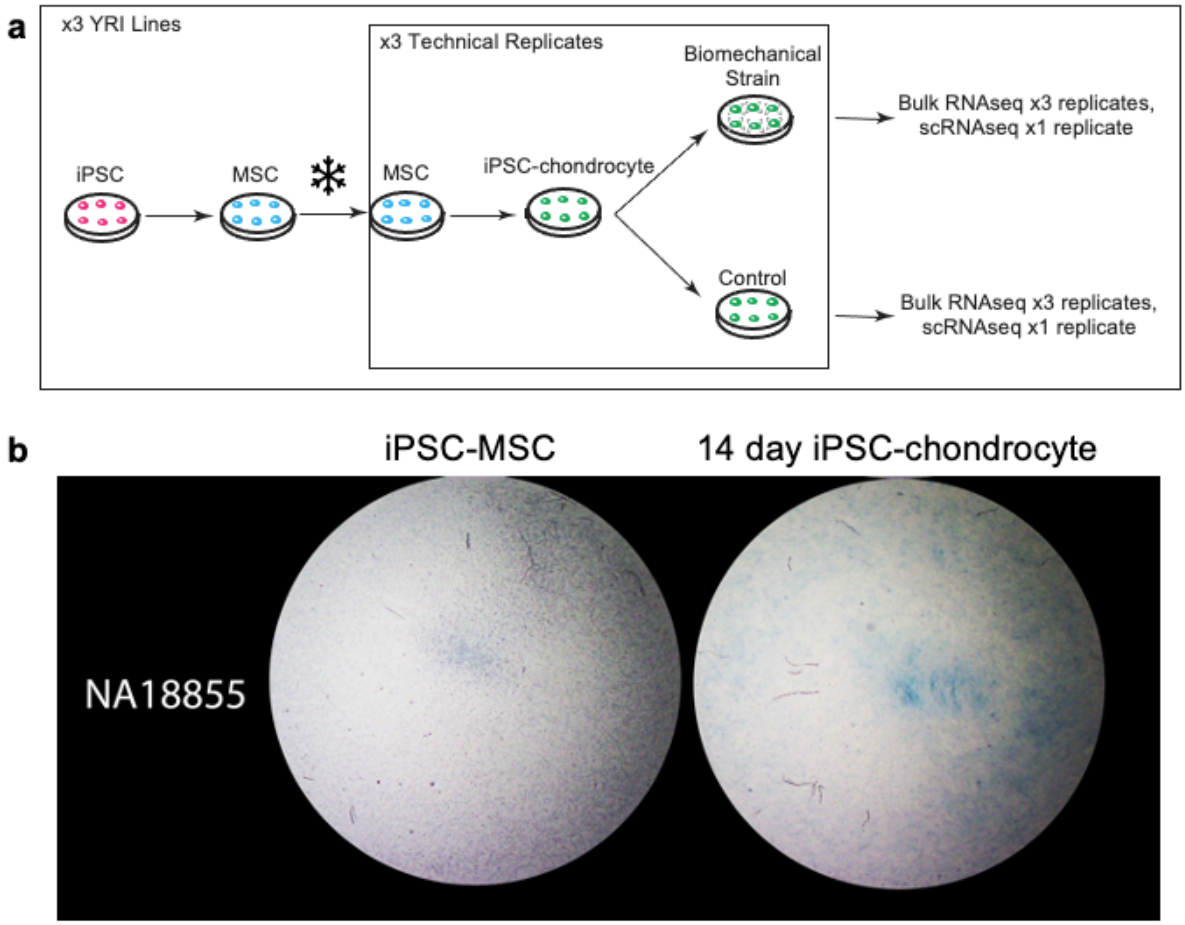


Figure 2.1: Description of *in vitro* biomechanical strain study design.

(a) In our study design, iPSCs generated from three Yoruba individuals were first differentiated along the chondrogenic lineage with an intermediate differentiation step into mesenchymal stem cells (MSCs). iPSC-derived MSCs from each individual were cryopreserved. For each replicate of the experiment, iPSC-MSCs from the same cryopreservation batch were differentiated into iPSC-chondrocytes over a period of 14 days. Subsequently, we treated iPSC-chondrocytes from each individual with 24 hours of a CTS treatment that is known to induce an OA-like phenotype. We simultaneously kept a second, matched control set of iPSC-chondrocytes in the same incubator for the same period of time, but without CTS treatment. We performed three technical replicates of this experiment, starting with MSCs from the same cryopreservation batch. Following strain-treated and control conditions, we extracted bulk RNA from all biological and technical iPSC-chondrocyte replicates and collected scRNA-seq data from one technical replicate from each cell line using the 10X Genomics Chromium Single Cell Gene Expression platform. **(b)** Representative images of Alcian blue staining of 14-day iPSC-chondrocytes and matched iPSC-MSCs demonstrating increased proteoglycan production in iPSC-chondrocytes. Images are cropped to show the central seeded area of wells of BioFlex Type I Collagen coated 6-well Culture Plate (seeded area diameter 25mm). Additional images for other cell lines are available in Supplementary Figure 2.

conditions more sensitively, we performed quantitative real-time reverse transcription PCR (RT-PCR) on control and strained iPSC-chondrocytes. We detected modest increases in expression of gene markers of chondrocyte hypertrophy *MMP1* and *MMP13* in response to CTS [69,108], while gene marker *TIMP2* did not change in expression, in line with previous reports (Supplementary Figure 2.3) [69].

2.3.2 iPSC-chondrocytes likely represent an early stage of chondrogenesis

As a next step in our analysis, we confirmed that iPSCs successfully differentiated to chondrogenic cells. Using standard staining protocols (see Methods), we demonstrated that our cells produce glycosaminoglycan ECM, a hallmark of chondrogenesis (Figure 2.1b; Supplementary Figure 2.2a). Our cells also produce the collagen COL2A1, a protein that is almost exclusive to cartilage tissues [109,110] (Supplementary Figure 2.2b). We also used our scRNA-seq data to address two major questions: First, what is the approximate proportion of iPSCs that differentiated into chondrogenic cells in each individual? Second, what is the relative maturity of iPSC-chondrocytes?

We expected that chondrocyte differentiation might result in heterogeneous populations of cells at different stages along the chondrogenic lineage and that iPSC-chondrocyte purity might differ across cell lines. We used scRNA-seq data to assess potential differences in differentiation efficiency among the three individuals. Following standardization and normalization (see Methods), unsupervised clustering of the single cell data revealed three clusters of cells in our untreated control samples (Figure 2.2a). The proportion of cell membership in each cluster is comparable across individuals (Figure 2.2b). Five percent of cells from individual NA18855 fall into cluster 2, along with 8% of cells from NA18856 and 7% of

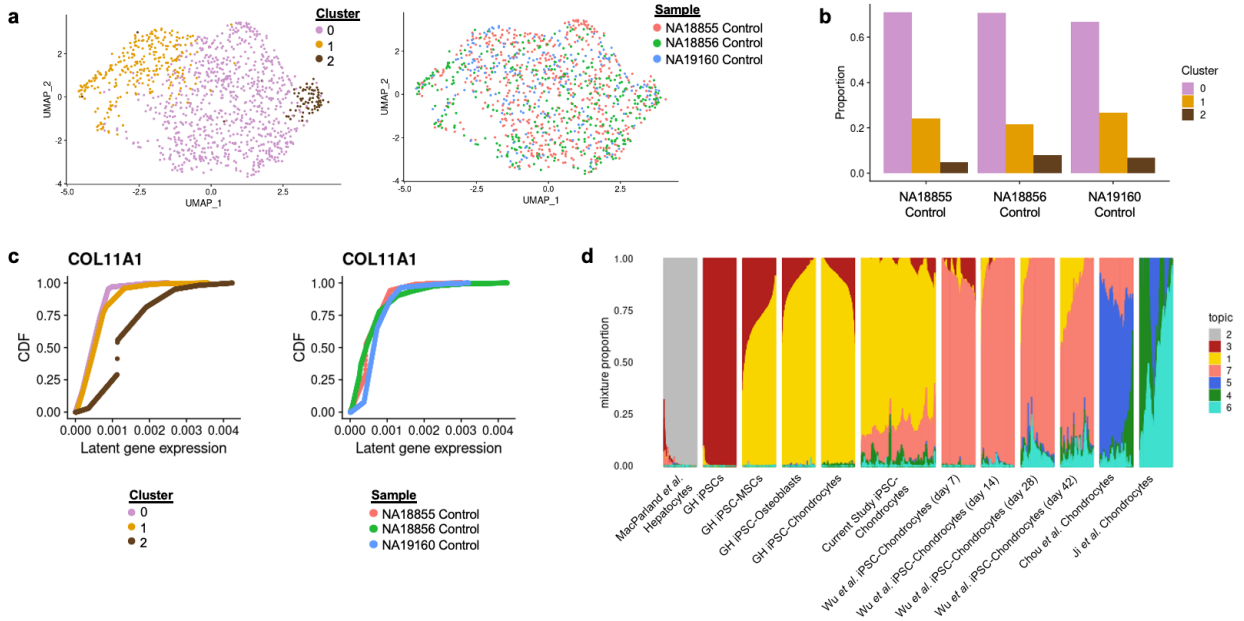


Figure 2.2: Characterization of cell type composition in iPSC-chondrocyte cultures.

(a) Uniform Manifold Approximation and Projection (UMAP) of normalized and integrated single cell RNA sequencing data from control samples from all three individuals included in the study. (Left) UMAP colored by sample label. Cells from different samples are well-integrated. (Right) UMAP colored by Seurat cluster determined from normalized gene expression data. **(b)** Proportion of membership of cells from each individual in each Seurat cluster. The relative membership in each Seurat cluster is comparable between individuals. **(c)** Cumulative distribution function (CDF) of marginal distributions of latent gene expression of *COL11A1* determined through fitting a Poisson adaptive shrinkage model to raw gene expression counts in each sample. (Left) CDF curves colored by Seurat cluster. Cells in cluster 2 contain a higher latent gene expression of *COL11A1* on average. (Right) CDF curves colored by Individual. **(d)** STRUCTURE plot representing the relative proportional membership of single cells (columns) in 7 different topics in a topic model fit to scRNA-seq data derived from iPSC-chondrocytes collected in this study ("Current Study iPSC-Chondrocytes"); matched iPSCs, iPSC-MSCs, iPSC-osteoblasts, and iPSC-chondrocytes collected from a single individual ("GH iPSCs, GH iPSC-MSCs, GH iPSC-Osteoblasts, GH iPSC-Chondrocytes"); primary hepatocytes ("MacParland *et al*. Hepatocytes"); a time-course of iPSC-Chondrocyte pellet culture differentiation through the use of pellet culture ("Wu *et al*. iPSC-Chondrocytes (day 7 – day 42)"; and primary adult chondrocytes described in two separate publications ("Chou *et al*. Chondrocytes", "Ji *et al*. Chondrocytes"). The ordering of individual cells within each study are determined through a one-dimensional tSNE algorithm applied to topic memberships of each cell. For each data source and cell type on the x-axis, a random subset of 800 cells is plotted with the exception of Current Study iPSC-Chondrocytes, for which all 1,815 cells are plotted.

cells from NA19160. Based on gene expression patterns, cluster 2 consists of cells that are most like chondrocytes; for example, these cells show high expression of *COL11A1*, an essential gene for normal cartilage collagen fibril formation [111] (Figure 2.2c; see Methods; additional genes in Supplementary Figure 2.4). In contrast, clusters 0 and 1 contain more MSC-like cells, characterized by expression of MSC genes and lower expression of chondrogenic genes. We found no substantial difference in *COL11A1* expression between individuals. Based on staining and gene expression patterns, iPSC-chondrocytes are undergoing chondrogenesis, and cell type composition does not differ substantially among the three individuals in this study.

In addition to discerning heterogeneity in our samples, we also evaluated the relative maturity of our iPSC-chondrocytes. To do this, we used topic modeling, analyzing our single cell data along with single cell datasets from multiple cell types, including primary adult chondrocytes (see Methods). Topic modeling is an unsupervised classification approach that, when applied to single cell gene expression data, allows one to find recurring patterns of gene expression, or topics, present across a collection of cells. By allowing each cell to have grades of membership in multiple topics simultaneously, rather than assigning cells to only one cluster [112], topic modeling can identify both discrete and continuous variation between cells.

A model fit with seven topics to the combined dataset shows that both iPSC-chondrocytes and primary chondrocytes are equally reliably distinct from unrelated cell types (e.g., hepatocytes). iPSC-chondrocytes retain a large proportion of gene expression patterns characteristic of iPSC-MSCs (topic 1), but they also possess certain gene expression patterns seen in adult primary chondrocytes and in iPSC-chondrocytes differentiated through a chondrogenic pellet (topics 4, 6, and 7) (Figure 2.2d). Topics 4-7 display relatively high levels of expression of several markers of chondrogenesis or cartilage fate, including *SOX9*, *SOX5*, *SOX6*

[113], and *COL9A1* [114] (Supplementary Figure 2.5). Furthermore, differential expression analyses across topics identified type IX collagen gene *COL9A3* [114] as highly occurring in topic 7 relative to all other topics (Supplementary Figure 2.5). Similarly, in topics 4 and 6, the chondrogenic marker *COMP* [115] has a high occurrence relative to all other topics. Thus, we conclude our iPSC-chondrocytes are likely in the early stages of chondrogenesis and are readily distinguishable from iPSCs and iPSC-MSCs.

2.3.3 Analysis of bulk RNA sequencing data

After confirming we can generate chondrogenic cells from iPSCs, we next sought to understand gene expression variation in this system. For this we focused on bulk RNA-seq data collected from all replicates. We generated an average of 22.3M raw reads per sample (s.d. 4M reads). We excluded one sample from further analyses because it displayed a particularly low percentage of mapped reads. We note the mapped reads from this sample cluster as expected with other technical replicates from the same individual and treatment (Supplementary Figure 2.6), but we still excluded it because it failed standard QC metrics. We filtered the remaining data for lowly expressed genes and standardized gene counts with respect to library size (see Methods).

As a first step of our analysis of the bulk data, we identified gene expression responses to strain treatment. We used the limma R package to fit a linear mixed model for each of the 10,486 expressed genes in the filtered bulk RNA-seq data, accounting for the random effect of experimental batch and the fixed effects for individual cell line, sex, treatment status, two factors of unwanted variation, and RIN score (see Methods). Using this model, we tested for differential expression between treated and untreated cultures. At an FDR of 0.05, 987 genes are

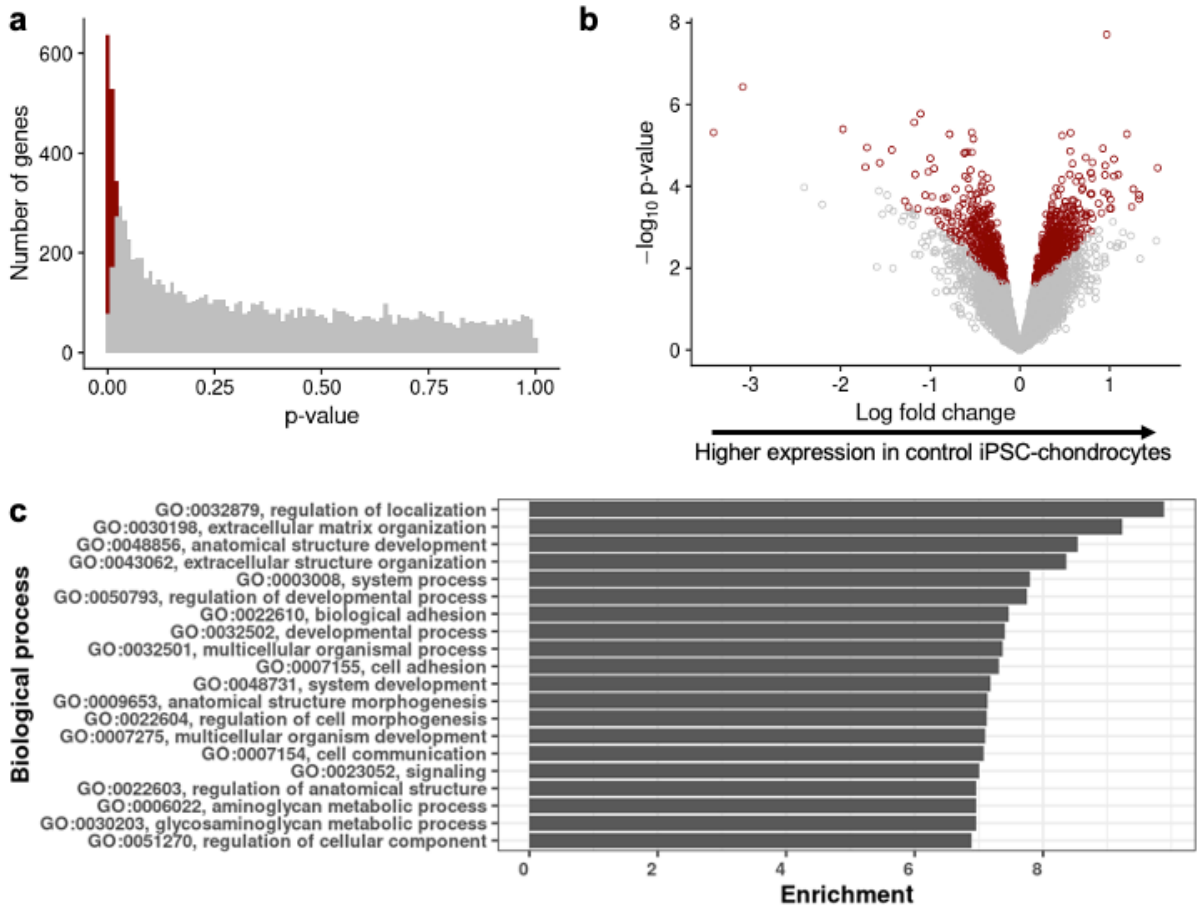


Figure 2.3: Results from linear mixed model differential expression (DE) analysis between treatment conditions.

The linear mixed model used to conduct the analysis also accounted for the random effect of experimental batch and the fixed effects of individual cell line, sex, treatment status, two factors of unwanted variation, and RIN score. **(a)** Histogram of raw p values from DE analysis conducted independently for each gene between control and strain-treated conditions. Highlighted in red are genes with an FDR-adjusted p value < 0.05 (987 of 10,486 tested genes). **(b)** Volcano plot of –log10 raw p values vs log-fold change between treatment conditions. Highlighted in red are genes with an FDR-adjusted p value < 0.05. Genes plotted to the right of 0 on the x-axis represent genes with higher average expression in control iPSC-chondrocyte samples compared to strain-treated samples. **(c)** Top 20 Biological processes enriched among DE genes compared to background set of 987 genes. These GO biological process terms include those related to extracellular matrix organization and metabolism of extracellular matrix structure.

significantly differentially expressed (DE) between treated and untreated chondrogenic cells (Figure 2.3a-b).

To evaluate the potential relevance of the DE genes to joint health and OA, we first considered their enrichment among gene ontology (GO) terms. The top 20 most highly enriched GO biological process terms include those related to ECM organization and metabolism of ECM structure (Figure 2.3c). These functions make intuitive sense given that ECM homeostasis is important for joint cartilage health; moreover, imbalances in this homeostasis are associated with OA [23,116].

We also determined whether the DE genes may be overrepresented in gene sets previously implicated in OA. We examined results from one of the largest GWAS for OA susceptibility, which identified 64 independent significant associations with OA [33]. We used a Fisher's exact test to assess enrichment of DE genes among a set of 553 genes located within 500 kb of the 64 associated loci. These 553 genes were also identified as having prior evidence of involvement in animal models of skeletal disease or human bone diseases [33]. We found that DE genes in our study are significantly enriched within the set of 553 genes previously associated with OA ($p = 0.002$).

Next, we evaluated results from a separate study, which profiled mRNA and protein samples in low-grade and high-grade osteoarthritic cartilage from 115 patients undergoing joint replacement [28]. Steinberg *et al*., 2019 found 409 genes with evidence of significant differential expression between patients with low-grade and high-grade osteoarthritic cartilage, at both the RNA and protein levels. Though causality is difficult to infer, this observation suggests at least a subset of these genes is involved in OA cartilage degradation. A Fisher's exact test reveals that

our DE genes are also significantly enriched among this gene set ($p = 0.02$). Of note, the two genes that overlap between the two external gene sets, *LTBP3* and *LAMC1*, are also DE our study. *LTBP3* is a regulator of the TGF-ß signaling family, which plays roles in cartilage formation and development [117]. *LAMC1* has been identified as a blood-based biomarker for detecting mild knee OA, with lower RNA expression identifying mild OA [118]. Based on these GO and gene set enrichment results we concluded that the DE genes identified between strain-treated and control iPSC-chondrocytes are relevant to joint health.

Due to differential power, highly expressed genes are more likely to be detected as DE than lowly expressed genes in any RNA sequencing dataset [119]. Therefore, it is possible the magnitude of expression of different genes in our data can explain our enrichment results. To assess the robustness of our findings, we permuted the labels of treatment condition among our samples and re-performed DE and enrichment analyses a total of ten times. In nine permutations, we failed to identify ECM-related GO terms among the top 20 enriched terms (one permutation revealed two ECM-related GO terms). Further, we did not find any enrichment of DE genes within the two OA-relevant gene sets using permuted data. As our permuted data do not display the same enrichment patterns as our actual data, we concluded our results are not due to differential power to detect DE.

2.3.4 Certain gene expression responses to stress are heterogeneous between individuals

In our DE analysis, we focused on identifying inter-treatment differences in gene expression rather than inter-individual differences. Ultimately, we would like to use this system to study gene-by-environment interactions, which occur at the intersection of inter-treatment and inter-individual differences. A gene-by-environment interaction occurs when the magnitude or

direction of gene expression response to an environmental stimulus is associated with an individual's genotype at a particular locus. The sample size of this current study is far too small to detect gene-by-environment interactions. Still, we identified several genes that exhibit inter-individual differences in expression in response to CTS (Figure 2.4).

For example, *MMP14* displays a different pattern of expression in each cell line before and after CTS (Figure 2.4a): *MMP14* expression remains constant between control and strain-treated NA18855 cells, is upregulated in strain-treated NA18856 cells, and is downregulated in strain-treated NA19160 cells. *MMP14* is expressed constitutively in adult joint cartilage and upregulated in diseased states [120]. In addition, *EXOSC8* and *COPG1* (Figure 2.4b-c) are both involved in the formation of secretory vesicles originating from the Golgi complex. These genes also display differences in direction or magnitude of gene expression response to CTS between individuals. If heterogenous responses to biomechanical stress exist more broadly and are associated with genotypic differences, this experimental system will be able to identify them in population-level eQTL studies.

2.3.5 Sources of variation in bulk RNA sequencing data

Thus far, our results show that CTS elicits a robust, joint health-relevant gene expression response in iPSC-chondrocytes, and that, anecdotally, this response can differ between individuals. Next, we sought to more generally evaluate the utility of this system for studying the effects of genetic variation and biomechanical stress on gene regulation. Specifically, we considered dynamic expression quantitative trait loci (dynamic eQTLs), which are genetic variants associated with a change in gene expression in response to a treatment. For our system to be useful in identifying dynamic eQTLs, individual differences should drive a substantial

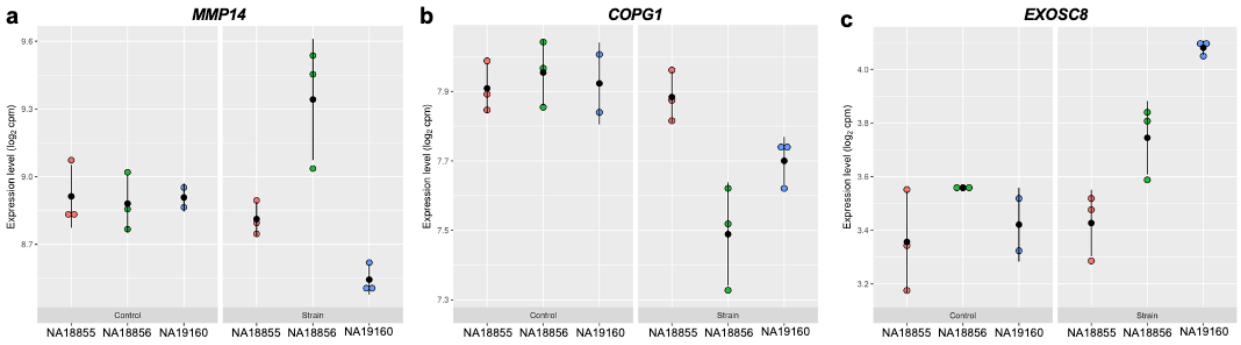


Figure 2.4: Examples of inter-individual differences in gene expression responses to cyclic tensile strain (CTS).

(a-c) *MMP14, COPG1,* and *EXOSC8* each demonstrate inter-individual differences in gene expression in our dataset. Dot plots are the expression level (log2 cpm) of each gene in each sample, with each individual and treatment condition plotted separately. Lines represent +/- one standard deviation. For each candidate gene, expression is relatively consistent between individuals in the control condition but differs in magnitude or direction between individuals in the strain condition (in response to CTS). These heterogenous responses do not all involve differing magnitudes of changes in the same direction. In the case of *MMP14,* while NA18855 does not seem to respond to CTS through altering *MMP14* expression, NA18856 responds through upregulation of this gene and NA19160 responds through downregulation. If heterogenous responses to CTS such as these exist more broadly between individuals and are associated with genotypic differences, they should be identifiable in population-level eQTL studies using this experimental system.

amount of the variation in gene expression response to the treatment. To quantify the contribution of individual differences to gene expression variation in iPSC-chondrocytes, we estimated how much of this variation is attributable to technical and biological factors. Our study design allows us to do so, as we collected bulk RNA-seq data from three independent technical replicates of each cell line.

First, we evaluated the contributions of experimental variables to major axes of variation in our bulk RNA-seq data by performing a principal components analysis (PCA; Supplementary Figures 2.7 and 2.8). Our results indicate the primary source of gene expression variation is individual (regression of PC1 by individual, $p = 1.34 \times 10\text{-}7$, regression of PC2 by individual, $p = 4.61 \times 10\text{-}4$). The second largest source of variation in the data is sex (regression of PC2 by sex, $p = 2.67 \times 10\text{-}4$), which is unsurprising given our study included one female and two male cell lines. Although treatment shows a minor correlation with PC2 ($R2 = 0.14$), PC3 ($R2 = 0.14$), and PC4 ($R2 = 0.3$), none of these correlations are statistically significant.

Encouragingly, we did not find technical replicate (or 'batch') to be significantly associated with any of the first five PCs in the data. Nevertheless, we took advantage of our replicated experimental design to account for two factors of unwanted technical variation in the data [121]; see Methods). Following this we observed the top three sources of gene expression variation are individual (regression of PC1 by individual, $p = 7.34 \times 10\text{-}8$, regression of PC2 by individual, $p = 3.98 \times 10\text{-}4$), sex (regression of PC2 by sex, $p = 1.79\text{x } 10\text{-}4$), and treatment effect (regression of PC3 by treatment, $p = 3.92 \times 10\text{-}2$), all three of which are significant (Figure 2.5a-b, Supplementary Figure 2.9).

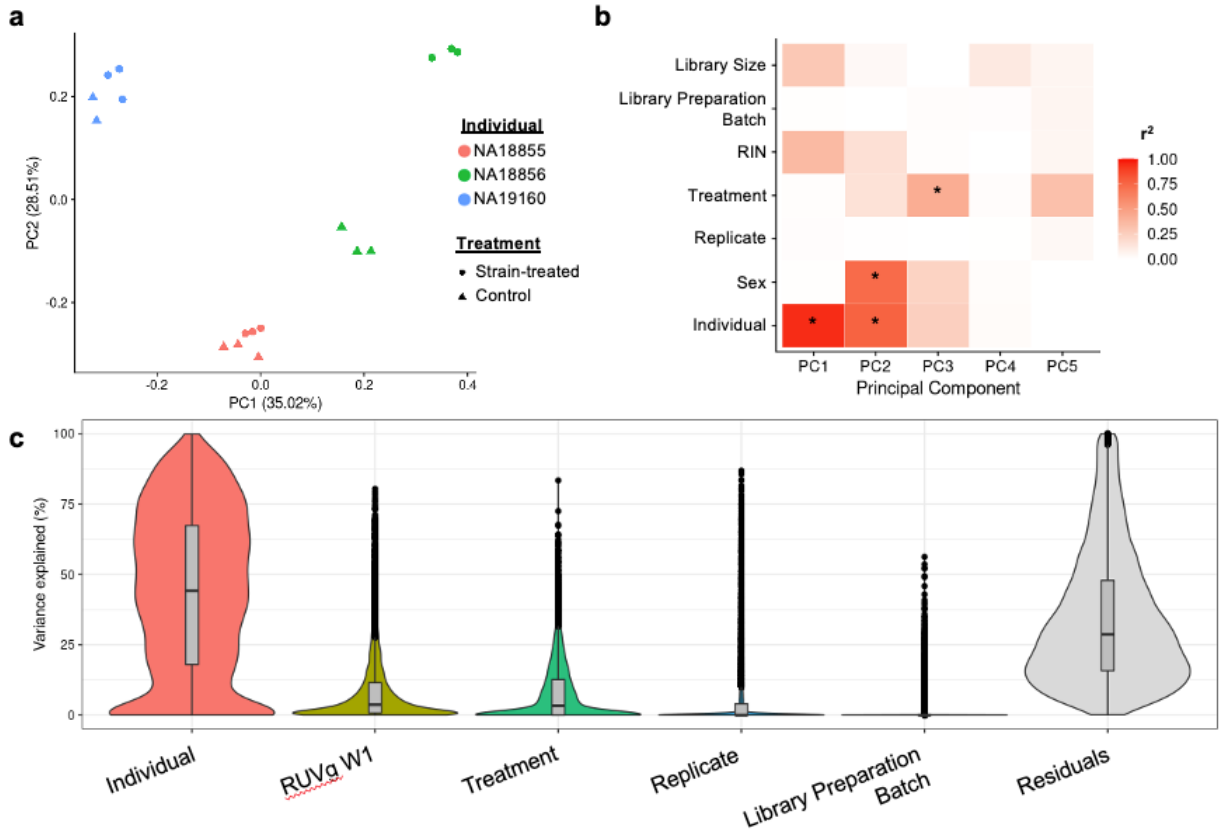


Figure 2.5: Characterization of sources of variation in bulk RNA sequencing data.

(a) Principal components (PC) plot of normalized and RUVs-corrected bulk RNA sequencing samples colored by individual and shaped by treatment condition. Samples largely separate by Individual, with strain-treated samples from NA18856 showing a large separation from control samples from the same Individual. **(b)** Correlation between each of the first 5 PCs and several experimental variables, determined through linear regression analysis on normalized and RUVs-corrected bulk RNA sequencing data. Significant regressions (Benjamini-Hochberg corrected FDR < 0.05) are highlighted with an asterisk. **(c)** Violin-box plots displaying the fraction of variation explained by a number of experimental variables of the study design, including a single factor of unwanted variation fit using RUVg. Variables are ordered from largest to smallest by the median fraction of variation explained except for Residuals. The boxplots indicate the median, inner quartile range (IQR) and 1.5 times the IQR. Data beyond this are plotted as points. Violin plots indicate the density of data points based on their width.

Next, we took a more systematic approach to modeling the contribution of biological and technical factors to gene expression variation. Our goal was to leverage the total amount of variation in our data rather than focusing only on a few major axes of variation, as in the PCA above. We quantified the contributions of several experimental variables to gene expression variation on the level of individual genes using a linear mixed model (see Methods). To do so, we modeled a single factor of unwanted variation in the data by using a set of 100 genes with the least amount of variation in the data as negative control genes [121] (see Methods). A single factor of unwanted variation was chosen as it was the minimum number of factors required to obtain correlations between gene expression residuals of technical replicates that were higher than those between samples from different individuals or treatments. We then included the filtered and normalized gene expression data and this single factor of unwanted variation in the model (see Methods; Figure 2.5c).

We determined that individual cell line contributes the largest amount of variance to the data (median of 42% variance explained). The additional factor of unwanted variation explains a median of 3.6% of the variance, and treatment explains a median of 3.5% of the variance. In contrast, technical replicate batch and cDNA library preparation batch explain a negligible amount of variance (median of 8.7x10-7 % and 3.5x10-7 % variance explained, respectively). We observed similar results when running a model that did not include the factor of unwanted variation (Supplementary Figure 2.10). Therefore, the biological variables of individual cell line and treatment contribute more to gene expression variation than technical variables. However, unwanted variation still seems to contribute to gene expression variation. Therefore, gene expression studies using this system should account for potential latent sources of variation.

2.3.6 A power analysis

Our results are encouraging for our goal of verifying the feasibility of using iPSC-chondrocytes to study gene-by-environment interactions in joint health. One possible way to study these interactions would be to use this system to map static eQTLs and dynamic response eQTLs (i.e., eQTLs that emerge in response to CTS). We conducted a power analysis to determine the potential impact of expanding this system to include 58 iPSC lines (Figure 2.6; we chose n=58 because this is the number of YRI iPSCs available to us). Under the assumptions of a simple linear regression to map eQTLs and a conservative Bonferroni correction for multiple testing (FWER = 0.05; see Methods), we estimated that a sample of 58 individuals will provide 80% power to detect eQTLs with a standardized effect size of 0.7 in each of the control and treatment conditions. This standardized effect size corresponds to a heritability of (0.7^2 / (0.7^2 +1) = 0.33). At this effect size, the power to detect eQTLs comes with a correspondingly low FPR of calling a dynamic eQTL (0.22).

To contextualize these results, we reanalyzed eQTL summary statistics from a set of previous dynamic eQTL studies that come from a variety of different research contexts [102,122,123] (Figure 2.6). In each of these studies, hundreds to thousands of genes in each treatment condition have at least one eQTL which meets the standardized effect size threshold of 0.7 above. While none of these examples perfectly recapitulates the results of our system and while the estimated effect sizes required to meet the threshold are fairly large, the fact these estimates are conservative and come from eQTL studies in three different stimulus conditions demonstrates the potential effectiveness of this approach.

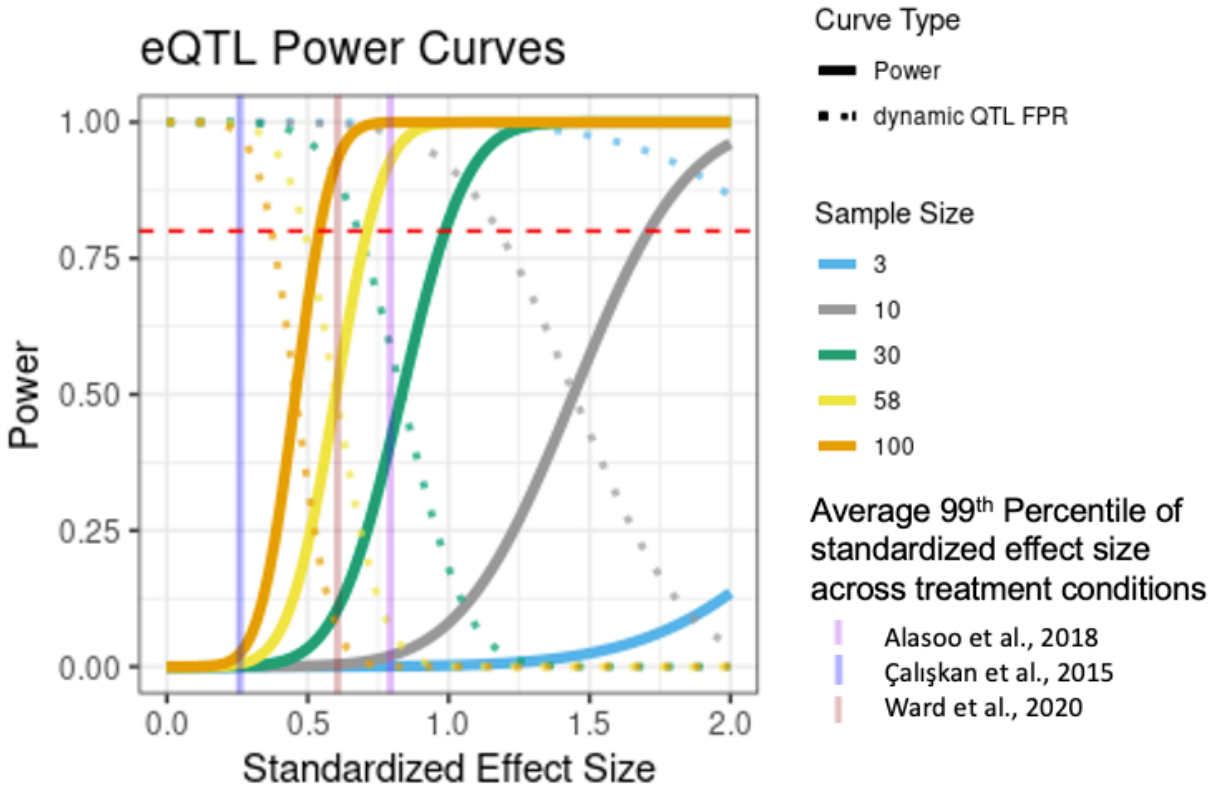


Figure 2.6: Power analysis for eQTL and dynamic eQTL study with two conditions.

Power curves are derived under the assumption of a simple linear regression for expression Quantitative Trait Locus (eQTL) mapping and plotted over standardized effect sizes (effect sizes divided by the phenotype standard deviation) for a range of sample sizes. Dynamic QTL false positive rates are computed as the probability of a SNP being called as significant in only one of two treatment conditions, assuming the standardized effect size was in fact identical in both conditions. The horizontal red line represents a power to detect eQTLs of 0.80. Vertical transparent lines represent the 99th percentile of the standardized effect size estimated from an empirical cumulative distribution function fit to eQTL summary statistics from 3 published dynamic eQTL studies from different contexts, with the mean value over all the conditions in each study plotted.

2.4 Discussion

We conducted a study to establish the feasibility of an *in vitro* iPSC-chondrocyte model for studying gene-by-environment interactions in joint health. Gene-by-environment interactions, particularly those related to biomechanical stress, may play a role in pathogenesis in joint related diseases such as OA. However, numerous ethical and logistical obstacles limit the study of these interactions and their effects on gene regulation in human chondrocytes. iPSC-chondrocytes may be a suitable alternative to circumvent these obstacles when paired with *in vitro* CTS models. Overall, our *in vitro* system allows for both the precise control of iPSC-chondrocyte environmental exposures and measurement of gene expression responses relevant to human joint health. While no *in vitro* model can completely accurately mimic *in vivo* disease, our results demonstrate this system has tremendous potential to increase our understanding of human joint health.

iPSC-chondrocytes are a valuable system to address the current lack of gene expression studies in human joint cells. Although iPSC-chondrocytes do not completely emulate mature, primary human chondrocytes, they do exhibit protein and gene expression patterns characteristic of both adult chondrocytes and developing chondrocytes. The relatively early differentiation stage of our cells may be due to a variety of factors, including a shorter differentiation time and the culturing of cells as a monolayer as opposed to a 3-dimensional pellet [124]. Nevertheless, iPSC-chondrocytes provide a unique opportunity to learn about gene regulation in human joints and the basis of adult joint disease phenotypes. For instance, the ability to generate cells along the trajectory between iPSC-MSCs and mature chondrocytes allows for gene expression studies at a level infeasible with human primary tissues. Furthermore, prior studies have shown that

studying iPSC-derived cells can uncover potentially important and transient forms of gene regulation masked in terminal cell types [20].

Our results point to a robust gene expression response to CTS in iPSC-chondrocytes. We detected 987 DE genes in our study between treated and control cultures. These DE genes are enriched for gene sets relevant to joint health and OA. Thus, our results highlight the potential of this system as a platform for gene expression studies of human joint cells that circumvent the limitations of primary tissues. Our observations also suggest that studying gene regulation in this relatively simple system may provide insight into more complex biological processes relevant to human joint disease.

One potential cause for reservation in our GWAS analysis is the mismatch in population ancestries between our DE results (African ancestry) and OA GWAS results (European ancestry). However, prior studies have found that genetic associations between causal variants and complex traits are largely shared between populations [125–127]. Furthermore, our analyses focus on the genes implicated in the OA GWAS results, rather than the causal variants themselves. As such, we expect the relevance of these gene sets to carry over more faithfully between populations. Finally, the population mismatch would likely have the effect of biasing our results towards the null rather than introducing false positive findings. Therefore, we believe the observed enrichment is meaningful despite the current lack of equivalent OA GWAS results from a more comparable African ancestry population.

We acknowledge that *in vitro* CTS models do not directly mimic the compressive biomechanical stress felt by joint chondrocytes *in vivo*. However, CTS models are recognized as a valid method for studying the effects of extra-physiological stresses in cultured cells. Others have previously used CTS models to measure responses of joint cells to controlled

biomechanical stress treatments [62–65,68–71,128]. Other groups have also developed models that use specific types and patterns of biomechanical strain to induce transcriptional and biochemical changes characteristic of early human OA [68–71]. Our RT-PCR and bulk RNA-seq results further attest to the utility of CTS as a model of biomechanical stress in studies of human joint health.

We also found that an *in vitro* iPSC-chondrocyte system may be useful for studying the effects of genetic variation on gene regulation; moreover, it offers a way to study how biomechanical stress interacts with genetic factors to affect gene regulation. Indeed, individual-level differences drive a substantial amount of gene expression variation in this system. Therefore, eQTL and dynamic eQTL studies would be feasible using iPSC-chondrocytes. We identified specific differences in the gene expression response to CTS between individuals in this study. As such, iPSC-chondrocytes may be fruitful for uncovering gene-by-environment interactions involved in pathogenesis of joint diseases.

Investigating dynamic and context-specific gene regulatory effects may reveal the mechanisms contributing to joint disease development and progression, as this approach has been successfully applied to a variety of other cell types and trait contexts [19,21,102–105,122,129]. Previous studies have found that dynamic eQTLs are more enriched for relevant significant GWAS alleles than non-dynamic ('standard') eQTLs, which show consistent effects between conditions [20,21,103,105]. Our power analysis suggests that a study with a few dozen individuals may grant sufficient power to detect many static and dynamic eQTLs. Dynamic eQTLs may be more useful for identifying candidate susceptibility genes in joint diseases than steady state eQTLs, and they may also improve our understanding of gene-by-environment interactions related to joint health and disease.

Future studies using the iPSC-chondrocyte system should account for the possibility that transcriptional heterogeneity between and within individual iPSC-chondrocyte lines may confound association results in an eQTL study. Our analysis of the scRNA-seq data from control iPSC-chondrocytes suggests that differentiation efficiency does not differ substantially between individuals. Nonetheless, it is possible that differentiation efficiency may differ for other individuals not included in this study. There may also still exist transcriptional heterogeneity between iPSC-chondrocytes in their response to CTS that bulk RNA-seq would not adequately capture. Measuring and accounting for transcriptional heterogeneity in iPSC-chondrocytes will also allow future gene expression studies to focus specifically on iPSC-chondrocytes, which represent only a minority of cells in each culture. This will increase power to detect both standard and dynamic eQTLs.

Our iPSC-chondrocyte system also facilitates investigations beyond those involving only human cells. The existence of panels of human and nonhuman primate iPSCs [130] introduces the possibility of inter-species comparisons of response to CTS. Comparative studies may help uncover gene-by-environment interactions that contribute to the differential prevalence of OA and other joint diseases observed across primate species [131–133].

We believe the *in vitro* iPSC-chondrocyte CTS model shows great promise when applied to studies of gene expression in human joints. We hope such a system enables future studies of gene regulation in joint cells and their connections to joint health and disease.

2.5 Methods

2.5.1 Chondrogenic differentiation

iPSCs in this study were previously generated from lymphoblastoid cell lines (LCLs) derived from three Yoruba individuals (NA18855, female; NA18856, male; and NA19160; male) [96]. These LCLs were originally derived from individuals collected as part of the HapMap project [134]. Undifferentiated iPSCs were cultured on Matrigel-coated plates (Corning 356230) in Essential 8 (E8) medium at 37°C, 5% CO2, and atmospheric O2 until iPSCs reached 30% confluency. E8 medium was subsequently changed to mesenchymal stem cell (MSC) culture medium, which consists of low glucose Dulbecco's Modified Eagle Medium (DMEM) with 20% stem cell-qualified fetal bovine serum (FBS), and 100 mg/mL Penicillin/Streptomycin. The MSC medium was changed every day for 3 days, at which point cells were 80-100% confluent. On day 3, cells were detached from the Matrigel-coated petri dishes using a 0.05% Trypsin/EDTA solution and cultured on uncoated polystyrene flasks in MSC medium. The medium was changed every 2-3 days until the cells reached 90% confluency. The cells were then sub-cultured at a ratio of 1:3 until passage 4, at which point cells were classified as iPSC-derived MSCs (iPSC-MSCs). iPSC-MSCs were cryopreserved with cryopreservation media (80% FBS, 10% MSC culture medium, 10% Dimethyl Sulfoxide) in liquid nitrogen at passage 5 to 7.

iPSC-MSCs were detached from culture flasks using 0.05% Trypsin/EDTA and seeded at a density of 250,000 cells/well onto the center of wells of BioFlex Type I Collagen coated 6-well Culture Plates (FlexCell International BF-3001C) using BioFlex cell seeders (FlexCell International). Cells were seeded using a regimen of 15% elongation for 2 hours followed by overnight culture in MSC culture medium. After seeding, cells were cultured in serum-free chondrogenic differentiation medium [92], consisting of high glucose DMEM, 100 mg/mL Penicillin/Streptomycin, 50mg/mL L-Proline, 200mM GlutaMax, 50mg/mL L-Ascorbic acid-2-

Phosphate, 11g/L Sodium pyruvate, 5mM Dexamethasone, 1x ITS Premix, and supplemented with 10 ng/mL TGF-β3. The chondrogenic medium was changed every 2-3 days for 14 days.

2.5.2 Standard phenotyping of iPSC-derived cells

Flow cytometry of iPSC-MSCs was performed using the BD Biosciences Human MSC Analysis Kit (BD Biosciences 562245), in combination with the Zombie Violet Fixable Viability Kit (BioLegend 423113). The Human MSC Analysis Kit assesses the surface markers CD90, CD105, CD73, CD34, CD45, CD11b or CD14, CD19, CD79α, and HLA-DR. In each flow experiment, matched iPSCs from the same line as each iPSC-MSC were included as a negative staining control. Samples were run on a BD LSRII Special Order System machine at the University of Chicago Cytometry and Antibody Technology Core Facility.

iPSC-chondrocytes were fixed using 4% paraformaldehyde in phosphate-buffered saline (PBS) before staining using Alcian blue and Nuclear Fast Red. Alcian blue binds proteoglycans, which are found in connective tissue, particularly in cartilage [135]. Stained iPSC-chondrocytes and matched iPSC-MSCs from the same individuals were imaged using an Olympus dissecting microscope.

2.5.3 Immunostaining for COL2A1

iPSC-chondrocytes differentiated in chondrogenic media from MSCs for 14 days in either monolayer or pellet culture, MSCs, and primary cartilage tissue were fixed using 4% paraformaldehyde in PBS. iPSC-chondrocyte pellets were generated as in Nejadnik *et al*., (2015) [92], fixed in 4% paraformaldehyde in PBS, dehydrated sequentially in 15% sucrose in PBS and 30% sucrose in PBS, and then embedded in optimal cutting temperature compound (OCT).

Primary human articular cartilage samples were obtained from a patient undergoing hip replacement surgery (University of Chicago BSD/UCMC IRB Protocol 19-0990). Under sterile conditions, cartilage scrapings were obtained from the medial portion of the femoral head and cut into small pieces using a scalpel. Samples were washed with PBS twice and flash frozen. Primary cartilage tissues were then thawed and embedded in OCT. Cartilage tissue and iPSC-chondrocyte pellets were sectioned on a cryostat to a thickness of 18 μm and 5 μm respectively and sections were mounted on slides prior to staining. Cells and sections were immunostained using a rabbit Collagen II polyclonal primary antibody (Thermo PA5-85108, RRID:AB_2792256) and a secondary antibody HRP/DAB detection IHC kit (Abcam ab64261, RRID:AB_2810213). Immunostained cells and sections were imaged using an EVOS microscope under the brightfield setting.

2.5.4 Cyclic tensile strain regimen

iPSC-chondrocytes were treated with a cyclic tensile strain (CTS) regimen that is known to induce an OA-like phenotype using the Flexercell FX6000 Tension System (Flexcell International) [68–71]. Plates were loaded onto the Flexercell baseplate (located in an incubator at 37°C, 5% CO2, and atmospheric O2), and a vacuum was used to deform the cell culture plate membrane and create uniform biaxial cyclic tensile strain. Specifically, 2.5% elongation (15kPa) of CTS was applied to the cells at a rate of 0.5 Hz for 24 hours.

2.5.5 Quantitative real-time reverse transcription PCR (RT-PCR) of chondrocyte hypertrophy-related marker genes

RNA was extracted from iPSC-chondrocytes following CTS or control treatments using the ZR-Duet™ DNA/RNA MiniPrep kit according to manufacturer instructions (Zymo D7001). To quantify target gene expression of GUSB, MMP1, MMP13, and TIMP2, we used RT-PCR with a QuantStudio 6 Flex Real-Time PCR System and SYBR Green reagents according to manufacturer instructions (Applied Biosystems). The cycle threshold (Ct) values were measured, and relative transcript levels were calculated for each target gene in each sample. The efficiency (E) of each PCR amplification reaction was calculated based on the slope of a linear curve of a series of dilutions of target DNA with known concentrations (E =10^(-1/slope)). Data were plotted as relative expression, which is calculated as E^(− ΔΔCt), using GUSB as a housekeeping gene in all cases [136].

2.5.6 Droplet-based single cell RNA sequencing

iPSC-chondrocytes were dissociated from adherent conditions into single cell suspension as follows: first, cells were rinsed once with 1X PBS. Then, 1mg/mL of collagenase II in 1X HBSS was added to cell culture wells at room temperature for 5 minutes. The collagenase II was neutralized with MSC medium and removed before further processing of the cells. Cells were rinsed once again with 1X PBS. A 0.25% Trypsin/EDTA solution was added to wells at room temperature for 2 minutes until cells detached. The trypsin was neutralized with MSC culture medium, and the cells were pelleted at 1000 rpm for 5 minutes and resuspended in FBS Stain Buffer. Cells were counted separately for each sample and combined in equal proportions before loading into a Chromium Single Cell A Chip kit according to manufacturer instructions (10X Genomics, 120236). To ensure that collection batch, individual, and treatment conditions were not confounded, samples were pooled strategically. One GEM well of a Chromium single cell

chip targeting a collection of 5000 cells contained NA19160 control, NA18856 control, and NA18855 strain-treatment cells. A second GEM well of the same Chromium single cell chip targeting a collection of 5000 cells contained NA19160 strain-treatment and NA18855 control cells. The cells collected from sample NA18856 strain-treatment were not processed due to viability issues. Single cell cDNA libraries were established following the 10x Genomics Chromium Single Cell protocol [137]. In brief, the RNA of the captured cells was released by lysis, barcoded via a reverse transcription process, and amplified to produce gene expression libraries. The libraries were sequenced to 100 base pairs, paired-end on one lane using the Illumina HiSeq4000 at the University of Chicago Genomics Core Facility according to manufacturer instructions.

2.5.7 Single cell data processing

FastQC (RRID:SCR_014583) was used to confirm that the reads were of high quality. Using an in-house computational pipeline, we extracted 10X cell barcodes and UMIs from raw scRNA-seq reads and mapped remaining reads to genes in the hg38 genome using STARsolo from the STAR software with default parameters (version 2.6.1b, RRID: RRID:SCR_004463) [138]. The software demuxlet was used to deconvolute sample identity of individual cell droplets and detect multiplets in multiplexed samples with default parameters [139]. Previously collected and imputed genotype data for the three Yoruba individuals from the HapMap and 1000 Genomes Project were used as input for demuxlet [134,140].

Processed gene count per cell barcode matrices were imported into R using the Seurat package (v3.2.0, RRID:SCR_007322) [141,142]. Data were filtered to remove cells with fewer than 2000 UMIs detected and more than 10% of reads mapping to mitochondrial genes. Cells

assigned as multiplets by demuxlet were also removed. A Uniform Manifold Approximation and Projection (UMAP) plot of the merged and unintegrated data shows that cells originating from the same individual cluster with each other (Supplementary Figure 2.11).

2.5.8 Integration of individual level scRNA-seq data and characterization of cell clusters

Filtered scRNA-seq data was integrated across individuals using Seurat. Cells that were assigned as singlets by demuxlet were treated as individual datasets. Specifically, we focused on just those datasets deriving from control cell culture conditions (n=3), as opposed to strain-treated conditions (n=2). Using Seurat, the SCTransform normalization function was applied to each of these datasets, and then datasets were integrated using integration anchors identified using the FindIntegrationAnchors function. Five-thousand features were selected as integration features for the SCT integration.

Seurat's FindClusters function was used with 38 gene expression principal components (PCs) and a resolution of 0.4 as parameters to perform unsupervised clustering of transformed and integrated data. Thirty-eight gene expression PCs were chosen by locating the elbow in an elbow plot of PCs. To characterize the resulting three clusters that emerged, a Poisson adaptive shrinkage model was fit to the raw count data from the cells in each pseudo-sample described above using the ashR package [143]. Poisson ashR models were fit separately for cell droplets assigned to each unsupervised cluster or separately for cell droplets from each individual. The cumulative density function of the inferred prior distributions for each of the fitted Poisson ashR models was plotted as in Sarkar and Stephens 2020 [144], for chondrogenic gene markers.

2.5.9 Topic modeling of single cell RNA sequencing data

An unsupervised topic model with k=7 topics was fit to the scRNA-seq raw count data from several published sources and data from iPSCs and iPSC-derived cell types generated by our laboratory. Briefly, single cell data from iPSCs, iPSC-MSCs, iPSC-chondrocytes, and iPSC-osteoblasts collected by our group from a single human cell line were combined with single cell data from primary human hepatocytes [145], iPSC-chondrocytes from an iPSC-chondrogenic pellet time-course [146], primary human chondrocytes [147,148], and the iPSC-chondrocytes from the present study.

First, the 10X Genomics Chromium Single Cell Gene Expression platform was used to collect scRNA-seq data from iPSCs, iPSC-MSCs, iPSC-chondrocytes, and iPSC-osteoblasts. All of these cells, which were derived from a single human cell line, were included in the topic modeling analysis. This human iPSC line was previously generated and characterized in our lab [130]. The same protocol used to generate other iPSC-MSCs in the current study was also used to generate iPSC-MSCs here, with the exception that DMEM:F-12 (Thermo fisher 11330032) was used instead of low glucose DMEM (see Methods). The same chondrogenic media formulation was used to differentiate iPSC-chondrocytes here as in this current study. iPSC-osteoblasts were generated by culturing iPSC-MSCs in osteogenic differentiation medium, consisting of high glucose DMEM (Gibco 11965092), 100 mg/mL Penicillin/Streptomycin (Corning 30002Cl), 10% stem cell-qualified fetal bovine serum (FBS, Thermo fisher 10567014), 50ug/mL Vitamin C, 100nM Dexamethasone, 10mM β-glycerophosphate, and 1uM Vitamin D. The osteogenic medium was changed every 2-3 days. Our iPSC-chondrocyte and iPSC-osteoblast protocols each included a total of 21 days of differentiation in their respective media before isolation and data collection, compared to 14 days for iPSC-chondrocytes in the current study.

Also included in the analysis were single-cell data from 3,490 hepatocytes published in MacParland et al., 2018, which were subset from a larger dataset of single cell results from whole liver homogenate [145]. Data from MacParland *et al*., 2018 are accessible using the R package HumanLiver and were originally obtained using the 10X Genomics Chromium Single Cell Gene Expression platform. These cells belong to clusters 1, 3, 5, 6, 14, and 15, identified in the original paper as showing enriched ALB (Albumin) expression, a hallmark of hepatocytes.

Additionally, data from iPSC-derived chondrocytes from a time-course of iPSC-chondrocyte pellet differentiation published in Wu *et al*., 2021 were obtained from GEO (GEO SRP290799) [146]. Data from single cells collected on day 7, day 14, day 28, and day 42 of differentiation were used to fit the topic model. Wu *et al*. chondrogenic pellets were treated with C59 for WNT inhibition during chondrogenesis to improve homogeneity of hiPSC chondrogenesis and avoid off-target cells.

Finally, data from 6,200 and 1,464 primary human chondrocytes were obtained from Chou *et al*., 2020 and Ji *et al*., 2018, respectively [147,148], for use in the topic modeling analysis. Cells from Chou *et al*. 2020 were isolated from the intact outer lateral tibial plateau of a single male individual and processed using the 10X Genomics Chromium Single Cell Gene Expression platform. Data from these cells were downloaded from GEO (GEO Sample GSM4626766). Cells from Ji *et al*., 2018 were obtained from 10 patients with OA undergoing knee arthroplasty and underwent a modified single cell tagged reverse transcription (STRT) protocol for single cell transcriptional data collection. Data from all cells included in the original study were used. In both primary chondrocyte studies, isolated chondrocytes were not cultured *in vitro* before processing for scRNA-seq.

Genes with non-zero counts in at least one cell in any of the six single cell datasets were included in the raw count matrix used to fit the topic model. A Poisson non-negative matrix factorization (NMF) model with 7 ranks was fit to the data using the fit_poisson_nmf function in the fastTopics R package with default parameters (v0.4.35) [149]. After fitting the Poisson NMF model, the fitted loadings and factors matrices were rescaled to sum to a total of 1 across each barcode for the loadings matrix and across each gene for the factors matrix to convert the Poisson NMF model into a topic model. The rescaled loadings matrix became the topic probabilities, and the rescaled factors matrix became the word probabilities in the resulting topic model.

The diff_counts_analysis function in fastTopics was applied to the topic model to evaluate differential expression of individual genes in each topic. Briefly, the function calculates a β statistic, which represents the log-fold change in relative occurrence of a gene in a single topic compared to its occurrence in all other topics. The function also calculates a standard error and z-score for each β statistic based on a Laplace approximation to the likelihood at the MLE.

2.5.10 Bulk RNA extraction and sequencing

RNA was extracted from cells following CTS or control treatments using the ZR-Duet™ DNA/RNA MiniPrep kit (Zymo D7001). RNA concentration and quality were measured using the Agilent 2100 Bioanalyzer. Library preparation was performed over two batches using the Illumina TruSeq RNA Sample Preparation Kit v2 (RS-122-2001 & -2002, Illumina). Samples were sequenced to 50 base pairs, single-end on one lane using the Illumina HiSeq4000 at the University of Chicago Genomics Core Facility according to manufacturer instructions. A minimum of 17,284,094 raw reads were generated per sample. We used FastQC to confirm that

the reads were of high quality. One bulk RNA-seq sample was found to have a very low proportion of mapped reads (38.40%) and was excluded from subsequent analyses.

2.5.11 Quantifying the number of bulk RNA-seq reads mapping to genes

Reads were mapped to the hg38 genome using STAR (version 2.6.1b) [138]. Gene expression levels were quantified using the featureCounts function in Subread (v1.6.5 RRID:SCR_009803) using standard parameters [150]. All downstream processing and analysis steps were performed in R (v3.6.1, RRID:SCR_001905) unless otherwise stated.

2.5.12 Transformation and normalization of bulk RNA-seq reads

Log2-transformed counts per million (CPM) were calculated from raw counts for each sample using the edgeR package (RRID:SCR_012802) [151]. Lowly expressed genes were filtered such that only genes with an expression level of log2(CPM) > 2.5 in at least 4 samples were kept for downstream analyses. For the remaining 10,486 genes, the raw read counts were normalized using the relative log expression (RLE) method to account for the median number of reads sequenced across samples.

2.5.13 Removing unwanted variation from bulk RNA-seq data

To account for batch effects arising between technical replicates before differential expression analysis, we modeled factors of unwanted variation using the RUVs correction method (RRID:SCR_006263) [121] with k=2. RUVs is a method that uses technical replicate samples to estimate factors of unwanted variation from RNA-seq data. Individual-treatment pairs were constant within replicate blocks, which are used for the RUVs correction. RUVg is distinct

from RUVs in that it uses negative control genes to estimate factors of unwanted variation from RNA-seq data rather than knowledge of technical replicate samples in the data.

2.5.14 Differential expression analysis with bulk RNA-seq data

Differential expression (DE) was measured using a linear-model-based empirical Bayes method in the limma R package (RRID:SCR_010943). The voom function from the limma package was also used to calculate weights to account for the mean-variance relationship in the RNA-seq count data.

Replicate batch was modeled as a random effect while treatment, individual, two RUVs coefficients, and RIN score were modeled as fixed effects in the linear mixed model for DE comparisons as in equation (1). The ashR package [143] was used to perform multiple testing correction on the DE tests using an adaptive shrinkage method. Genes with an FDR-adjusted p value < 0.05 were considered DE.

Equation 1:

$$Y \sim \beta_0 + \beta_1 * treatment + \beta_2 * individual + \beta_3 * sex + (1|replicate) + \beta_4 * RUV_{W1} + \beta_5 * RUV_{W2} + \beta_6 * RIN + \varepsilon$$

2.5.15 Enrichment of DE genes in biological pathways and OA-related gene sets

Using topGO (RRID:SCR_014798), we assessed enrichment of Gene Ontology (GO) biological processes among DE genes. A Kolmogorov-Smirnov test using ashR adjusted p values was used for assessing enrichment of GO processes, and the top 20 most enriched terms were reported. To test for enrichment of sets of OA-related genes in our DE genes [28,33], a Fisher's

exact test was used. In all enrichment tests, the background gene set was the complete set of genes tested for DE in our analyses (n = 10,486 genes).

There is the possibility that the enrichment of DE genes for GO categories and outside gene sets is driven by differential power due to some genes having higher expression in our samples. We therefore repeated the DE and enrichment analyses ten times, permuting the treatment condition labels for the samples each time.

2.5.16 Analysis of sources of variation in bulk RNA-seq data

Principal component analysis (PCA) was performed on the normalized log2(CPM) values from above. A linear regression analysis was then performed between each of the top 5 PCs and several biological and technical variables. These variables included number of reads sequenced, library preparation batch, RIN score, treatment condition, replicate, and individual. P values from the regression were corrected using the Benjamini Hochberg (BH) procedure. Results with a BH-adjusted p value < 0.05 were considered significant.

The variancePartition (RRID:SCR_019204) package was applied to the filtered and RLE-normalized CPM values [152]. variancePartition uses a linear mixed model to quantify the contribution of variance from different sources. Our linear mixed model included variation due to individual cell line, treatment status, replicate batch, and library preparation batch. In addition, a single coefficient of unwanted variation was determined using the RUVg correction method [121] with k=1; this coefficient was also included in the model. The RUVg correction method estimates factors of unwanted variation in RNA-seq data through negative control genes, which have the lowest variation in expression between samples. The 100 least variable genes in the data ranked by coefficient of variation were used as the set of control genes for the RUVg correction.

2.5.17 Power curves for expression QTL (eQTL) and dynamic eQTL mapping

To ascertain the power to detect eQTLs and dynamic eQTLs across a range of sample sizes and standardized effect sizes, we followed the example presented in Ward *et al*., 2021 [123]. In brief, for the power analysis we assumed a simple linear regression for eQTL mapping and a conservative Bonferroni correction for multiple testing (FWER=0.05). Standardized effect sizes are defined as the true additive effect size of genotype on gene expression divided by the phenotype standard deviation. To estimate the false positive rate of calling a dynamic eQTL, we computed the probability of a SNP being called as significant in only one of the two treatment conditions, assuming the standardized effect size was in fact identical in both conditions.

2.5.18 Reanalysis of previous dynamic eQTL studies

We used summary statistics from eQTL mapping in three prior dynamic eQTL studies [102,122,123] to determine standardized effect sizes for eQTL association tests in each treatment condition. Briefly, p values from association tests were converted into Z-scores using the appropriate quantile function. Z-scores were then converted to standardized effect sizes by adjusting for the square root of the sample size of the study. For summary statistics from Alasoo et al., 2018, and Caliskan *et al.*, 2015, an adaptive shrinkage model was fit to the distribution of effect sizes and standard errors using ashR [143]. The ashR posterior estimates of effect sizes and standard deviations were used to compute the standardized effect size. Standardized effect size thresholds for at least 0.8 power under a sample size of 10, 30, 58, or 100 individuals were determined as described above. The number of genes with at least one association test that meets each of these thresholds in each condition were tabulated. Empirical distribution functions were

fit to the distributions of the standardized effect sizes from each condition in each of the three studies. The 99th percentile of these standardized effect sizes was determined from the empirical distribution function.

2.5.19 Data and code availability

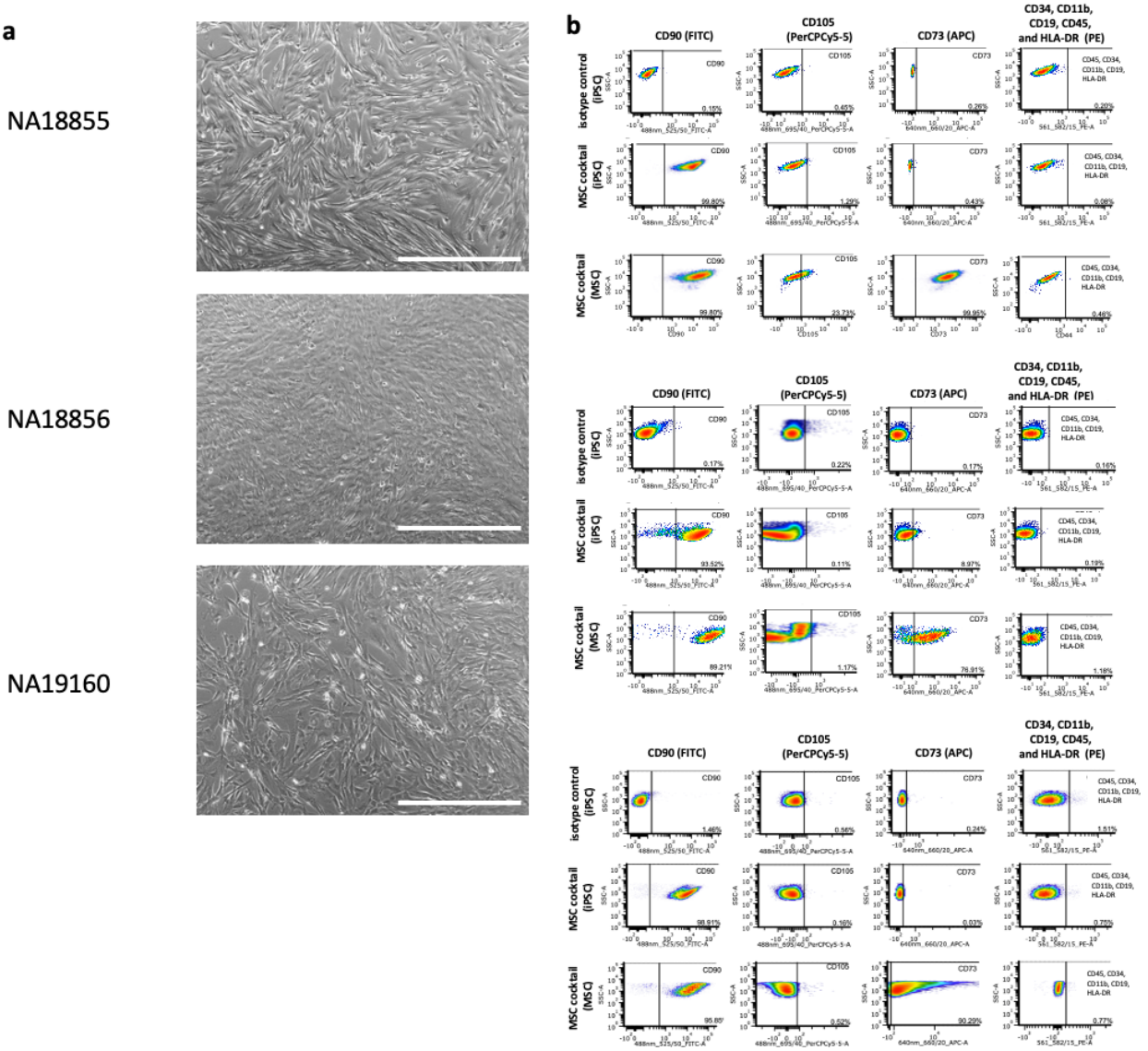
The data have been deposited in NCBI's gene expression omnibus and are accessible through: Characterizing gene expression responses to biomechanical strain in an *in vitro* model of osteoarthritis. Accession number: GSE165874; https://identifiers.org/geo:GSE165874 and Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model. Accession number: GSE167240; https://identifiers.org/geo:GSE167240. All computational scripts and analysis pipelines can be found on GitHub at: https://github.com/anthonyhung/invitrostrain_pilot_repository and in webpage format at: https://anthonyhung.github.io/invitrostrain_pilot_repository/index.html.

2.6 Acknowledgements

We thank Natalia Gonzales for comments on the manuscript. We thank Abhishek Sarkar, Michelle Ward, Kenneth Barr, and other members of the Gilad Lab for helpful discussions. This work was completed in part by using computational resources provided by the University of Chicago Research Computing Center. This work was supported by the US National Institutes of Health (R35GM131726 to Y.G., F30AG071412 to A.H., and F32AR075397 to G.H.). A.H. is supported by T32 GM007281 to the University of Chicago and an ARCS Foundation Scholar award from ARCS Illinois.

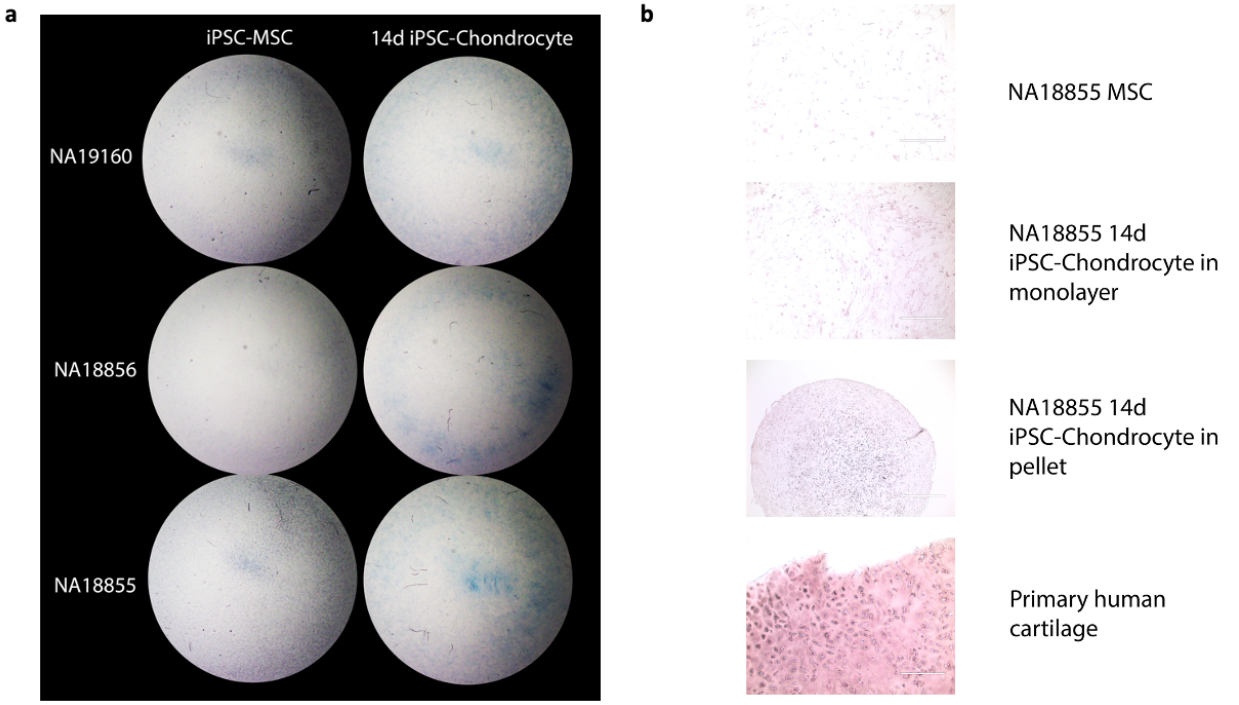
2.7 Supplementary Information

2.7.1 Supplementary Figures



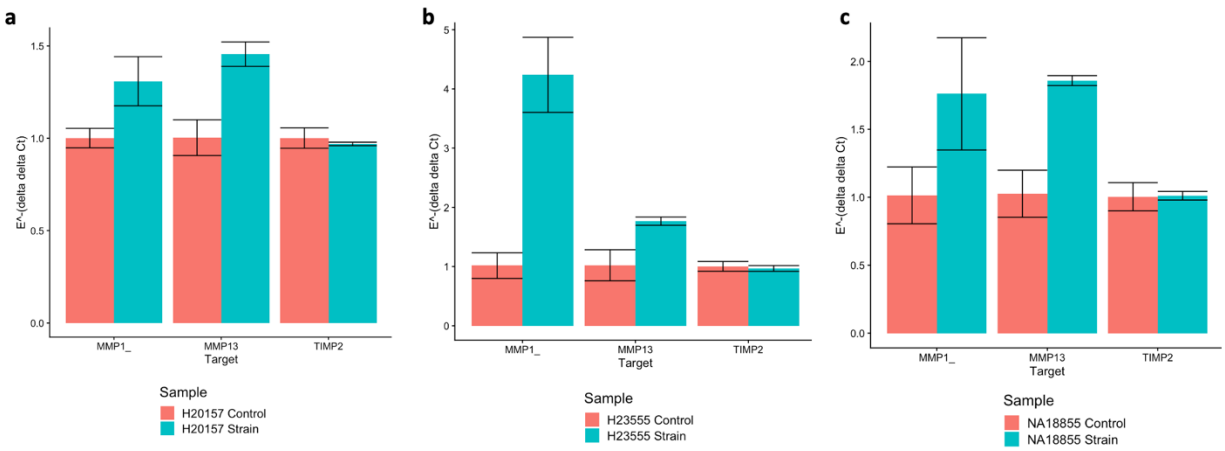
Supplementary Figure 2.1: iPSC-MSC characterization.

(a) EVOS microscope images of iPSC-derived MSCs from all three individuals demonstrate elongated morphology characteristic of MSCs. (Scale bar 250um) **(b)** Flow cytometry plots from labeling of iPSC-MSCs and matched iPSCs for cell-surface markers characteristic of MSCs. Compared to iPSCs, iPSC-MSCs demonstrate increased labeling of CD90 and CD73, while they do not show increased staining of several negative markers of MSCs (CD45, CD34, CD11b, CD19, and HLA-DR). iPSC-MSC staining of CD105, a third positive marker of MSCs, is variable across the three individuals.



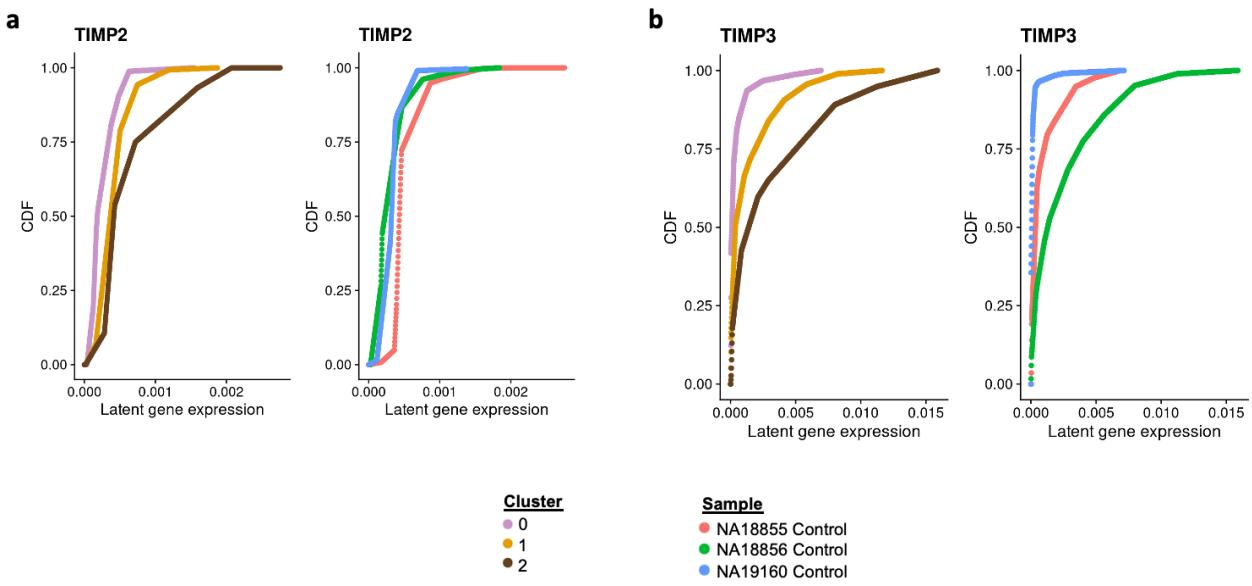
Supplementary Figure 2.2: Staining of 14-day iPSC-chondrocytes and matched iPSC-MSCs.

(a) Brightfield dissecting microscope images are cropped to show circular central seeded area of cells on BioFlex Type I Collagen coated 6-well Culture Plates (diameter 25mm). **(b)** EVOS microscope brightfield images of MSCs, 14-day iPSC-chondrocytes, and primary human cartilage stained for COL2A1 by immunocytochemistry/immunohistochemistry.



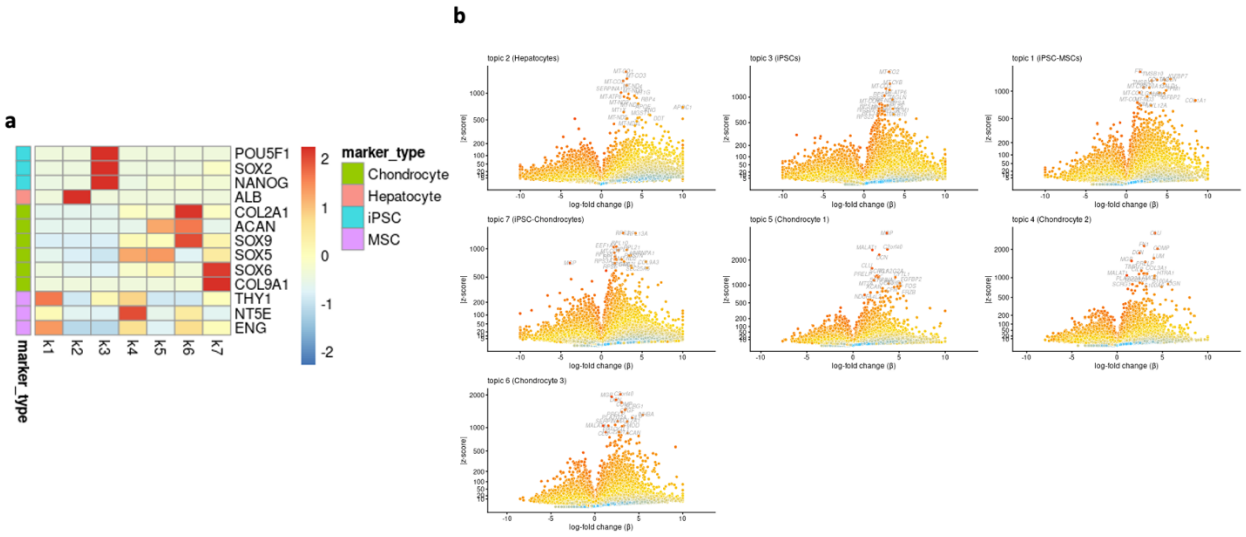
Supplementary Figure 2.3: Effect of cyclic tensile strain on the expression of chondrocyte hypertrophy markers in iPSC-chondrocytes.

Bar plots of relative expression of *MMP1, MMP13, and TIMP2* measured by RT-PCR in control and CTS-treated iPSC-chondrocytes from **(a)** H20157, **(b)** H23555, and **(c)** NA18855. H20157 and H23555 are differentiated from human fibroblast-derived iPS cell lines generated in Romero *et al*., 2015. Relative expression values are calculated via the delta-delta Ct method, using *GUSB* as a housekeeping gene in all cases and the efficiency (E) of each PCR amplification calculated based on the slope of a linear curve of a series of dilutions of target DNA. Error bars represent standard deviation from 3 technical replicates of the RT-PCR reaction.



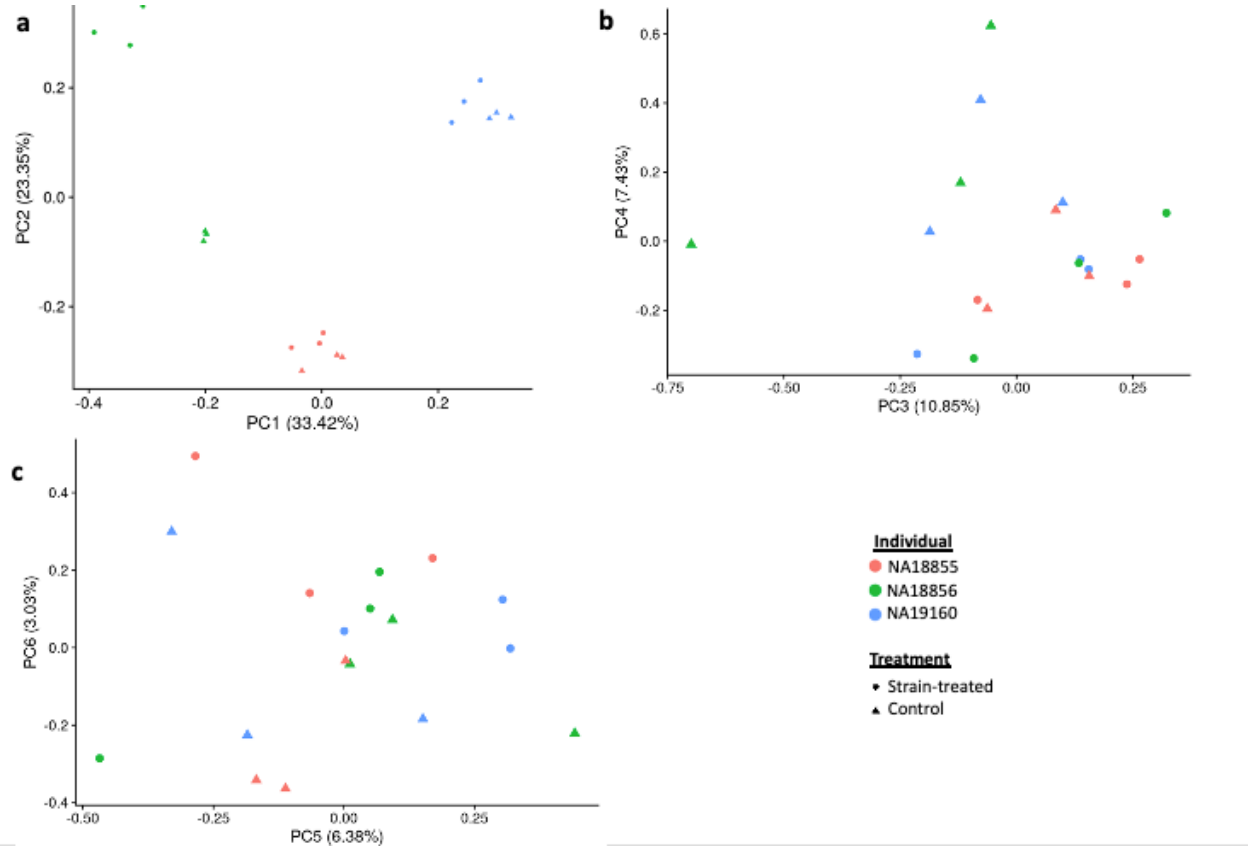
Supplementary Figure 2.4: Additional gene markers of single cell clusters.

Cumulative distribution function (CDF) of marginal distributions of latent gene expression of **(a)** *TIMP2* and **(b)** *TIMP3* determined through fitting a Poisson adaptive shrinkage model to raw gene expression counts in each sample. In each panel, (Left) CDF curves are colored by Seurat cluster. Cells in cluster 2 contain a higher latent gene expression of *TIMP2* and *TIMP3* on average. (Right) CDF curves colored by Individual.



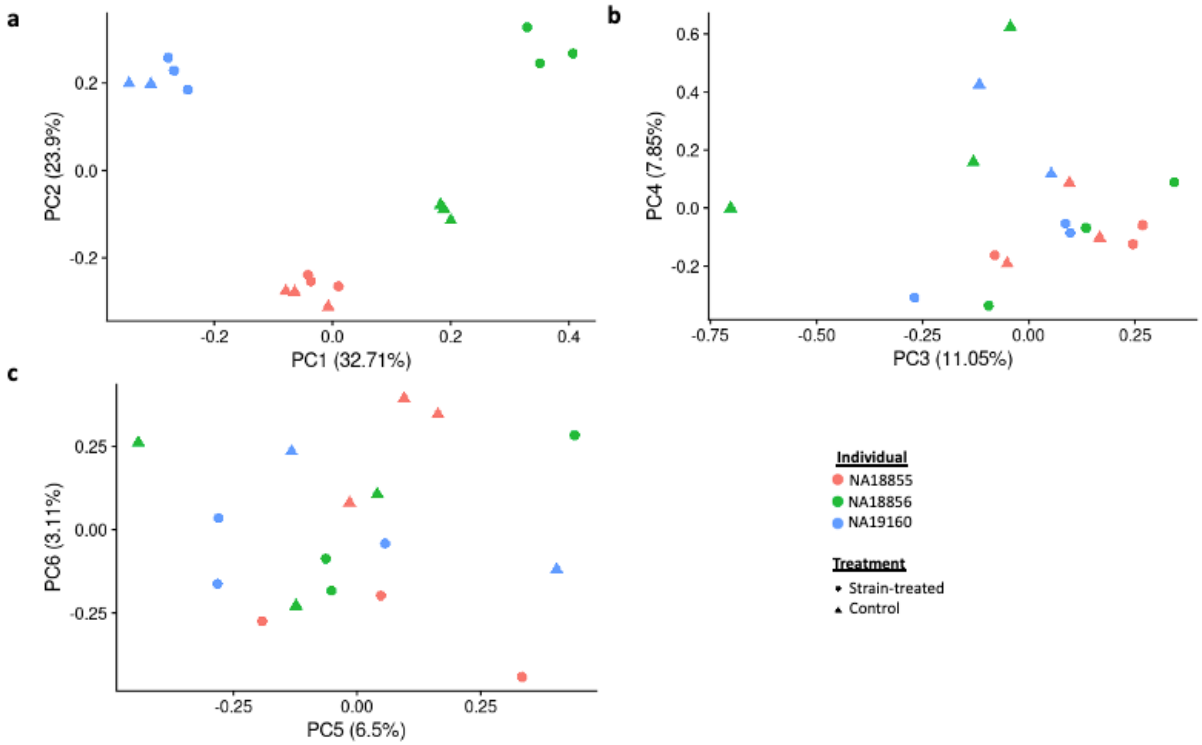
Supplementary Figure 2.5: Characterization of topics from topic model.

(a) Heatmap of word probabilities for selected marker genes for different cell types across a topic model containing seven topics. Heatmap cells are colored by word probabilities after applying a standard normal centering and scaling across each row. Colored bars on the left side of the heatmap denote the category of selected marker genes. Topic k2 demonstrates the highest word probability for hepatocyte marker gene *ALB*, while k3 demonstrates highest word probabilities for iPSC marker genes. Topic k1 demonstrates high word probabilities for MSC marker genes. Topics k4, k5, k6 and k7 demonstrate high word probabilities for several chondrocyte marker genes. **(b)** Volcano plots for log2 fold change relative occurrence (β) of individual genes in one topic compared to all other topics. |Z-scores| represent the absolute value of the z-scores for the β statistic. Positive β represent relative increase in occurrence of the gene in the topic compared to all other topics. Genes with |z-scores| above the 0.999 quantile are labeled in each plot.



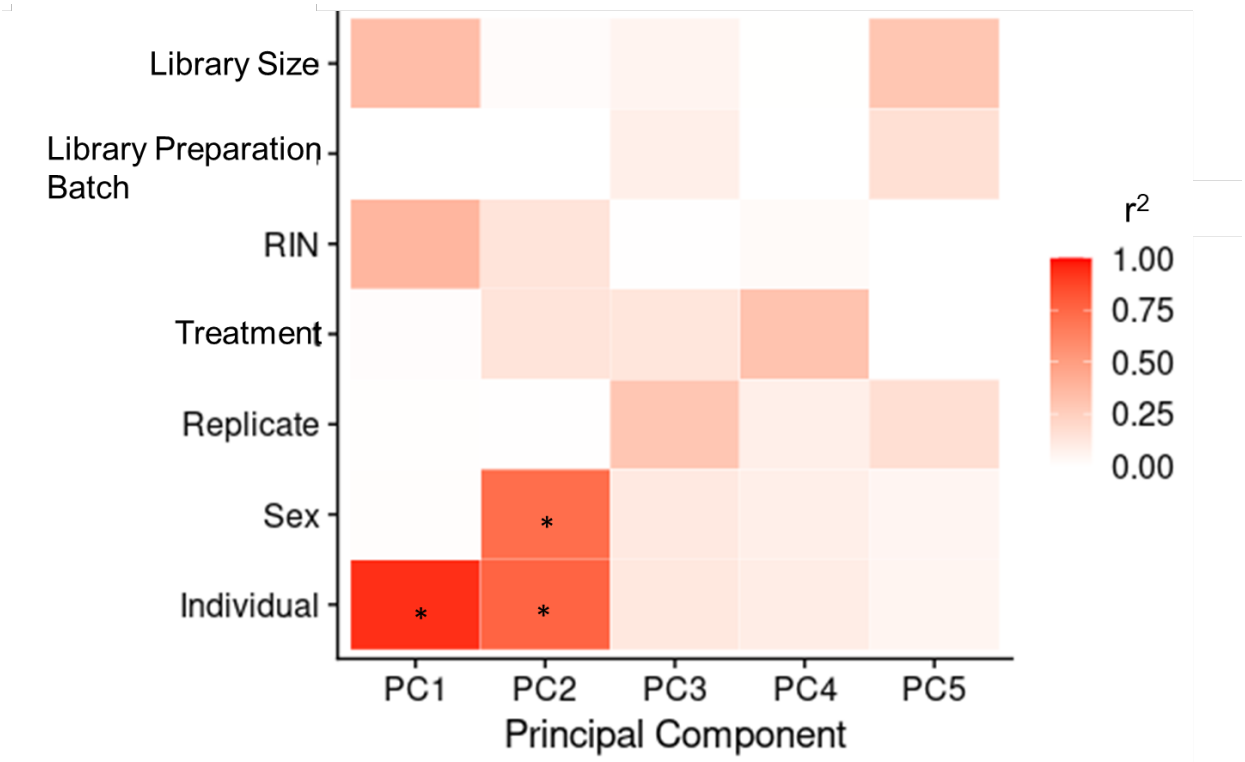
Supplementary Figure 2.6: PCA plots for normalized and filtered bulk RNA-seq data before RUVs correction including sample that failed QC.

(a-c) Principal components analysis (PCA) plots for principal components 1-6 are plotted for data including the single bulk RNA sequencing sample that was removed during quality control, corresponding to cells from individual NA19160 in the control condition from the first technical replicate of the experiment. When included in the PCA, this sample clusters as expected with corresponding samples from the other two technical replicates.



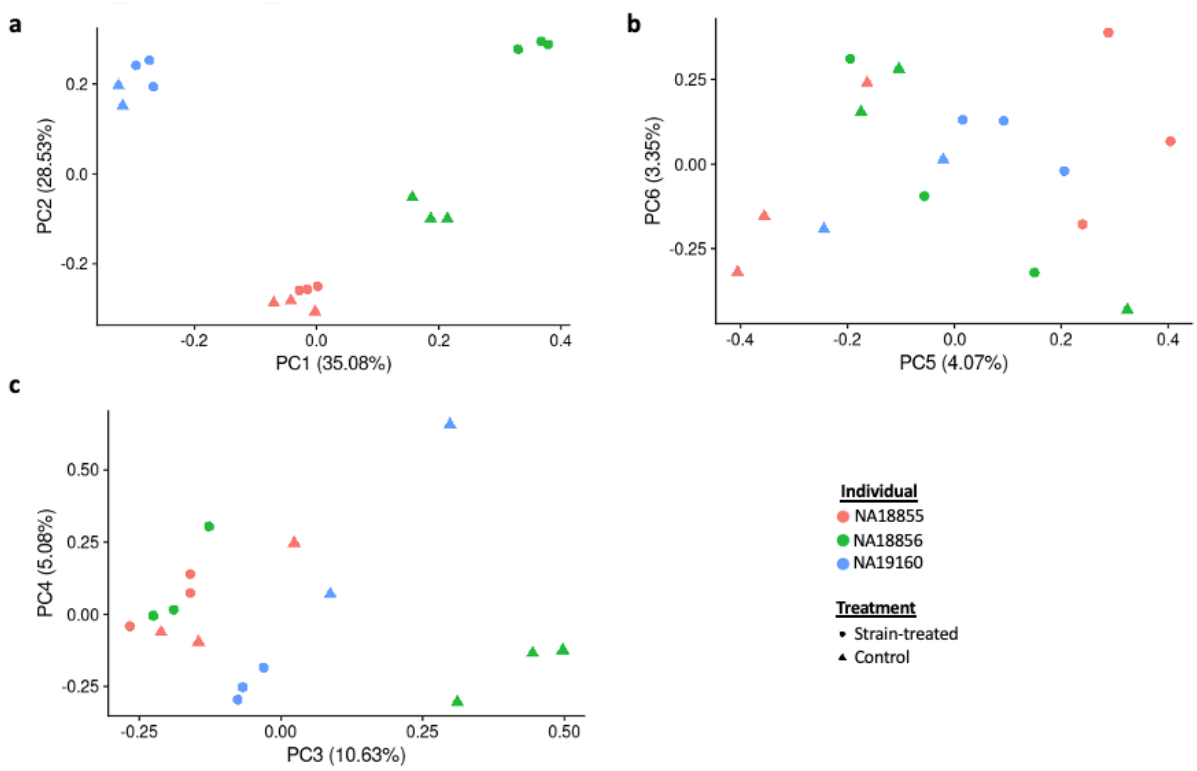
Supplementary Figure 2.7: PCA plots for normalized and filtered bulk RNA sequencing data before RUVs correction.

(a-c) PCA plots for principal components 1-6 are plotted for data absent RUVs correction.



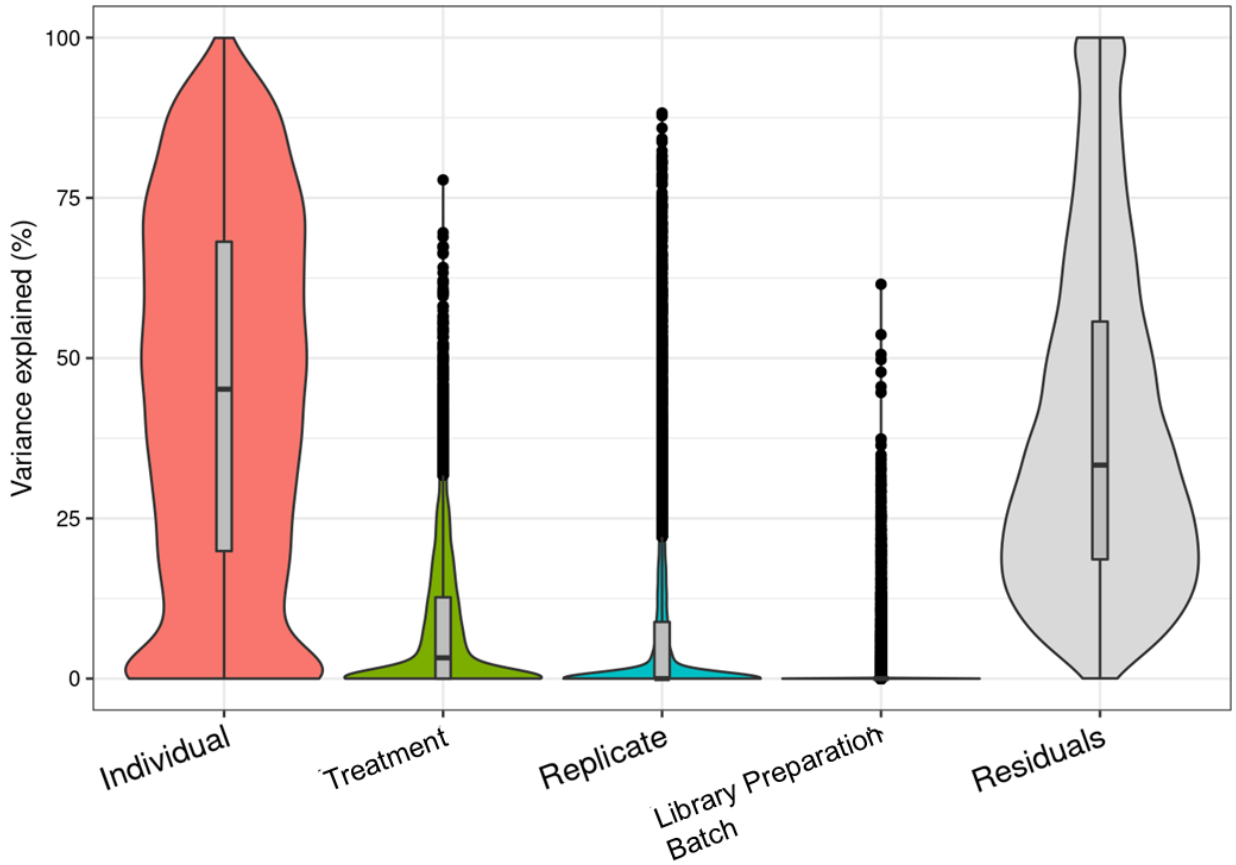
Supplementary Figure 2.8: Correlation between each of the first 5 PCs and several experimental variables with normalized and non RUVs-corrected bulk RNA sequencing data.

Significant regressions (Benjamini-Hochberg corrected FDR < 0.05) are highlighted with an asterisk.



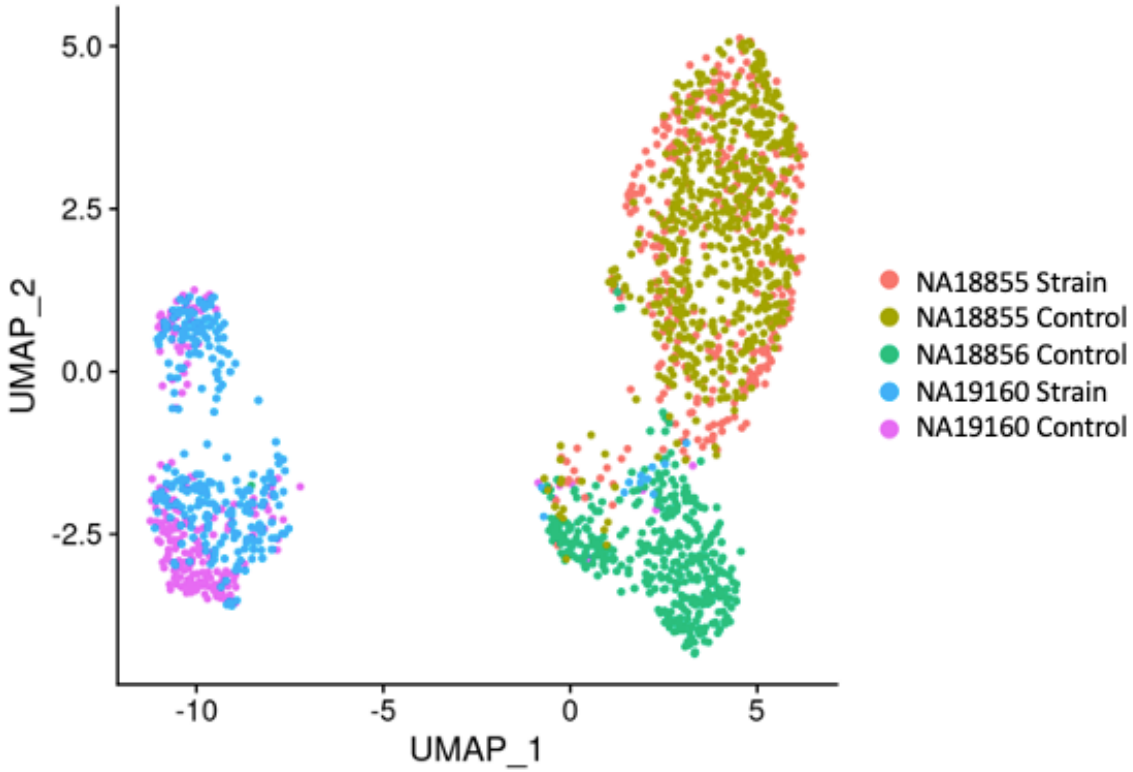
Supplementary Figure 2.9: PCA plots of additional PCs for normalized and filtered bulk RNA sequencing data after RUVs correction.

(a-c) PCA plots for PC3-6 for normalized and filtered bulk RNA sequencing data. PCA plot for PC1-2 is duplicated in Figure 2.3.



Supplementary Figure 2.10: Variance partition results on normalized and filtered bulk RNA sequencing data without the inclusion of RUVg factors of unwanted variation.

Individual and treatment both explain a larger proportion of variance of individual genes on average compared to technical factors of replicate and library preparation batch.



Supplementary Figure 2.11: Uniform Manifold Approximation and Projection (UMAP) plot of merged scRNA-seq data.

Raw scRNA-seq data from the two 10x GEM wells were merged and filtered to retain cells with fewer than 10% of reads coming from mitochondria and containing at least 2000 features. After log normalization, a UMAP plot was created for the data based on 2000 variable features and 50 gene expression principal components. No integration was applied to the data across the two GEM wells. Cells originating from the same individual cluster with each other.

CHAPTER 3

DETERMINING THE GENETIC BASIS OF GENE EXPRESSION RESPONSES TO TWO *IN VITRO* MODELS OF OSTEOARTHRITIS

3.1 Abstract

Osteoarthritis (OA), a painful degenerative disease of joint tissues that often leads to disability, results from the interplay between genetic and environmental factors. To study gene-by-environment interactions in OA and joint health, we established a panel of induced-pluripotent stem cell-derived chondrocytes and paired these iPSC-chondrocytes with OA-relevant *in vitro* inflammatory cytokine and mechanical stress treatments. We measured gene expression levels from iPSC-chondrocyte lines derived from 22 genotyped individuals following exposure to a control treatment and the two OA-relevant treatment conditions and characterized gene regulatory changes between treatments. We further identified 1070 genes with *cis* expression quantitative trait loci (eQTLs) across the three conditions as well as 257 genes with dynamic eQTLs, which have different effects across conditions. This study of an OA-relevant cell type exposed to two OA-relevant environments represents a step towards understanding the role of gene-by-environment interactions in joint health.

3.2 Introduction

Osteoarthritis (OA) is the most common form of arthritis, affecting over an estimated 32.5 million adults in the United States [153]. OA impacts all cells and tissues of the joint, leading to degeneration of cartilage, inflammation of the synovial capsule, and structural changes to subchondral bone [154]. Patterns of gene expression have been observed to differ between

healthy and OA cartilage, reflecting activation of biological pathways in this disease and potential gene dysregulation in diseased joint tissues [27,28]. Genetic variation plays a meaningful role in this disorder, with the heritability of OA estimated to be between 40-70% [29,30]. Genome-wide association studies (GWAS) of OA have linked over one hundred variants to OA, most of which are found in non-coding regions of the genome and thought to impact the regulation of genes [31,32].

Efforts to map regulatory quantitative trait loci, including gene expression quantitative trait loci (eQTLs), have sought to provide functional annotations for trait-associated variants as a way to connect non-coding variants to their associated phenotypes. Enrichment of gene expression, splicing, and chromatin accessibility QTLs has been found among a wide range of trait-associated GWAS variants [12,13], suggesting that these intermediate molecular phenotypes may contribute to complex traits. Nevertheless, completely bridging the gap between genotype and phenotype is difficult. Even the Genotype-Tissue Expression (GTEx) project, a large eQTL mapping effort which has identified tens of thousands of eQTLs across dozens of tissues and hundreds of donors in steady-state conditions, has only confirmed that 43% of GWAS disease-associated loci are also classified as eQTLs [17].

As gene regulation is highly dynamic and context specific, looking for eQTLs in specific cell types that have different environmental or developmental contexts may help to increase the number of detectable eQTLs and to close the trait-associated variant-eQTL gap. Non-steady state environmental contexts relevant to disease may also represent selective pressures that have shaped inter-population and inter-individual genetic variation underlying disease associations [19]. Fittingly, eQTLs associated with gene expression in certain developmental or environmental states but not others (often referred to as dynamic eQTLs) have the potential to

explain trait-associations better than static eQTLs identified in higher-powered studies [16,20,21]. Dynamic eQTL studies have been performed to great effect in various cell types and contexts, including drug-induced cardiotoxicity [21], cardiomyocyte differentiation [20], vitamin D exposure [101], and response to infection [19,102–105].

Despite these advances, similar efforts to examine dynamic eQTLs and other regulatory QTLs relevant for OA and joint health have been limited given the relative inaccessibility of joint cells and tissues from healthy donors. As an alternate, there are protocols to differentiate induced pluripotent stem cells (iPSCs) into cells relevant to joint health and disease, such as chondrocytes [92,100] which are the primary cells of cartilage. Additionally, cultured joint cells can be exposed to *in vitro* environmental treatments like inflammatory cytokine (e.g., IL-1β) exposures [77,78,83–87] and biomechanical strain [68–71] in order to induce phenotypes characteristic of human OA. Pairing such iPSC-derived chondrogenic differentiations with *in vitro* OA perturbations would allow researchers to circumvent many of the obstacles of primary tissue collection and to more thoroughly investigate of regulatory and dynamic QTLs related to OA.

Along this line, we designed a study to further explore the genetic determinants of gene expression responses to an OA-relevant cell type to two OA-relevant environments. Specifically, we collected bulk RNA-sequencing data from a panel of iPSCs-derived chondrocytes from 22 genotyped individuals exposed to a control treatment, an inflammatory cytokine treatment, and a biomechanical stress treatment. We used these data to study the gene expression responses of iPSC-chondrocytes to the two different conditions, including shared and condition-specific responses. We then leveraged the genetic diversity of the panel to better understand genetic associations with gene regulation by mapping eQTLs in each condition. We further investigated

gene-by-environment interactions on gene regulation by identifying dynamic eQTLs. Lastly, to better interpret the biology of both static and dynamic eQTLs, we overlapped our results with several external datasets pertaining to OA and joint health. We evaluated the enrichment of mapped eQTLs and eGenes for functional categories and heritability of GWAS related to OA and joint health. Overall, this study represents a step towards understanding the contributions of genetics, environment factors, and gene-by-environment interactions towards inter-individual risk for joint disease.

3.3 Results

We differentiated iPSC-MSC lines from 22 iPSCs derived from Yoruba YRI individuals from the HapMap project [134]. We then further differentiated iPSC-MSCs towards iPSC-chondrocytes over 14 days after seeding on BioFlex Type I Collagen coated 24-well Culture Plates (see Methods). We assigned iPSC-chondrocytes from each individual that were seeded from the same flask of iPSC-MSCs and differentiated concurrently to one of three treatment groups: an untreated control group, an IL-1β treated group, and a cyclic tensile strain treated group (see Methods). We sought to use this system to better understand the patterns of gene regulation and the nature of dynamic genetic control of gene expression across treatments in iPSC-chondrocytes. To this end, we collected and sequenced bulk total RNA from all three treatment groups for each individual simultaneously over each of seven batches, with all treatment groups for each individual included in the same batch (Figure 3.1a).

A principal components analysis and a pairwise correlation analysis of the bulk RNA sequencing data both demonstrate strong separation of samples by their individual of origin

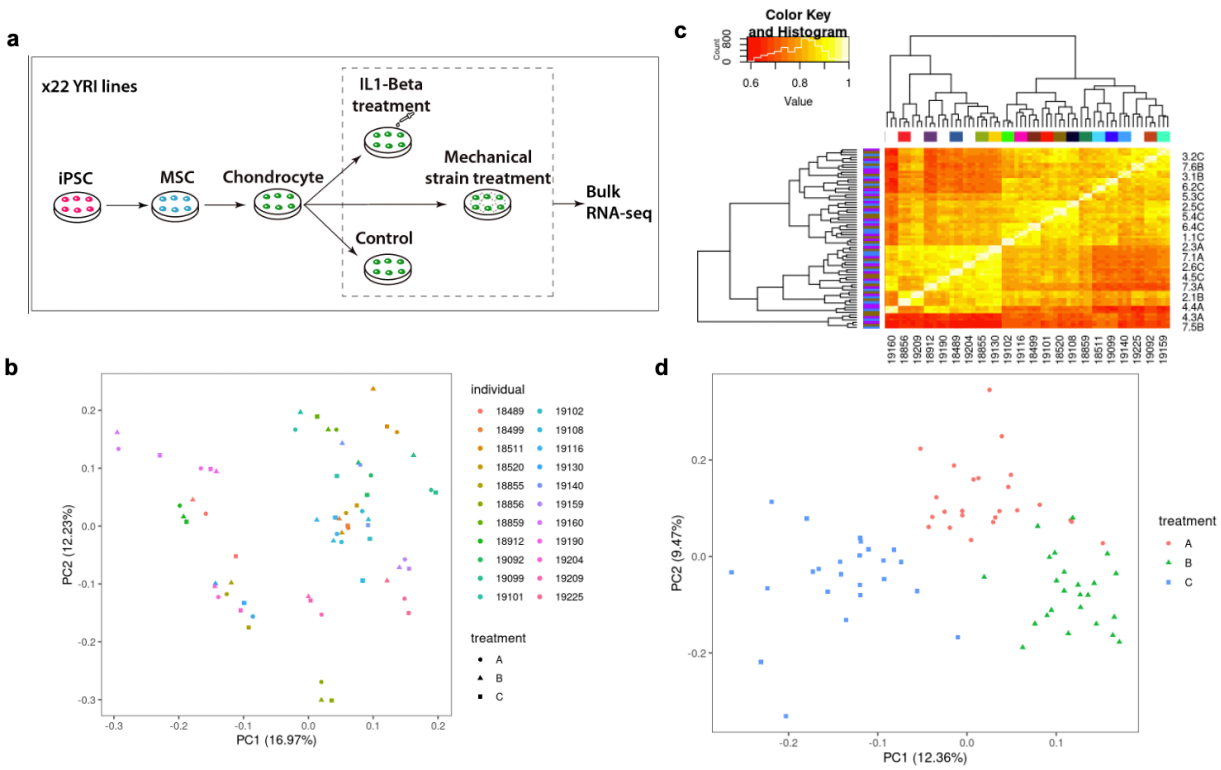


Figure 3.1: Study design and bulk RNA sequencing data.

(a) In our study design, iPSCs generated from 22 Yoruba individuals were first differentiated along the chondrogenic lineage with an intermediate differentiation step into mesenchymal stem cells (**MSCs**). iPSC-MSCs from each individual were differentiated into iPSC-Chondrocytes over a period of 14 days on flexible bottom cell culture plates. Subsequently, we treated separate iPSC-chondrocyte cultures from each individual with 3 treatments in the same cell culture incubator: 24 hours of no treatment, 24 hours of exposure to IL-1β cytokine, and 24 hours of exposure to a cyclic tensile strain treatment that is known to induce an OA-like phenotype. Following 24 hours of treatment, we extracted bulk RNA from all iPSC-chondrocyte replicates simultaneously. **(b)** Principal components analysis (PCA) plot of the first two principal components of normalized and filtered expression data collected across batches of the experiment. Each sample dot is shaped according to treatment (A = control, B = IL-1β, C = CTS) and colored by individual. **(c)** Pairwise Pearson correlations between normalized and filtered expression data from all samples reordered according to dendrogram correlation weights between samples. Columns are colored corresponding to individual, and rows are colored corresponding to treatment. **(d)** A PCA plot of the first two principal components of normalized expression data that has been corrected for the effects of individual, collection batch, RIN score, and sex through the use of a linear mixed model. Samples are colored and shaped according to treatment status (A = control, B = IL-1β, C = CTS).

(Figure 3.1b-c). Upon investigating the correlation of several experimental variables to top principal components of variation in the data, our results indicate the primary source of gene expression variation is individual (regression of PC1 by individual, R2 = 0.98, regression of PC2 by individual, R2 = 0.98, regression of PC3 by individual, R2 = 0.92). The second largest source of variation in the data is collection batch (regression of PC1 by collection batch, R2 = 0.21, regression of PC2 by collection batch, R2 = 0.52, regression of PC3 by collection batch, R2 = 0.49. RIN score also contributes to variation in the data (regression of PC3 by RIN, R2 = 0.31). Although treatment shows a minor correlation with PC3 (R2 = 0.05), this correlation is not statistically significant.

The strong contribution of individual, collection batch, and RIN score to the variation in the data demonstrates the need to account for these variables when performing downstream analyses seeking to examine differences between the variable of interest, treatment. To demonstrate the ability to robustly segregate samples from different treatment groups after accounting for these variables, we fit a linear mixed model to the gene expression data accounting for individual, collection batch, RIN score, and sex to the data (Equation 1). We then extracted the residuals from the fitted model, which represent gene expression data after accounting for the variables included in the linear mixed model. A principal components analysis of the corrected data demonstrates a clear separation of samples by treatment on the first two major axes of variation (Figure 3.1d).

3.3.1 Gene expression changes in response to *in vitro* OA treatments

To identify gene expression differences between treatments while accounting for confounding experimental variables, we fit a linear mixed model accounting for individual, collection batch, sex, and RIN variables to test for differential expression between the control samples and each of the two treatment groups (see Methods). At an FDR of 0.05, 2881 genes are significantly differentially expressed between control and IL-1β samples (Figure 3.2a-b). These differentially expressed genes are enriched for gene ontology biological processes corresponding to immune responses to cytokines and external stimuli. At an FDR of 0.05, 3449 genes are significantly differentially expressed between control and CTS samples (Figure 3.2c-d). These differentially expressed genes are enriched for gene ontology biological processes corresponding to connections between cells and between cells and the extracellular matrix, as well as extracellular matrix organization and metabolism.

3.3.2 Shared and condition specific responses to *in vitro* OA treatments reveal patterns of gene expression activated across different cellular environments

We next sought to evaluate the sharing of differential gene expression patterns across each of our two comparisons (control-IL-1β and control-CTS). To more accurately determine the sharing of effects between comparisons and to leverage the information that can be gained by sharing across the two comparisons, we fit a multivariate adaptative shrinkage (mash) model to the DE results (see Methods). Congruent with increased power, the number of differentially expressed genes in each comparison (defined as a mash posterior lfsr < 0.05) increased, with 3291 genes significantly DE in the control-IL-1β comparison and 4564 genes significantly DE in the control-CTS comparison. From the posterior estimates obtained after fitting the mash model to the data, we defined three groups of DE genes corresponding to effects unique to each of the

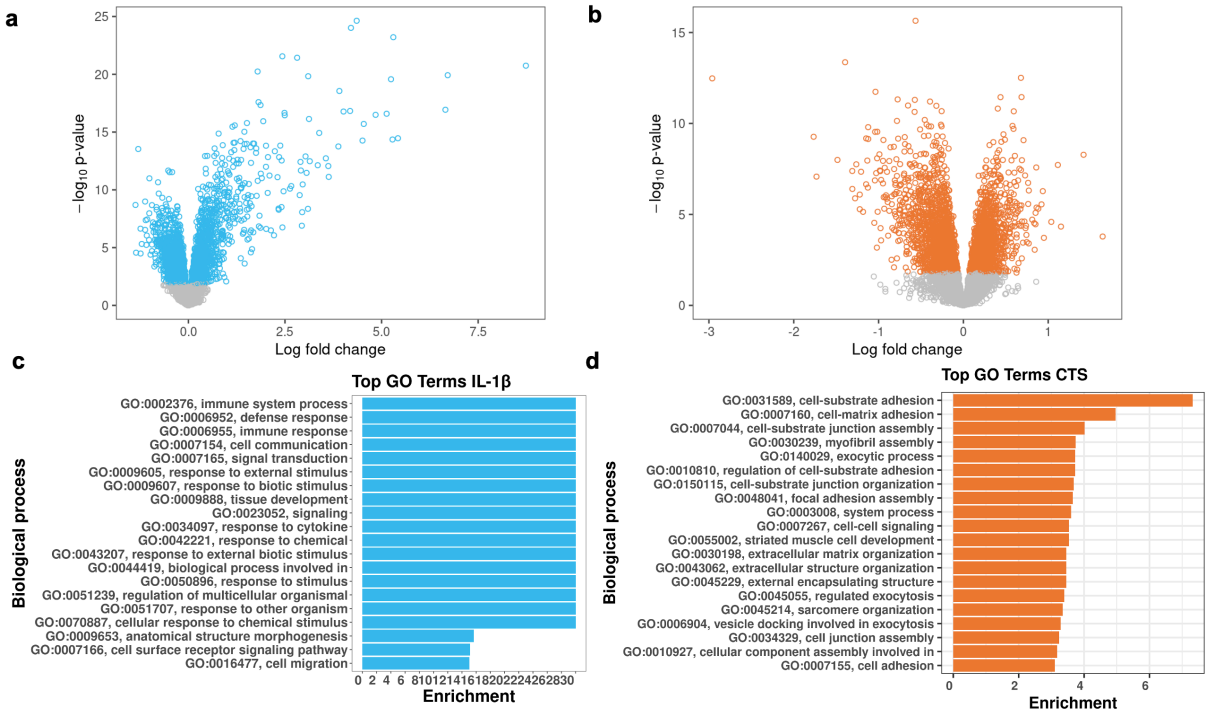


Figure 3.2: Differential expression results for IL-1β and CTS treatments.

(a and b) Volcano plot of –log10 raw p values vs log-fold change between control and IL-1β (a) and control and CTS (b) conditions**.** Highlighted in red are genes with an FDR-adjusted p value < 0.05. Genes plotted to the right of 0 on the x-axis represent genes with higher average expression in IL-1β or CTS-treated iPSC-chondrocyte samples compared to control samples. **(c-d)** Top 20 Biological processes enriched among DE genes compared to background set of 10,821 genes included in the analysis of control and IL-1β (c) and control and CTS (d) comparisons.

two comparisons ("IL-1β specific effects" and "CTS specific effects") and effects shared between the two treatments ("shared-effects").

We defined IL-1β specific effects as those genes significant at a local false sign rate (lfsr) < 0.05 in the IL-1β comparison and without a significant lfsr or an effect size within a factor of 2 difference in the CTS comparison. CTS specific effects were those genes with a significant lfsr in the CTS comparison and without a significant lfsr or a comparable effect observed in the IL-1β comparison within a factor of 2. Shared effects were those that had a significant lfsr in both conditions or were significant in one condition and had a comparable effect size within a factor of 2 in the other condition. Effects that were significant in both conditions also must have been in the same direction of effect to be classified as shared. Based on these definitions, we determined there were 1536 DE genes specific to the control-IL-1β comparison, 2485 DE genes specific to the control-CTS comparison, and 2234 DE genes shared between the two comparisons.

Shared DE genes are enriched for GO terms corresponding to extracellular matrix organization and cartilage and joint development (Table 3.1, Figure 3.3). While control-IL-1β specific DE genes are enriched for GO terms corresponding to immune and inflammatory response, bone development, and extracellular matrix disassembly, control-CTS specific DE genes are not enriched for any terms corresponding to extracellular matrix organization or cell-cell adhesions (Table 3.1, Figure 3.3). Furthermore, while shared and control-IL-1β specific DE genes are enriched for genes previously implicated in two differential gene expression studies between healthy and OA human cartilage [155] and damaged vs intact OA cartilage [156], control-CTS specific DE genes are not found to be enriched for either of these external DE gene sets (Figure 3.3). Thus, gene expression pathways activated specifically by CTS treatments in chondrogenic cells and that are not also activated by inflammatory cytokine treatment may not

Category of gene set	GO Term	Enrichment p value	Example genes
Shared	GO:0030198 extracellular matrix organization	0.00045	*COL11A1, COL14A1*
Shared	GO:0001501 skeletal system development	0.0041	*EXT1*
Shared	GO:0003417 growth plate cartilage development	0.00149	*COL27A1*
Shared	GO:0032330 regulation of chondrocyte differentiation	0.00362	*BMP4, BMP6*
Shared	GO:1905049 negative regulation of metallopeptidase activity	0.00765	*TIMP2, TIMP3*
IL1B specific	GO:0045087 innate immune response	4.10E-09	*ADAM15, CCL2, CCL20*
IL1B specific	GO:0006954 inflammatory response	3.90E-07	*STAT3, CXCL8*
IL1B specific	GO:0060348 bone development	0.00434	*SMAD5, BMPR2, TGFB3*
IL1B specific	GO:0071260 cellular response to mechanical stimulus	0.00469	*MAP3K2, PIEZO1, SOX9*
IL1B specific	GO:0022617 extracellular matrix disassembly	0.00223	*MMP1, MMP13, ADAMTS5*
CTS specific	GO:0045454 cell redox homeostasis	0.0016	*NQO1, PRDX1*
CTS specific	GO:0030262 apoptotic nuclear changes	0.00876	*BLCAP, ACIN1*

Table 3.1: Curated list of significantly enriched GO terms amongst shared and specific differentially expressed gene sets.

Shared, IL1B specific, CTS specific differentially expressed gene sets are defined based on the posterior estimates of local false sign rate and effect size from mash. GO terms were annotated as significantly enriched if the Fisher's exact test p value < 0.01. Gene expression plots of some example genes are located in Figure 3.3.

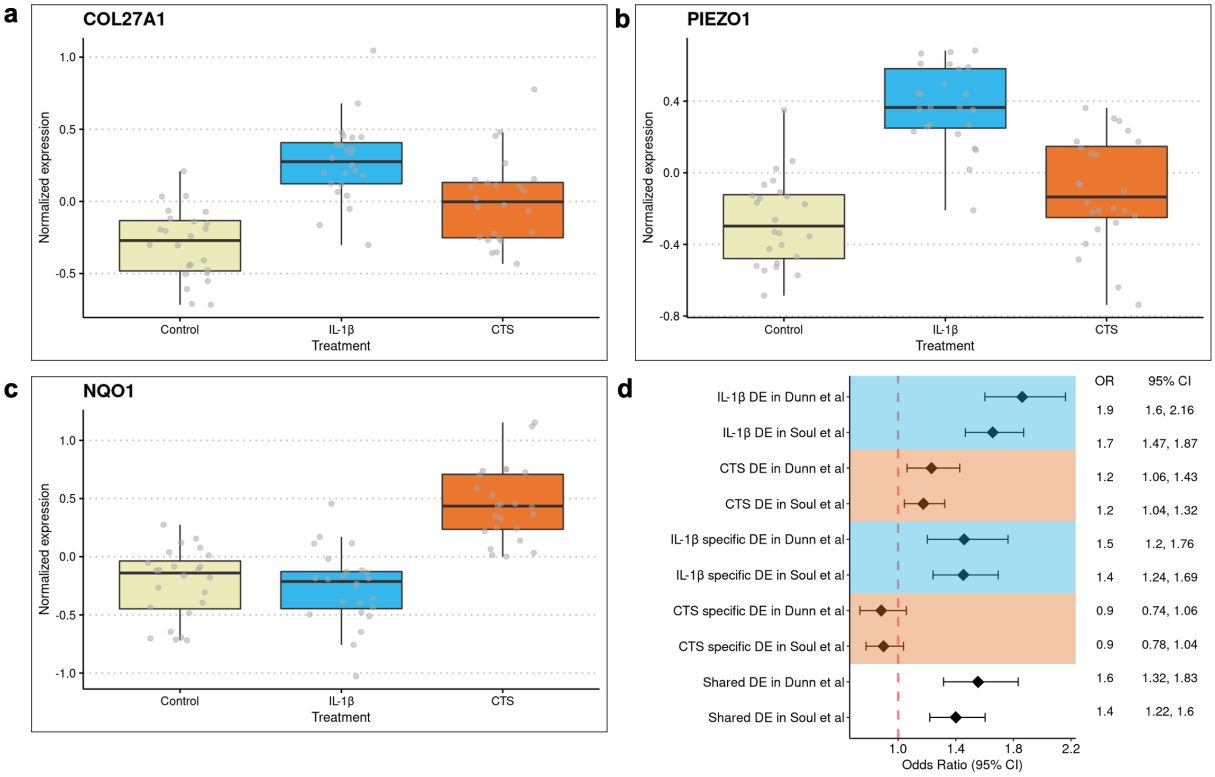


Figure 3.3: Shared and specific patterns of differential expression.

(a-c) Box plots of normalized expression accounting for several experimental variables by a linear mixed model fit using equation (1) to the data. In each plot, samples are separated by treatment group (A = control, B = IL-1β, C = CTS). **(a)** *COL27A1* is a shared differentially expressed gene across both control-IL-1β and control-CTS comparisons (both A-B and A-C comparisons are significant). **(b)** *PIEZO1* is an example of a control-IL-1β specific differentially expressed gene (only the A-B comparison is significant). **(c)** *NQO1* is an example of a control-CTS specific differentially expressed gene (only the A-C comparison is significant). **(d)** Forest plot of enrichments of two publicly-available osteoarthritis-related differential expression datasets amongst shared ("Shared DE"), control-IL-1β specific ("IL1B specific DE"), control-CTS specific ("CTS-specific DE"), control-IL-1β DE ("IL1B DE"), and control-CTS DE ("CTS DE") gene sets generated in this current study. Odds ratios and 95% confidence interval estimates of the odds ratios are plotted for each enrichment tests.

adequately reflect OA diseases processes. However, CTS treatments as a whole are still a relevant model for OA, as the set of genes found to be differentially expressed (not uniquely differentially expressed) in CTS vs control are enriched for both external DE gene sets (Figure 3.3).

3.3.3 Dynamic eQTLs are revealed following environmental perturbations

With our RNA sequencing data collected from genotyped individuals, we sought to identify eQTLs in each of our treatment conditions. Using the TReCASE framework, an approach that leverages total and allele-specific read count information for QTL mapping [157], followed by multiple testing correction, we identified 965 genes with an eQTL (eGenes) in at least one condition (control: 379 eGenes, IL-1β: 351 eGenes, CTS: 389 eGenes; FDR < 0.01).

We further sought to identify dynamic response eQTLs, which display condition-specific effects. Due to the sample size of our study, we expect incomplete power to detect eQTLs in any condition. Determining dynamic eQTLs by analyzing a simple overlap of the sets of eQTLs in each condition would therefore lead to an overestimation of the number of dynamic eQTLs, by underestimating the true amount of sharing of eQTLs between conditions. To better account for incomplete power in eQTL detection, we defined a shared eQTL as one which is significant in one of the three conditions (FDR < 0.01) and with a nominal p value < 0.0548 (see Methods) in another condition. Defining dynamic eQTLs as those either significant in control (FDR < 0.01) but not in any other conditions (p value > 0.0548), or as those significant in IL-1β and/or CTS (FDR <0.01) but not in control (p value > 0.0548), we identify 262 dynamic eQTLs corresponding to 257 unique genes (dynamic eGenes; Figure 3.4, Supplementary Figure 3.3).

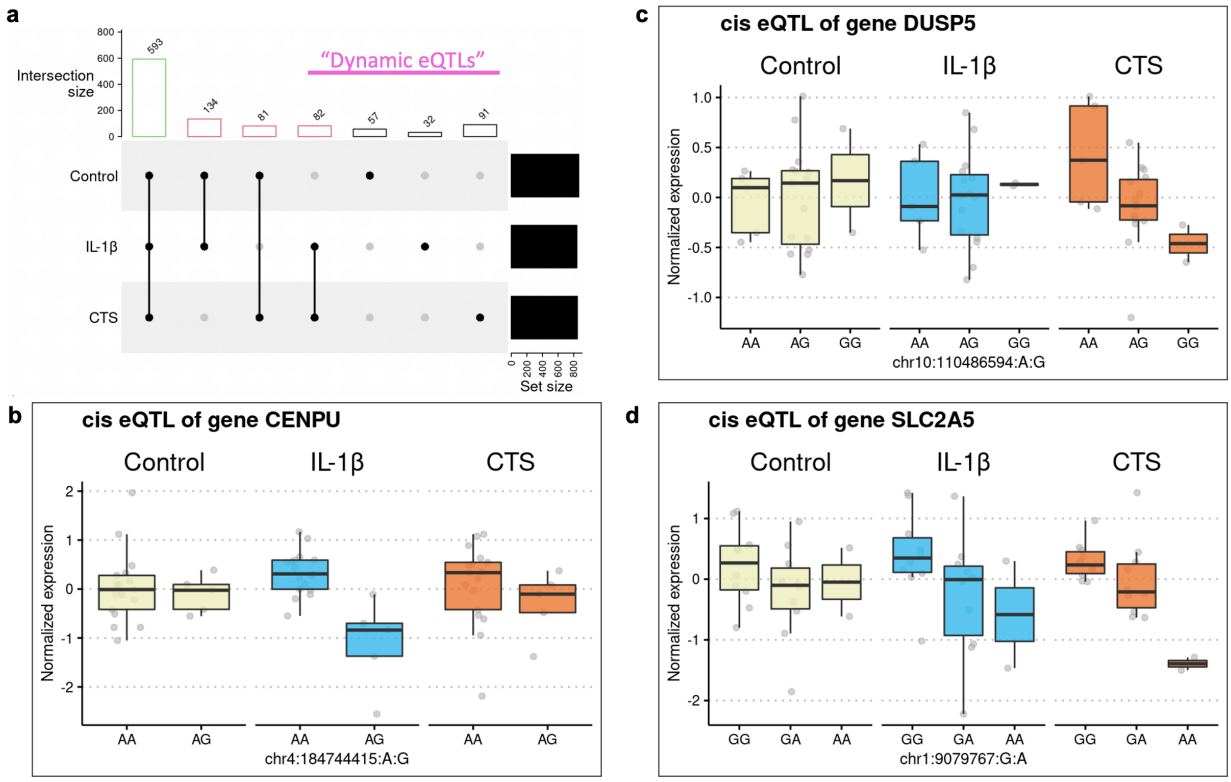


Figure 3.4: Sharing of eQTLs between conditions and dynamic eQTLs.

(a) Upset plot of distinct sets of eQTLs after analysis of sharing of eQTLs across conditions by an FDR < 0.01 in the primary condition and a more lenient secondary cutoff of p value < 0.548 for sharing in other conditions (see Methods). Each column corresponds to the size of the set of shared eQTLs across the conditions with filled-in pegs in the lower plot. Set sizes of eQTLs found to be significant at either the primary cut-off of FDR < 0.01 or at the cutoff of p < 0.548 in the secondary sharing analysis are plotted to the right of the plot. eQTLs that are either specific (not shared with any other condition) to one condition, or shared between the IL-1B and CTS but not control condition, are termed dynamic eQTLs. **(b-d)** Box plots of example dynamic eQTLs. The expression counts of the gene, log normalized and adjusted for library size and 7 gene expression PCs, are plotted on each y axis. Box plots are broken down by genotype at the SNP listed on the x axis and faceted by treatment condition group. (b) Represents an IL-1β condition-specific dynamic eQTL, (c) represents a CTS condition-specific dynamic eQTL, and (d) represents an IL-1β-CTS dynamic eQTL.

3.3.4 Dynamic eQTLs are more specific than nondynamic eQTLs

Previous analyses of eQTLs across dozens of different tissues in static conditions by the GTEx project have found a high degree of eQTL sharing between tissues [17]. Despite the lack of either cartilage or bone within the GTEx dataset, we sought to contextualize the specificity of the eQTLs identified in iPSC-chondrocytes by considering the naïve overlap between dynamic and nondynamic iPSC-chondrocyte eQTLs with the sets of eQTLs identified in each GTEx tissue. We find that in general, more closely related tissues to chondrocytes, including mesoderm derived tissues such as adipose and skeletal muscle, display higher degree of overlap of eQTLs with both dynamic and nondynamic eQTLs identified in iPSC-chondrocytes than more distantly related tissues such as brain or liver (Figure 3.5; Supplementary Figure 3.4). Furthermore, in almost all GTEx tissues, the proportional overlap observed for nondynamic eQTLs was higher than that of dynamic eQTLs, further reinforcing the context-specificity of dynamic regulatory effects. Although the overlaps of eQTLs with GTEx are likely underestimations in each tissue due to incomplete power, we expect this to only result in a bias towards an observation of greater overlap in GTEx tissues with more eQTL mapping power. We do not expect the incomplete power issue to impact the result of dynamic eQTLs being more specific than nondynamic eQTLs.

3.4 Discussion

While gene expression is impacted by genetic variation, some patterns of gene regulation are only revealed during and therefore can only be studied in cells undergoing perturbed states. In order to study the genetic control of gene regulation of cartilage in both static and dynamic conditions, we differentiated iPSC-chondrocytes and treated these cells with control, pro-

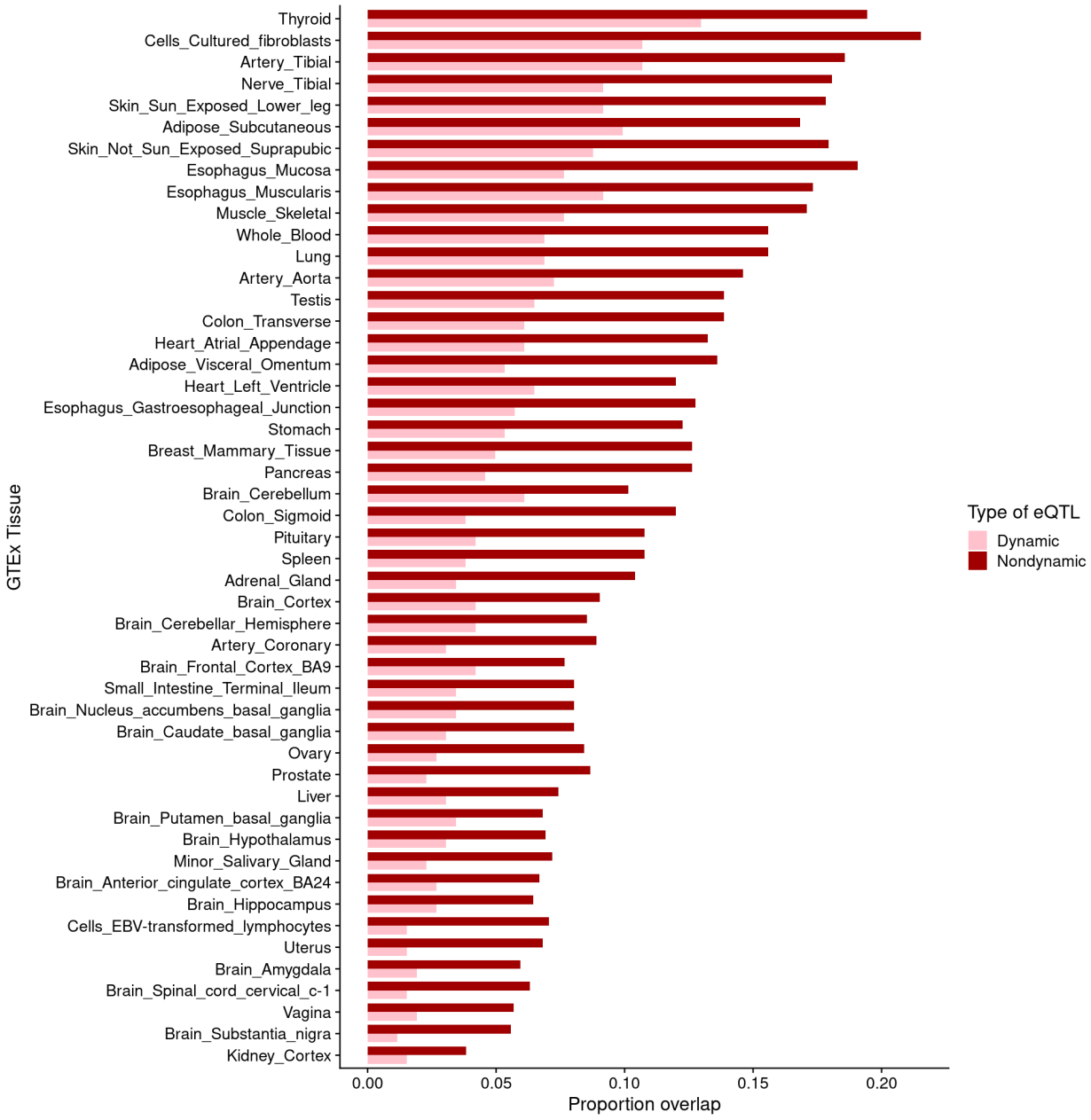


Figure 3.5: Dynamic eQTLs are more specific than nondynamic eQTLs.

Proportion of dynamic and nondynamic eQTLs (SNP-gene test pairs) that are also detected in GTEx tissues. The Y axis is ordered according to the size of the sum of the proportion of the two eQTL overlaps for each tissue, from largest to smallest. Due to incomplete power in GTEx tissues, the proportion of overlap is likely an underestimate of the true overlap of effects.

inflammatory, and biomechanical stress conditions. The responses of the cells, as well as the dynamic patterns of gene regulation uncovered through quantitative trait locus mapping, are in a disease-relevant cell type in disease-relevant conditions. The gene-by-environment interactions revealed by comparing the effects of eQTLs across conditions represent context-specific gene regulation that may help to explain the basis for interindividual differences in susceptibility to disease.

3.4.1 Relevance and limitations of *in vitro* OA models for human disease

While *in vitro* treatment models of OA seek to recreate environmental conditions seen in the disease to induce phenotypes mimicking human OA, no model of disease is perfect at recapitulating all aspects of *in vivo* biology. Here, we present the responses of iPSC-chondrocytes to two differing dimensions of treatment types: a physical treatment in the form of mechanical stress and a chemical treatment in the form of the pro-inflammatory cytokine IL-1β. Both of these treatments have been used in primary tissue culture of cartilage to produce phenocopies of OA *in vitro*. Despite the obvious differences between the conditions in a cell culture dish and those within synovial capsules, the patterns of gene expression activated by these individual treatments hold relevance to OA, as demonstrated by the enrichment of activated genes in sets of genes that are known to be altered during the course of the joint disease.

Additionally, monolayer culture of iPSC-chondrocytes, as done in this study, necessarily places chondrogenic cells in an unfamiliar context; monolayer culture of primary chondrocytes is known to lead to transcriptional and phenotypic changes characteristic of hypertrophy and dedifferentiation [158]. Therefore, despite the detected expression of several chondrogenic genes amongst our collected samples (Supplementary Figure 3.1), the relative immaturity of our iPSC-

chondrocytes is not surprising. Three-dimensional cell culture of iPSC-chondrocytes during differentiation and treatment would improve the maturity of cells [159], but this would also introduce additional heterogeneity within and between cultures. Therefore, the monolayer cultures presented here represent a first step towards understanding the cells of human joints without the need for invasive sampling of volunteers.

3.4.2 Shared and specifically activated gene expression patterns in methods of OA induction

By comparing the sets of differentially expressed genes in the two OA induction treatments and using a statistically robust method to better estimate sharing of effects across treatments, we identify patterns of gene activation that are shared and unique across the two *in vitro* OA treatments.

The set of differentially expressed genes that are shared between comparisons is enriched for annotations corresponding to development of skeletal system components and organization of the extracellular matrix. The same set of genes is also enriched amongst two publicly available OA-related DE gene sets, demonstrating relevance to joint disease. Given that both IL-1β and CTS are used as models of OA and are known to induce expression of catabolic genes and hypertrophic differentiation in cartilage [68–71,77,83–87], these gene expression response patterns and enrichments are in line with expectations.

Furthermore, as expected, the set of genes uniquely activated by IL-1β treatment is enriched for pathways related to immune and inflammatory responses. However, pathways related to development of bone, disassembly of the extracellular matrix, and proteoglycan biosynthetic processes are also uniquely impacted by the cytokine treatment. For example, the *SMAD5* gene, involved in signaling of bone morphogenic protein *BMP* in chondrogenesis and

endochondral bone formation [160], is specifically downregulated in iPSC-chondrocytes in the context of IL-1β treatment. Interestingly, genes annotated with the "cellular response to mechanical stimulus" gene ontology term such as the mechanically activated ion channel *PIEZO1,* are also found to specifically respond to IL-1β treatment.

In contrast, the control-CTS specific DE gene set is enriched for pathways related to cellular stress and apoptosis, but not enriched for terms related to extracellular matrix metabolism or amongst the two external OA DE sets. This lack of enrichment is not due to insufficient numbers of unique DE genes in this set, as at 2234 genes, the set is larger than that of the control-IL-1β specific gene set. Nor does this lack of enrichment due to a lack of relevance of CTS treatment, as the entire set of CTS-impacted genes (not restricted to those uniquely impacted by CTS) is enriched for relevant terms and gene sets (Figure 3.2; Figure 3.3).

Taken together, the nature of the enrichments amongst genes specifically impacted by IL-1β and the lack of enrichment amongst the set of genes specifically impacted by CTS treatment suggest that the form and degree of mechanical stress used in these experiments induces gene expression changes that are relevant to joint health and osteoarthritis, but that these changes may not necessarily be specific to mechanical stress. In comparison, inflammatory cytokine treatments may not only directly impact inflammatory pathways in iPSC-chondrocytes but may also impact these cells in a way that sensitizes them to mechanical stress. Indeed, prior studies of another inflammatory cytokine IL-1α in cartilage have found similar enhancements of *PIEZO1* expression and potential increased function of the resulting protein [161]. As *PIEZO1/2* are both highly expressed in OA cartilage and respond particularly to injurious strains [162,163], there may be interactions between these two environmental treatments that merit further study. Another relationship between mechanical stress and inflammation in cartilage is the fact that

mechanical loading of cartilage may have a protective effect against the matrix-degrading phenotypes induced by cytokine administration [164]. Although not performed here, a combination treatment encompassing simultaneous exposure to both treatments could not only help refine the understanding of the shared pathways activated by the two treatments individually, but also the potential interactions between the two treatments on the transcriptional level and their potential impacts on cartilage degradation in OA.

3.4.3 Dynamic eQTLs reveal context-specific genetic control of gene regulation

We identify several classes of dynamic eQTLs in our data, depending on the conditions in which they are detected (Figure 3.4, Supplementary Figure 3.3). One example of an IL-1β specific eQTL affects *CENPU* (Figure 3.4b). This gene has been identified as a marker for chondrogenic progenitor cells (CPCs) [148], thought to migrate to sites of damaged cartilage in OA to participate in tissue repair. As inflammation in the joint in response to acute injury has conflicting effects, stimulating both repair and continued degeneration of tissue [165], inter-individual differences in the ability to recover from cartilage damage in the early stages of joint disease may be influenced by genetic regulation of genes activated by inflammation.

CTS-specific dynamic eQTLs display effects in CTS conditions, but not in other conditions. Interestingly, we identify dynamic eQTLs in CTS that affect genes related to cellular responses to inflammation. (Figure 3.4c; Supplementary Figure 3.3e). *DUSP5* is known to regulate cellular inflammation, has reduced expression in OA cartilage, and negatively regulates IL-1β induced inflammation in a rat chondrocyte model of OA [166]. As in the differential expression analysis, the interplay between different perturbations relevant to OA and their effects

on gene regulation may present key insights into their role in contributing to OA disease progression.

IL-1β – CTS specific dynamic eQTLs display effects under perturbation but not in steady-state control conditions. An example of such an eQTL is one affecting *SLC2A5* expression (Figure 3.4d). This gene is differentially expressed and possesses enhancers that display differential chromatin accessibility between damaged and intact OA cartilage [156,167]. One potential mechanism behind dynamic eQTLs such as this may be a lack of expression or gene regulatory activity of a particular gene in static conditions. Only under perturbations such as inflammation or mechanical stress, or under conditions of cartilage damage in disease, is the regulation of these genes visible, and therefore subject to identification via QTL or other analyses.

Overall, dynamic eQTLs offer a view into gene regulation that is not possible when studying cells and tissues under steady-state conditions alone. We also identify nondynamic eQTLs in the data, including eQTLs with effects found across in all three conditions. Some of these shared eQTLs affect the expression of genes responsible for the production or modification of extracellular matrix protein components (Supplementary Figure 3.3g-h). In particular, the lysyl hydroxylase 1 gene (*PLOD1*) catalyzes the conversion of lysine to hydroxylysine in collagens and is associated with Ehlers-Danlos Syndrome, a connective tissue disorder that results in generalized joint hypermobility and skin fragility [168,169]. Similarly, chondroitin polymerizing factor 2 (*CHPF2*) is affected by a nondynamic shared eQTL and is involved in the polymerization of chondroitin, a structural component of cartilage. Nondynamic eQTLs display their effects more generally across multiple environmental conditions and may impact more

general pathways in chondrocytes, including developmental pathways that are dysregulated in a variety of joint diseases not directly related to OA.

The difference in overlap between our two sets of eQTLs identified in perturbed and unperturbed iPSC-chondrocytes and steady-state GTEx eQTLs demonstrates the context-specific nature of the gene regulatory programs that dynamic eQTLs represent. The fact that many eQTLs in iPSC-chondrocytes do overlap with both related and unrelated tissues in this external dataset is not altogether unsurprising, given the large amount of overlap in eQTL discovery between GTEx tissues themselves. However, despite the sizable statistical power of GTEx, the ability to detect eQTLs in this study that are not identified in this external dataset highlight the potential for additional discoveries when searching in other cell types and cell states.

A recent study investigating eQTLs mapped in matched low-grade and high-grade OA cartilage samples from 115 patients found 32 genes displaying so-called “differential eQTLs”, with an effect that could be detected in either low-grade or high-grade cartilage but not both groups [170]. While this represents a method of investigating disease context-specific genetic control of gene regulation, the sharing of the joint space and overall synovial environment between low-grade and high-grade samples in this study may mask differences in gene regulation that would be observed between truly healthy and diseased tissue, potentially explaining the low number of dynamic eQTLs identified. Despite its limitations, modeling the cellular environment *in vitro* is a potential way to control for the cellular environment to more specifically interrogate gene-by-environment interactions.

In summary, here we profiled the response to inflammatory cytokine exposure and mechanical stress in iPSC-derived chondrogenic cells. We find shared and specific patterns of

responses to different *in vitro* OA treatments and examples of hidden patterns of genetic control of gene regulation that are made visible by changes to the cellular environment.

3.5 Methods

3.5.1 Samples

We selected 22 individuals from the Yoruba YRI HapMap population [134]. iPSCs were previously reprogrammed from lymphoblastoid cell lines from these individuals [96]. This number of individuals has been sufficient to identify eQTLs in previous studies [123,171]. Experiments were performed to avoid confounding of technical variables with the variable of interest of condition (Figure 3.1).

3.5.2 iPSC-MSC and iPSC-chondrocyte differentiation

Mesenchymal stem cells (MSCs) were differentiated from iPSCs as previously described [172]. iPSC-MSCs have been previously shown to display phenotypes and cell surface markers characteristic of primary MSCs [172]. A random subset of 16 iPSC-MSC lines used in this study were characterized through flow cytometry to display cell surface markers characteristic of primary MSCs. iPSC-MSCs were detached from culture flasks using 0.05% Trypsin/EDTA and seeded at a density of 92,000 cells/well onto the center of wells of BioFlex Type I Collagen coated 24-well Culture Plates (FlexCell International HTPW-3001C) using BioFlex cell seeders (FlexCell International). Cells were seeded using a regimen of 5% elongation for 2.5 hours. After seeding, cells were cultured in serum-free chondrogenic differentiation medium [92], consisting of high glucose DMEM, 100 mg/mL Penicillin/Streptomycin, 50mg/mL L-Proline, 200mM GlutaMax, 50mg/mL L-Ascorbic acid-2-Phosphate, 11g/L Sodium pyruvate, 5mM

Dexamethasone, 1x ITS Premix, and supplemented with 10 ng/mL TGF-β3. The chondrogenic medium was changed every 2-3 days for 14 days. iPSC-chondrocytes differentiated for 14 days have been previously shown to display Alcian blue and *COL2A1* immunostaining demonstrating production of proteoglycans and collagen II [172], the latter being specifically expressed in human cartilage. iPSC-chondrocytes also demonstrate expression of several chondrogenic-related genes (Supplementary Figure 3.1).

3.5.3 Cyclic tensile strain regimen and Interleukin-1 beta cytokine treatment

iPSC-MSCs from each individual were seeded simultaneously on BioFlex Type I Collagen coated 24-well Culture Plates and differentiated for 14 days in chondrogenic media. Plates containing iPSC-chondrocytes from each individual were then assigned to one of three treatment groups: an untreated control group, a cyclic tensile strain treated group, and an IL-1β treated group. Cells from all groups were placed in the same incubator (at 37°C, 5% CO2, and atmospheric O2), during the 24 hours of treatment.

iPSC-chondrocytes in the control group were fed with serum-free chondrogenic differentiation medium for 24 hours.

iPSC-chondrocytes in the cyclic tensile strain group were fed with serum-free chondrogenic differentiation medium and treated with a cyclic tensile strain (CTS) regimen that is known to induce an OA-like phenotype using the Flexercell FX6000 Tension System (Flexcell International) [68–71]. Plates were loaded onto the Flexercell baseplate and a vacuum was used to deform the cell culture plate membrane and create uniform biaxial cyclic tensile strain. Specifically, 1.8% elongation (15kPa) of CTS was applied to the cells at a rate of 0.5 Hz for 24 hours. This strain loading regimen has been previously shown to induce the expression of

chondrocyte hypertrophy-related marker genes in iPSC-chondrocytes by quantitative real-time reverse transcription PCR (RT-PCR) compared to matched untreated controls [172].

iPSC-chondrocytes in the IL-1β treated group were fed with serum-free chondrogenic differentiation medium containing IL-1β (10 ng/ml) for 24 hours. This concentration of IL-1β has been shown to induce an OA-like phenotype in primary chondrocytes [173–177].

3.5.4 Bulk RNA extraction and sequencing

RNA was extracted from cells following control, CTS, and IL-1β treatments using the ZR-Duet™ DNA/RNA MiniPrep kit (Zymo D7001). RNA concentration and quality were measured using the Agilent 2100 Bioanalyzer. Library preparation was performed over three batches using the Illumina TruSeq Stranded mRNA library prep kit (20020594, Illumina). Samples were sequenced to 28+89 base pairs, paired-end on three SP flow-cells using the Illumina NovaSeq at the University of Chicago Genomics Core Facility according to manufacturer instructions. A minimum of 10 million reads mapped to the hg38 genome were generated per sample. We used FastQC to confirm that the reads were of high quality.

3.5.5 Quantifying the number of bulk RNA-seq reads mapping to genes

Reads were mapped to the hg38 genome using STAR (version 2.6.1b) [138]. Gene expression levels were quantified using the featureCounts function in Subread (v1.6.5 RRID:SCR_009803) using standard parameters [150]. All downstream processing and analysis steps were performed in R (v4.2.0, RRID:SCR_001905) unless otherwise stated.

3.5.6 Transformation and normalization of bulk RNA-seq reads

Log2-transformed counts per million (CPM) were calculated from raw counts for each sample using the edgeR package (RRID:SCR_012802) [151]. Lowly expressed genes were filtered such that only genes with an expression level of log2(CPM) > 3 in at least 8 samples were kept for downstream analyses. Genes originating from the mitochondrial or sex chromosomes were removed from downstream analyses. Raw read counts from 10,821 genes that passed these filters were normalized using the upper quartile normalization method to account for the number of reads sequenced across samples for each gene [178].

3.5.7 Analysis of sources of variation in bulk RNA-seq data

Principal component analysis (PCA) was performed on the normalized log2(CPM) values from above. A linear regression analysis was then performed between each of the top 5 PCs and several biological and technical variables. These variables included individual cell line, treatment, RIN score, collection batch, number of passages as a feeder iPSC, number of passages as a feeder-free iPSC, and number of passages as an iPSC-MSC. P values from the regression were corrected using the Benjamini Hochberg (BH) procedure. Results with a BH-adjusted p value < 0.05 were considered significant.

Pairwise Pearson correlations were computed between normalized log2(CPM) values between samples. Correlation values were plotted in a heatmap using the gplots::heatmap.2 function in R, and dendrograms organizing the samples by correlation weights between samples were used to reorder samples in the heatmap.

3.5.8 Regressing out experimental variables in RNA sequencing data

A linear mixed model was fit to the gene expression data using the dream function of the variancePartition R package to trimmed mean of M-values normalized read counts (RRID:SCR_019204). Collection batch and individual were modeled as random effects while sex and RIN score were modeled as fixed effects in the linear mixed model for DE comparisons as in equation (1). Residuals of the model fit to the gene expression data, representing measured gene expression after accounting for the variables included in the linear mixed model, were extracted and examined using PCA.

Equation 1:

$$Y \sim \beta_1 * sex + (1|individual) + (1|collection\ batch) + \beta_4 * RIN + \varepsilon$$

3.5.9 Differential expression analysis of bulk RNA-seq data

Differential expression (DE) was measured using a linear-model-based empirical Bayes method in the dream function of the variancePartition R package to trimmed mean of M-values normalized read counts that passed filtering steps (RRID:SCR_019204).

Collection batch and individual were modeled as random effects while treatment, sex and RIN score were modeled as fixed effects in the linear mixed model for DE comparisons as in equation (2). Genes with an adjusted p value < 0.05 in comparisons between control-CTS treatment and control-IL-1β treatment were considered DE.

Equation 2:

$$Y \sim \beta_1 * treatment + (1|individual) + (1|collection\ batch) + \beta_4 * RIN + \beta_5 * sex + \varepsilon$$

3.5.10 Enrichment of DE genes in biological pathways

Using topGO (RRID:SCR_014798), we assessed enrichment of Gene Ontology (GO) biological processes among DE genes. For DE genes in each of the two treatment comparisons, a Kolmogorov-Smirnov test using adjusted p values was used for assessing enrichment of GO processes, and the top 20 most enriched terms were reported. For assessing the GO enrichments of shared and unique gene sets between the two treatment comparisons, the weight01 algorithm was used with a Fisher's exact test. Terms with a p value < 0.01 were considered significantly enriched.

3.5.11 Analysis of sharing of DE genes between conditions

A multivariate adaptive shrinkage (mash) model was fit to the DE genes detected in the control-CTS treatment and control-IL-1β treatment to estimate sharing of significant effects between the two conditions [179]. To account for correlations between tests from the two comparisons due to both groups sharing the same control samples, the correlations between null tests from both comparisons were estimated. The mash model was fit using the null correlation matrix, data-driven covariance matrices determined through empirical Bayes matrix factorization and PCA, as well as canonical covariance matrices to best account for multiple potential patterns of covariance between tests across the two comparisons.

3.5.12 Enrichment of DE gene sets in OA-related DE gene sets

To test for enrichment of sets of OA-related genes in our DE genes, a two-sided Fisher's exact test was used. In all enrichment tests, the background gene set was the complete set of genes tested for DE in our analyses (n = 10,821 genes). Two differential gene expression studies between healthy and OA human cartilage [155] and damaged vs intact OA cartilage [156] were

used to assess enrichment amongst DE gene sets. Specifically, the first gene set comprised 2692 DE genes detected in Soul et al., 2018 from comparisons between RNA-Seq collected from total knee replacement cartilage from 60 patients with OA and 10 control patients without OA. The second gene set comprised 1575 differentially expressed genes detected in Dunn et al., 2016 between damaged and intact knee cartilage from 8 OA patients undergoing total knee replacement.

3.5.13 Expression quantitative trait locus mapping

We used the TReCASE framework in the asSeq software package to leverage both total read counts and allele-specific expression in our RNA sequencing samples to identify *cis*-eQTLs in iPSC-chondrocytes separately in control, IL-1β, and CTS conditions [157,180]. We selected the TReCASE method to maximize the power to detect eQTLs given the relatively small sample size of the study through using allele-specific expression information. Phased genotypes for hg38 from our 22 individuals were obtained from the 1000 Genomes project [140]. Total read counts for each gene in each sample were obtained, and allele specific counts mapped to non-overlapping exons comprising haplotypes for each gene were obtained using asSeq functions. We required over half of samples to have at least 5 mapped counts for a given gene for it to be tested within the TReC model, and for at least 5 samples to have at least 10 allele specific counts for a gene to be tested in the ASE model. Seven gene expression PCs (computed using the log(raw total count + 1 / library size) of genes in each sample) and the total library size of each sample were included as covariates in the eQTL mapping procedure. We tested variants 50 kb upstream and 50 kb downstream of the TSS, resulting in 3,508,367 shared tests across all three conditions (control: 3,710,311 tests, IL-1β: 3,721,204 tests, CTS: 3,753,050 tests). For

downstream analyses, we subsetted the eQTL mapping results to include those coming from the set of tests that were included across all three conditions.

TReCASE p values were adjusted on two hierarchical levels, first using eigenMT to account for multiple SNP-gene tests for each gene and LD between SNPs [181]. Then, the SNP-gene pair with the lowest p value for each gene was selected from each gene. A secondary p value adjustment was performed across the selected SNP-gene pairs over all genes using the Benjamini-Hochberg procedure [182]. SNP-gene tests with an FDR < 0.01 were considered significant. This FDR threshold was chosen to maximize the amount of sharing of eQTL effects determined between conditions (see "Identification of dynamic eQTLs") in order to increase the reliability of dynamic eQTL calling.

3.5.14 Identification of dynamic eQTLs

To overcome incomplete power in our dataset when calling dynamic eQTLs, we employed a strategy utilizing Storey's $\pi 1$ method [183]. Specifically, we first identified significant eQTLs using an FDR cutoff of < 0.01 separately in each of the three conditions. We then determined and plotted the union of all nominal p values assigned to these tests in the other two conditions, conditioning on an FDR > 0.01 in each of the other two conditions (the tests are not also called as significant in the other two conditions). These p values are expected to follow a uniform distribution if we have complete power to detect eQTLs. We further plotted the p value distribution of all tests in all conditions with an FDR < 0.01 (Supplementary Figure 3.2). The enrichment of p values on the low-end of the histogram in the first distribution of p values over the second distribution is due to the sharing of effects between conditions. To determine the secondary nominal p value cutoff at which to determine sharing of effects between conditions,

we obtained the intersection of the density functions of the two distributions. This intersection (p value: 0.0548) represents the point at which values deviate from the null distribution under a model assuming no sharing between conditions.

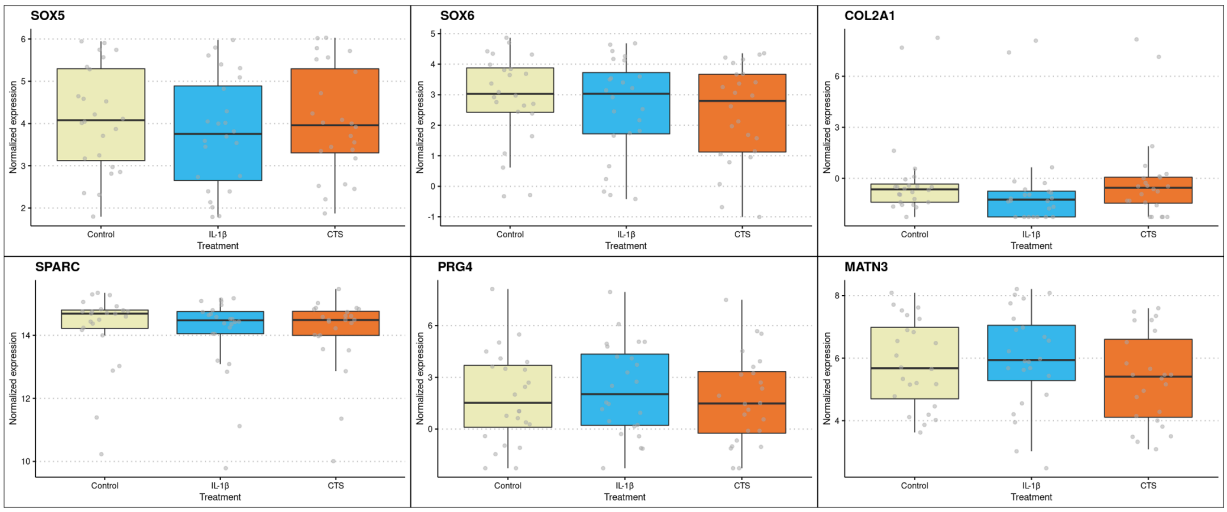
We iterated through this procedure, choosing a range of FDR thresholds between 0.1 and 0.001, in order to determine the FDR cutoff at which sharing of eQTL effects between conditions is maximized. We chose this value (FDR < 0.01) as it maximized pairwise sharing of eQTLs between conditions to increase the robustness of called dynamic eQTLs

3.5.15 Integration of eQTLs with GTEx data

eQTL summary statistics from the Genotype Tissue Expression Project (GTEx, v8) were obtained [17]. eGenes, eSNPs, and significant test pairs from GTEx from each of the 49 tissues were intersected with our eQTL sets.

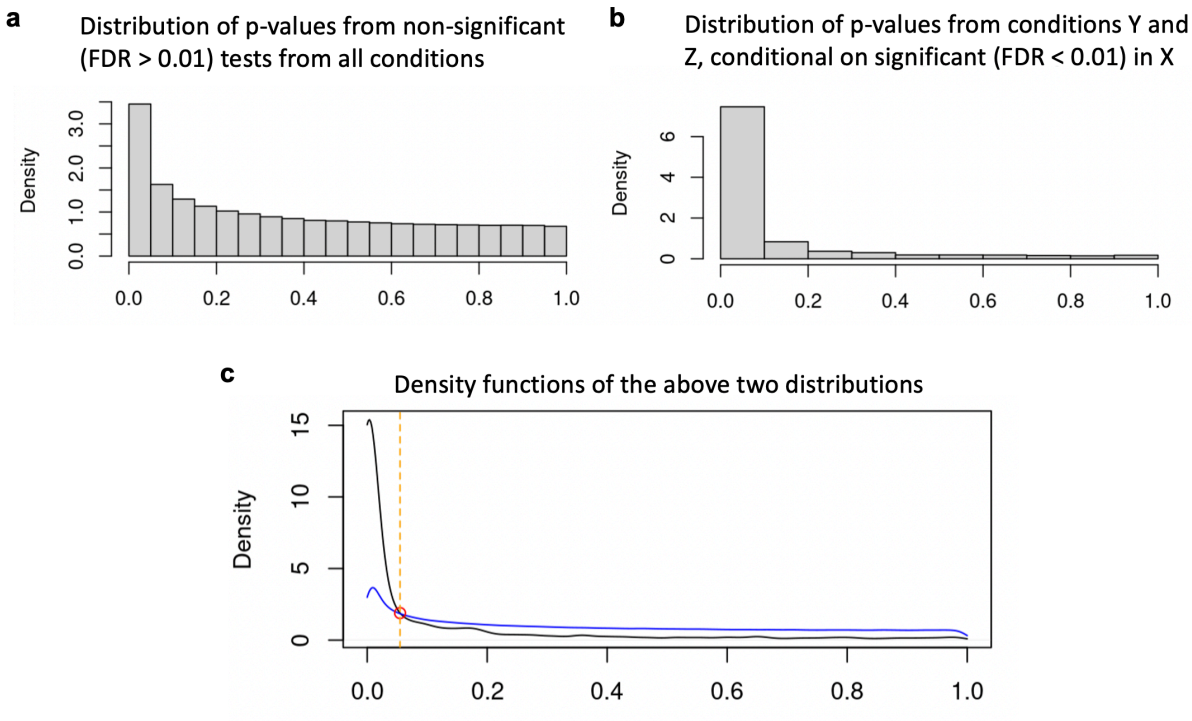
3.6 Supplementary Information

3.6.1 Supplementary Figures



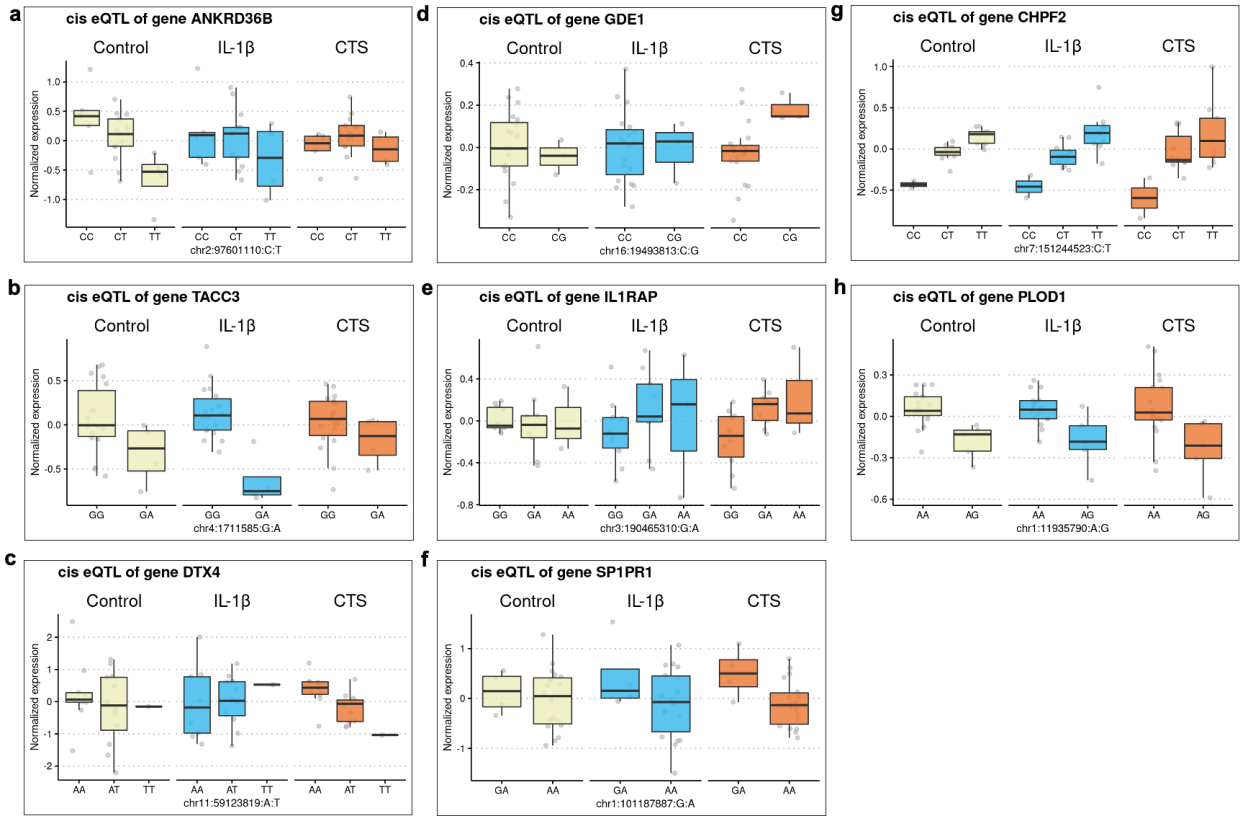
Supplementary Figure 3.1: Expression levels of several chondrogenic genes across RNA sequencing samples.

Y axis in each plot is gene expression counts per million reads normalized by upper quantile normalization across libraries. X axis in each plot is bulk RNA sequencing samples grouped by condition (A= control, B = IL-1β, C = CTS).



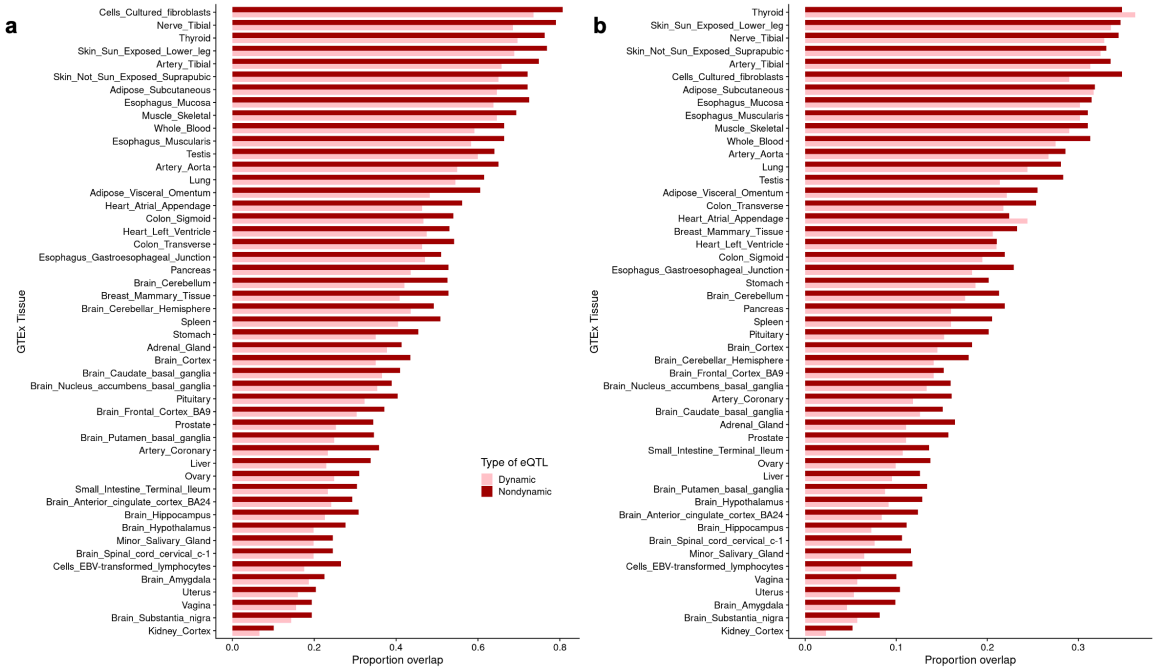
Supplementary Figure 3.2: Graphical depiction of the method used to determine the secondary cutoff for calling dynamic eQTLs.

(a) Histogram of nominal p values from all tests in all conditions that were not significant (FDR > 0.01). **(b)** Distribution of nominal p values of non-significant tests (FDR > 0.01) conditioned on being significant in a different condition (FDR < 0.01). For example, a test called as significant in control and not called as significant in IL-1B or CTS would have its p values from the IL-1B and CTS tests included in this histogram. **(c)** The density functions of the two histograms from (a) and (b) are plotted, and the intersection of the two plots, signifying the point at which the p values in distribution in (b) deviate from the null, is marked with a red circle and orange vertical line. We used this point (p value 0.0548) as our second more lenient cutoff when assessing sharing of eQTLs between conditions.



Supplementary Figure 3.3: Additional eQTL box plots.

Box plots of additional example dynamic eQTLs. The expression counts of the gene, log normalized and adjusted for library size and 7 gene expression PCs, are plotted on each y axis. Box plots are broken down by genotype at the SNP listed on the x axis and faceted by treatment condition group. **(a)** Represents a control condition-specific dynamic eQTL, **(b)** represents an IL-1β condition-specific dynamic eQTL, **(c-e)** represent CTS condition-specific dynamic eQTLs, **(f)** represents an IL-1β-CTS dynamic eQTL, and **(g-h)** represent two eQTLs with effects detected across all treatments.



Supplementary Figure 3.4: Additional plots of GTEx overlap with dynamic and nondynamic eQTLs.

Overlap of dynamic and nondynamic (a) eGenes or (b) eSNPs and those detected in GTEx tissues. The Y axis is ordered according to the size of the sum of the proportion of the two overlaps for each tissue, from largest to smallest.

CHAPTER 4

DISCUSSION

Together, the two projects presented in this thesis leverage the power of induced pluripotent stem cell (iPSC) derived cell types and *in vitro* treatments to explore the impacts of gene-by-environment interactions on gene regulation. First, in chapter 2, I characterize an experimental system for studying gene regulation changes relevant to joint health and osteoarthritis (OA). Although the cells that resulted from iPSC-derived chondrogenic differentiation in this work were relatively immature, following an *in vitro* biomechanical strain treatment, they displayed gene expression responses relevant to adult disease processes. Additionally, this system induced robust gene expression responses that we could reliably detect across cells with different genetic backgrounds. In chapter 3, I present a project that leverages and builds upon this experimental system to characterize dynamic gene regulation in static conditions and two environmental perturbations relevant to OA. Gene expression changes stimulated by both *in vitro* OA treatments suggest that inflammatory pathways play shared and important roles in active disease progression and response of joint cells to pathological environmental conditions. Simultaneously, the presence of genetic control of gene expression unique to each environmental perturbation reinforces the importance of cell context in understanding the genetic basis of complex diseases such as OA, and these context-specific genetic impacts on gene regulation may help to explain inter-individual disease risk in OA and other joint diseases.

4.1 Relevance of cell type and environmental context to disease

The difficulty of sampling skeletal tissues, like bone and cartilage, has hampered efforts to investigate the role of gene regulation in skeletal traits and diseases. As an alternative to primary cells, I used iPSCs, which can be obtained relatively readily from many genetically diverse individuals and subsequently differentiated into multiple cell lineages, to obtain chondrogenic cells to study regulatory patterns in cartilage. Further, I exposed these iPSC-chondrocytes to controlled environmental conditions that can be adjusted to reflect contexts relevant to osteoarthritis. Specifically, the work presented here investigates one joint-related cell type (iPSC-chondrogenic cells) in two disease-relevant contexts that model aspects of the OA joint (biomechanical strain and inflammatory cytokine exposure). While the *in vitro* treatment models of OA used in this work seek to recreate environmental conditions seen in the disease to induce phenotypes mimicking human OA, I acknowledge that neither model of disease is a perfect copy of *in vivo* biology. At the same time, each of these treatments have been used in primary tissue culture of cartilage to produce phenocopies of OA *in vitro*. The patterns of gene expression activated by these individual treatments hold relevance to OA, as demonstrated by the enrichment of activated genes in sets of genes that are known to be altered during the course of the joint disease. Thus, while imperfect, the system constructed here has much relevance to human joint health and holds many advantages compared to alternative models, thus serving as a good first step into investigating gene regulation in OA.

4.1.1 Advantages of an *in vitro* system for modeling gene regulation in OA

In vitro systems offer the ability to finely control cellular environments and are therefore advantageous for research of gene regulation patterns, as such patterns can be highly context

dependent. However*, in vivo* models of osteoarthritis, especially those in mice, exist and see wide use in OA studies and offer genetically and experimentally tractable ways to study OA [184]. Furthermore, human studies of diseased OA joints sampled after joint replacement surgery also offer a view into the biology of this disease in the entire synovial capsule, unabstracted as in they are in OA models that contain only one or a few cell types [170]. At the same time, these human and animal *in vivo* studies contend with several issues when attempting to study the precise impact of cellular environments on gene regulation.

For one, animal models of OA do not necessarily develop OA in the same way that humans do. Animal models of OA typically require the destabilization of joints through surgical resection of joint ligaments to lead to mechanical changes within the joint environment [184], making them more a closer model of post-traumatic OA. While animals of many species do naturally develop OA in the wild and captivity [184], naturally occurring animal models of OA require lengthy experimental time to allow for animals to reach maturity and further develop disease, greatly increasing the financial and temporal cost of experiments and making short-term experiments difficult. In contrast, *in vitro* OA models of primary or differentiated human joint cells allow for the induction and observation of OA phenotypes on a faster time scale while also ensuring the species-relevance of findings to humans.

Human OA tissue samples that are available for sampling in OA research studies are typically collected at the end stages of disease as opportunistic samples from joint replacement surgery. These tissues are therefore often characterized by extensive joint tissue degeneration. The severity of disease in these samples make it difficult to understand the early disease process of OA. Furthermore, it is difficult to determine whether gene expression changes that can be observed in late stage OA are the drivers or results of the disease process. Furthermore,

confounding from medication status of patients, environmental exposures not measured by researchers, and logistical difficulties associated with high quality sample collection and preservation can all introduce noise to gene-by-environment studies in human subjects. A recent study mapped eQTLs separately in matched low-grade and high-grade OA cartilage samples from 115 patients and found 32 genes displaying so-called "differential eQTLs", with an effect that could be detected only in one group of cartilage but not both groups [170]. As the synovial environment of low-grade and high-grade samples in this study is the same within each individual, differences in gene regulation that would be observed between truly healthy and diseased tissue are masked in this study design. This fact potentially explains the low number of differential eQTLs identified despite the large sample size of the study. Despite its limitations, modeling the cellular environment *in vitro* is a potential way to control for the cellular environment to more specifically interrogate gene-by-environment interactions. Thus, *in vitro* studies such as the one conducted in this thesis work can detect many more examples of robust dynamic genetic control of gene regulation with relatively small sample sizes [21,123].

Overall, the ability to finely control the cellular environment of human cells *in vitro* makes such experimental systems tractable and advantageous for gene-by-environment studies of gene regulation in human disease.

4.1.2 Mechanical strain compared to other mechanical stresses in OA models

CTS models do not directly mimic the compressive or shear biomechanical stresses experienced by joint chondrocytes *in vivo.* However, the previous extensive use of CTS as a model for studying the effects of extra-physiological stresses in cultured cells validates its use as an OA model [62–65,68–71,128]. Furthermore, the patterns of sharing observed on both the

differential expression and eQTL levels between CTS perturbation and IL-1β treatment suggest that there are pathways that are conserved even across types of treatments, and likely will be shared pathways between types of mechanical treatments as well. Furthermore, the lack of compressive force models of OA that can be applied to monolayer cultures of chondrocytes necessitated the use of CTS in this work. Adapting the system to include three-dimensional differentiated iPSC-chondrocytes is possible, but would also introduce additional undesired consequences such as heterogeneity of cultures.

4.1.3 Monolayer vs three-dimensional culture of iPSC-chondrocytes

Monolayer culture of iPSC-chondrocytes necessarily places these chondrogenic cells in an unfamiliar context compared to being embedded as they would be in primary cartilage tissues. In fact, monolayer culture of primary chondrocytes is known to lead to transcriptional and phenotypic changes due to hypertrophy and dedifferentiation [158]. Therefore, the relative immaturity of our iPSC-chondrocytes observed in chapter 2 is not surprising. Three-dimensional cell culture of iPSC-chondrocytes during differentiation and perturbation treatments would improve the maturity of cells and potentially the relevance of the findings to primary disease [92,146]. However, three-dimensional culturing would also introduce additional heterogeneity within and between cultures compared to the relative uniformity of monolayer cultures, due to the differential permeability of cellular aggregates to differentiation factors in differentiation media or oxygen as well as cytokine treatments [185]. Therefore, while the monolayer cultures presented here represent a first step towards understanding the cells of human joints without the need for invasive sampling of volunteers, cellular aggregates may further improve the power of this system to model primary human OA. Application of single-cell sequencing methods to

deconvolute three-dimensional cultures into cell types of interest would be a further option to ameliorate the heterogeneity of such culture systems.

4.1.4 Alternative environmental contexts relevant in skeletal biology

In general, every cell type and environmental perturbation will vary in relevance and explanatory power for a given complex trait. Still, there may be groupings of cell types and conditions that produce common regulatory responses and shared phenotypic effects. Therefore, it is important to consider multiple environmental factors that may be relevant for particular traits to clarify core responses versus truly context-specific ones. With respect to joint health, biomechanical forces and inflammation are particularly influential, which is why they were examined in this thesis. Additionally, several other environmental factors affect skeletal development and pathogenesis, such hypoxia and cell-cell signaling – both of which can be further explored using similar cell culture systems.

Hypoxia may be a particularly interesting environmental context to examine with respect to the skeletal system, as the synovial environment which contains cartilage, bone, synovial, and infiltrating immune cells can become hypoxic in response to inflammatory, metabolic, and mechanical perturbations. This hypoxia can lead to structural and pathogenic changes to joint tissues, similar to that observed in rheumatoid arthritis and OA [186–188], and may be a part of the disease process of these conditions. Studying how gene regulation is impacted by low oxygen environments and identifying which regulatory changes are unique to joint cell types as compared to other cell types, like heart, muscle, and neuronal cells, could improve our understanding of how transient and chronic hypoxia impacts and contributes to pathogenic states specifically in the joints.

In addition to the joints, several other regions within the skeletal system experience different environmental exposures. One critically important area is the bone marrow niche in long bones, where different skeletal and hematopoietic cells occupy the same space. In particular, mesenchymal stem cells (MSCs) and their derivatives interact with and modulate the development of immature immune cells. Such cell-cell signaling can impact aspects of health and disease. Establishing a model of this unique bone marrow signaling environment may be possible by co-culturing iPSC-derived MSCs with hematopoietic stem cells or other related immune cells. Given that the bone marrow is also a mechanically dynamic environment, it would be especially interesting to investigate how mechanical stimulation impacts the signaling capacity of iPSC-derived MSCs with immune cells, how such responses may contribute to disease initiation and progression, and how such environmental factors may mediate their phenotypic effects through interactions with underlying genetic variation.

4.2 Validation of eQTL findings

The eQTLs mapped in this work have the potential to help explain the causative mechanisms behind OA GWAS loci. Should the same genetic locus be simultaneously associated with disease risk and gene expression changes in chondrocytes, these gene expression changes may mediate the effect of genetic variation on susceptibility. However, in order to definitively establish this connection, we will need to take additional steps to validate our eQTL findings.

A first step will be to overlap eQTL variants with OA GWAS associated variants. In order to overcome issues with linkage disequilibrium between adjacent genetic variants in the human genome, applying methods to colocalize causal genetic variants across different traits will

be necessary for this work [189,190]. The differences in population between the participants in the vast majority of OA GWAS, which have been almost exclusively performed in European ancestry subjects, and the population of Yoruba individuals surveyed in my work, will potentially pose a problem for determining overlaps in causal loci. However, as genetic variants used to map both GWAS loci and eQTLs are common and therefore likely to be shared across populations [191], I do expect an effort to overlap European OA GWAS loci and eQTLs mapped in Yoruba iPSC-chondrocytes to be fruitful. Alternatively, determining an overlap of iPSC-chondrocyte eGenes and genes implicated in OA GWAS through examining genes that are located close to GWAS variants may be less affected by population differences but still yield interesting insight by serving as additional evidence to connect GWAS variants to genes they may impact. Further integration of the eQTL data with functional annotation data encompassing chromatin accessibility, transcription factor binding, and three-dimensional chromatin structure from cartilage could lend further evidence as to the functional relevance of eQTL variants. For example, a dynamic eQTL identified in chapter 3 affects expression of *SLC2A5*, a gene that is differentially expressed between and possesses gene regulatory elements that display differential chromatin accessibility between damaged and intact OA cartilage [156,167]. One potential reason for the ability to detect this eQTL in only perturbed conditions may be the lack of expression or gene regulatory activity of the gene in unperturbed conditions. Understanding potential mechanisms underlying eQTLs or dynamic eQTLs can aid in discovery of additional instances of genetic control of gene regulation in these cells.

Functional studies also can certify eQTLs as causally impacting gene expression. For example, CRISPR/Cas9 genome editing strategies to precisely introduce genetic changes at mapped eQTLs and observing their impacts on gene expression would be one way to establish

conclusively the direct connection between genetic variation and gene expression in chondrocytes [192].

Despite the drawbacks of many *in vivo* models of OA, validation of my findings in such models would be warranted to overcome some of the experimental blind spots of our system. One step to validate the ability of this iPSC-chondrocyte model to mimic both chondrogenic cells and diseased cells would involve studying patients with OA. One such study could involve collecting primary chondrocytes from patients who have OA and are undergoing hip or knee replacement. Gene expression and phenotypic measurements from these primary cells could then be compared to gene expression patterns in iPSC chondrocytes reprogrammed blood or skin cells collected from those same patients. This comparison would allow for an assessment of the ability of iPSC derived cells to model primary cells. A further comparison of the gene expression profiles of these cells following CTS or IL-1β treatment would allow us to anchor our observations of iPSC-chondrocytes with primary disease cells collected from the same individuals. As a way to overcome the degraded nature of primary late stage OA chondrocytes, differentiating primary MSCs collected from these patients into chondrocytes using differentiation protocols would allow us to generate more comparable cell types to match against iPSC-chondrocytes. Such a study would lend more credence to the claim that these iPSC models can be used to understand the gene expression behaviors of primary cells without the need for invasive sampling.

Additionally, animal models could be used to validate the disease implications of dysregulation of genes in an organismal context. Induced overexpression or knock down expression of eGenes with eSNPs that overlap OA GWAS associated variants in animal joints

and observing the impacts of these gene expression perturbations on joint health would help to connect the impact of these variants on gene regulation with their impact on phenotype.

4.3 Clinical implications of findings for understanding of joint health

Understanding how inter-individual differences influence joint health could identify new therapeutic targets to treat OA. The gene-by-environment interactions revealed by comparing the effects of eQTLs across conditions represent context-specific gene regulation that contribute to these inter-individual differences in susceptibility to this disease. The analysis of differential expression and differential gene regulation between the individuals in my studies reveals insights into the interplay between inflammatory and mechanical environments in OA, contributing to an ongoing discussion in the field about the relative importance of each factor in OA development.

4.3.1 Interplay between inflammatory and mechanical environments in OA

In chapter 3, I evaluated the gene expression responses of iPSC-chondrocytes to two different types of OA induction treatments, namely IL-1β and CTS. While both inflammatory and mechanical factors are known to play contributing roles in OA risk, a current topic of research within the OA field surrounds whether this disease is primarily the result of the impacts of one factor or the other [193–195]. My results from both differential expression and eQTL analyses from these two different forms of environmental perturbations highlight the interplay between the factors and the necessity to consider both when evaluating the disease.

First, by comparing the sets of differentially expressed genes in the two OA induction treatments, I identified shared and unique patterns of gene activation. Around a third of the union of differentially expressed genes in either comparison between perturbation and control are

differentially expressed in the same direction in both comparisons (2234 shared DE genes out of 6255 DE genes in either comparison). This shared set of DE genes is enriched for annotations corresponding to skeletal system development and extracellular matrix organization and amongst two publicly available OA-related DE gene sets, reinforcing their relevance to joint disease. Since the two models used in this work are both models of OA and are known to induce OA like phenotypes [68–71,77,83–87], these gene expression response patterns and enrichments are in line with expectations. However, the fact that they activate similar gene expression response programs further suggests that they may act through common pathways in producing OA phenotypes.

Interestingly, even looking to those genes that are specifically DE in one perturbation but not the other, I can observe evidence of the cross-talk between inflammation and mechanical stress. Genes specifically responding to IL-1β treatment are enriched for the "cellular response to mechanical stimulus" gene ontology term. One example is the mechanically activated ion channel *PIEZO1*, which is found to respond specifically to IL-1β treatment in my data. Prior studies of response of cartilage to inflammatory cytokines have found similar enhancements of *PIEZO1* expression on the RNA level and potential increased function of the resulting protein within inflamed cartilage tissue cultures [161]. *PIEZO1/2* are both highly expressed in OA cartilage compared to healthy cartilage and act particularly in response to injurious mechanical stresses [162,163], joint inflammation may also predispose joints to respond negatively to mechanical stimuli in a disease-promoting fashion. At the same time, the relationship between these two factors may not always be synchronistic. Indeed, moderate mechanical loading of cartilage may have a protective effect against the matrix-degrading phenotypes induced by cytokine administration [164].

In the eQTL analysis from the same study, I find several examples of CTS-specific dynamic eQTLs that are associated with changes in expression of genes that have been previously shown to be related to cellular responses to inflammation. For example, the gene *DUSP5* is known to regulate cellular inflammation, has reduced expression in OA cartilage compared to healthy cartilage, and has been shown to negatively regulate IL-1β induced inflammation in a rat chondrocyte cell culture model of OA [166]. As in the differential expression analysis, the interplay between different perturbations relevant to OA and their effects on gene regulation may present key insights into their role in contributing to OA disease progression.

Based on these observations from the current studies, it is difficult to disentangle the impacts of mechanical stress from those of inflammation and *vice versa*. Instead, use of a liability model encompassing the cumulative impacts of both genetic variation and multiple environmental factors may be more accurate when approaching the issue of defining osteoarthritis [196]. Combination treatments whereby chondrocytes are exposed simultaneously to both mechanical stress and inflammatory cytokines at different dosages could help refine the understanding of the shared pathways activated by the two treatments individually and the interplay between these two environmental factors in a dynamic human joint context. This type of study may also further reveal the interactions between the two treatments on the transcriptional level and their potential impacts on cartilage degradation in OA.

4.4 Skeletal traits and comparative biology

While human cell lines provide genetic backgrounds and cellular environments relevant for certain studies of human-specific gene regulation, the utility of animal models should not be

overlooked. This is especially true for our closest nonhuman primate (NHP) relatives, which can provide additional evolutionary context for understanding the basis for human trait variation and disease. While the general functions of the skeleton are conserved across primate species, the different behaviors and evolutionary histories of humans and NHPs have led to specialized skeletal features, such as the bicondylar angle of the femur that enables is bipedal locomotion in humans. Gene regulation may provide insight into how such morphological skeletal differences develop and evolve across species. Additionally, examining regulatory patterns in NHP skeletal cells may help to clarify aspects of human disease. For example, there are cases of rhesus macaques and chimpanzees developing conditions similar to those observed in human OA [132]. Although the prevalence of this degenerative joint disease in wild and captive NHPs is not well documented, evaluating how gene regulation may contribute to or protect against OA in NHPs will better inform our understanding of human OA.

Unfortunately, like human skeletal tissues, NHP skeletal tissues are difficult to access. An alternative way to investigate comparative gene regulation dynamics in skeletal cell types would be to use existing panels of matched human and chimpanzee iPSCs [197,198]. Studies using these primate iPSCs and derived cell types, including skeletal cells, have been fruitful in advancing our understanding of gene regulatory variation in NHPs and how these patterns compare to those in humans [197–200]. Extending the cell culture systems developed and characterized in this thesis to other NHP iPSCs would help to build a complete evolutionary picture of skeletal cell regulatory responses in the context of joint health and disease.

4.5 Future directions in dynamic skeletal cell gene regulation studies

While the work presented here contributes to and expands our understanding of context-specific gene regulation in the skeleton, there is still much to be discovered by exploring additional dimensions of developmental time and spatial location. In this work, growing cells in a two-dimensional monolayer promoted homogeneity in culture conditions and treatment responses – desirable qualities for transcriptional assays at a bulk resolution. However, recent advances in single-cell technologies alongside increasing numbers of protocols to culture three-dimensional cellular aggregates, including aggregates of iPSCs known as embryoid bodies, have enabled more complex and potentially more efficient experiments to survey dozens of developmental trajectories simultaneously [201]. The potential for such systems for dynamic gene regulation studies is substantial.

In particular for skeletal and related cell types, spheroid cultures of aggregated MSCs (mesenspheres) have been previously generated to enhance the differentiation potential and immunomodulatory capabilities of MSCs [202,203]. Mesenspheres of differing sizes and compositions can undergo directed differentiation into bone, cartilage, fat, and muscle cells [159,204,205]. Alternatively, it may be possible to promote spontaneous and undirected differentiation in mesenspheres. This spontaneous differentiation would generate heterogenous cellular aggregates (mesenchymal bodies) that have broad samplings of mesodermal cell types, including potential transient lineages that are currently impossible to obtain from primary sampling of adult skeletons either because they only exist at early developmental stages or at very small proportions in adult tissues. Additionally, these mesenchymal bodies could be paired with environmental treatments to further expand the range of regulatory dynamics that can be studied in mesodermal cell types. Although mesenchymal bodies would likely not recreate the

precise structure and context of *in vivo* mesodermal biology, their heterogeneity is a strength, and the potential for cross-cell signaling in this model may lead to further insights.

4.6 Concluding remarks

While much has been done to advance the knowledge of gene regulation in diverse cell types and tissues, we have only scratched the surface in terms of understanding the dynamic nature of how gene expression is controlled throughout time and space. My work in this dissertation contributes to this effort by sampling – with the help of iPSC differentiation – an area of the cellular differentiation space that has been difficult to access using primary tissues, namely chondrogenic cells. The observations between mechanical and inflammatory perturbations of chondrogenic cells in this work adds to current discussion about the relevance of these environmental factors in OA development. The field of genomics is in an exciting moment, with ever-increasing technological and methodological advances that are enabling researchers to survey the vast landscape of gene regulation. With each discovery, the field of Human Genetics continues to move closer and closer to better understanding the connections between human genotypes and phenotypes.

References

1. Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., *et al*. (2007). The Diploid Genome Sequence of an Individual Human. PLoS Biol. *5*, e254.

2. Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. Nat. Rev. Methods Primer *1*, 1–21.

3. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., *et al*. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

4. Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., *et al*. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.

5. Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., *et al*. (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. *50*, 956–967.

6. Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. *9*, 477–485.

7. Lappalainen, T., and Dermitzakis, E.T. (2010). Evolutionary history of regulatory variation in human populations. Hum. Mol. Genet. *19*, R197-203.

8. Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

9. Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Impact of Regulatory Variation from RNA to Protein. Science *347*, 664–667.

10. Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. TIG *24*, 408–415.

11. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., *et al*. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

12. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. PLOS Genet. *6*, e1000888.

13. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. Science *352*, 600–604.

14. Shi, H., Kichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. *99*, 139–153.

15. Cusanovich, D.A., Billstrand, C., Zhou, X., Chavarria, C., De Leon, S., Michelini, K., Pai, A.A., Ober, C., and Gilad, Y. (2012). The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. Hum. Mol. Genet. *21*, 2111–2123.

16. Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, C.A., Alazizi, A., Pazokitoroudi, A., Sankararaman, S., Wen, X., *et al*. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. eLife *10*, e67077.

17. The GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

18. Umans, B.D., Battle, A., and Gilad, Y. (2020). Where Are the Disease-Associated eQTLs? Trends Genet. *0*. Available at: https://www.cell.com/trends/genetics/abstract/S0168-9525(20)30209-2 [Accessed October 19, 2020].

19. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.-C., Freiman, A., Sams, A.J., Hebert, S., *et al*. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell *167*, 657-669.e21.

20. Strober, B.J., Elorbany, R., Rhodes, K., Krishnan, N., Tayeb, K., Battle, A., and Gilad, Y. (2019). Dynamic genetic regulation of gene expression during cellular differentiation. Science *364*, 1287–1290.

21. Knowles, D.A., Burrows, C.K., Blischak, J.D., Patterson, K.M., Serie, D.J., Norton, N., Ober, C., Pritchard, J.K., and Gilad, Y. (2018). Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. eLife *7*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6010343/ [Accessed December 1, 2019].

22. Lawrence, R.C., Felson, D.T., Helmick, C.G., Arnold, L.M., Choi, H., Deyo, R.A., Gabriel, S., Hirsch, R., Hochberg, M.C., Hunder, G.G., *et al*. (2008). Estimates of the Prevalence of Arthritis and Other Rheumatic Conditions in the United States, Part II. Arthritis Rheum. *58*, 26–35.

23. Lorenzo, P., Bayliss, M.T., and Heinegård, D. (2004). Altered patterns and synthesis of extracellular matrix macromolecules in early osteoarthritis. Matrix Biol. J. Int. Soc. Matrix Biol. *23*, 381–391.

24. Dreier, R. (2010). Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. Arthritis Res. Ther. *12*, 216.

25. Sharif, M., Whitehouse, A., Sharman, P., Perry, M., and Adams, M. (2004). Increased apoptosis in human osteoarthritic cartilage corresponds to reduced cell density and expression of caspase-3. Arthritis Rheum. *50*, 507–515.

26. Blanco, F.J., Guitian, R., Vázquez-Martul, E., de Toro, F.J., and Galdo, F. (1998). Osteoarthritis chondrocytes die by apoptosis. A possible pathway for osteoarthritis pathology. Arthritis Rheum. *41*, 284–289.

27. Li, H., Yang, H.H., Sun, Z.G., Tang, H.B., and Min, J.K. (2019). Whole-transcriptome sequencing of knee joint cartilage from osteoarthritis patients. Bone Jt. Res. *8*, 288–301.

28. Steinberg, J., Southam, L., Butterfield, N.C., Roumeliotis, T.I., Fontalis, A., Clark, M.J., Jayasuriya, R.L., Swift, D., Shah, K.M., Curry, K.F., *et al*. (2019). Decoding the genomic basis of osteoarthritis. bioRxiv, 835850 [Preprint] [cited 2019 Nov 12] Available from: https://doi.org/10.1101/835850.

29. Valdes, A.M., and Spector, T.D. (2008). The Contribution of Genes to Osteoarthritis. Rheum. Dis. Clin. N. Am. *34*, 581–603.

30. Reynard, L.N., and Loughlin, J. (2013). Insights from human genetic studies into the pathways involved in osteoarthritis. Nat. Rev. Rheumatol. *9*, 573–583.

31. Boer, C.G., Hatzikotoulas, K., Southam, L., Stefánsdóttir, L., Zhang, Y., Coutinho de Almeida, R., Wu, T.T., Zheng, J., Hartley, A., Teder-Laving, M., *et al*. (2021). Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell *184*, 4784-4818.e17.

32. Aubourg, G., Rice, S.J., Bruce-Wootton, P., and Loughlin, J. (2022). Genetics of osteoarthritis. Osteoarthritis Cartilage *30*, 636–649.

33. Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J., *et al*. (2019). Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank. Nat. Genet. *51*, 230–236.

34. Pregizer, S.K., Kiapour, A.M., Young, M., Chen, H., Schoor, M., Liu, Z., Cao, J., Rosen, V., and Capellini, T.D. (2018). Impact of broad regulatory regions on Gdf5 expression and function in knee development and susceptibility to osteoarthritis. Ann. Rheum. Dis. *77*, 450–450.

35. Richard, D., Liu, Z., Cao, J., Kiapour, A.M., Willen, J., Yarlagadda, S., Jagoda, E., Kolachalama, V.B., Sieker, J.T., Chang, G.H., *et al*. (2020). Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell *181*, 362-381.e28.

36. Vingård, E. (1994). Sport and the Development of Osteoarthrosis of the Hip. Sports Med. *18*, 1–3.

37. Vingård, E., Alfredsson, L., and Malchau, H. (1997). Osteoarthrosis of the hip in women and its relation to physical load at work and in the home. Ann. Rheum. Dis. *56*, 293–298.

38. Vingård, E., Alfredsson, L., and Malchau, H. (1998). Osteoarthrosis of the Hip in Women and Its Relationship to Physical Load from Sports Activities. Am. J. Sports Med. *26*, 78–82.

39. Buckwalter, J.A. (1995). Osteoarthritis and articular cartilage use, disuse, and abuse: experimental studies. J. Rheumatol. Suppl. *43*, 13–15.

40. Palmoski, M.J., Colyer, R.A., and Brandt, K.D. (1980). Joint motion in the absence of normal loading does not maintain normal articular cartilage. Arthritis Rheum. *23*, 325–334.

41. Palmoski, M., Perricone, E., and Brandt, K.D. (1979). Development and reversal of a proteoglycan aggregation defect in normal canine knee cartilage after immobilization. Arthritis Rheum. *22*, 508–517.

42. Lindhorst, E., Vail, T.P., Guilak, F., Wang, H., Setton, L.A., Vilim, V., and Kraus, V.B. (2000). Longitudinal characterization of synovial fluid biomarkers in the canine meniscectomy model of osteoarthritis. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *18*, 269–280.

43. Carlson, C.S., Guilak, F., Vail, T.P., Gardin, J.F., and Kraus, V.B. (2002). Synovial fluid biomarker levels predict articular cartilage damage following complete medial meniscectomy in the canine knee. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *20*, 92–100.

44. Altman, R.D., Tenenbaum, J., Latta, L., Riskin, W., Blanco, L.N., and Howell, D.S. (1984). Biomechanical and biochemical properties of dog cartilage in experimentally induced osteoarthritis. Ann. Rheum. Dis. *43*, 83–90.

45. McDevitt, C.A., and Muir, H. (1976). Biochemical changes in the cartilage of the knee in experimental and natural osteoarthritis in the dog. J. Bone Joint Surg. Br. *58*, 94–101.

46. Elliott, D.M., Guilak, F., Vail, T.P., Wang, J.Y., and Setton, L.A. (1999). Tensile properties of articular cartilage are altered by meniscectomy in a canine model of osteoarthritis. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *17*, 503–508.

47. Setton, L.A., Mow, V.C., Müller, F.J., Pita, J.C., and Howell, D.S. (1994). Mechanical properties of canine articular cartilage are significantly altered following transection of the anterior cruciate ligament. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *12*, 451–463.

48. Ratcliffe, A., Billingham, M.E., Saed-Nejad, F., Muir, H., and Hardingham, T.E. (1992). Increased release of matrix components from articular cartilage in experimental canine osteoarthritis. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *10*, 350–358.

49. Eyre, D.R., McDevitt, C.A., Billingham, M.E., and Muir, H. (1980). Biosynthesis of collagen and other matrix proteins by articular cartilage in experimental osteoarthrosis. Biochem. J. *188*, 823–837.

50. Brouwer, G.M., van Tol, A.W., Bergink, A.P., Belo, J.N., Bernsen, R.M.D., Reijman, M., Pols, H. a. P., and Bierma-Zeinstra, S.M.A. (2007). Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum. *56*, 1204–1211.

51. Sharma, L., Song, J., Felson, D.T., Cahue, S., Shamiyeh, E., and Dunlop, D.D. (2001). The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA *286*, 188–195.

52. Hadley, N.A., Brown, T.D., and Weinstein, S.L. (1990). The effects of contact pressure elevations and aseptic necrosis on the long-term outcome of congenital hip dislocation. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *8*, 504–513.

53. Maxian, T.A., Brown, T.D., and Weinstein, S.L. (1995). Chronic stress tolerance levels for human articular cartilage: two nonuniform contact models applied to long-term follow-up of CDH. J. Biomech. *28*, 159–166.

54. Anderson, D.D., Chubinskaya, S., Guilak, F., Martin, J.A., Oegema, T.R., Olson, S.A., and Buckwalter, J.A. (2011). POST-TRAUMATIC OSTEOARTHRITIS: IMPROVED UNDERSTANDING AND OPPORTUNITIES FOR EARLY INTERVENTION. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *29*, 802–809.

55. Lindberg, H., Roos, H., and Gärdsell, P. (1993). Prevalence of coxarthrosis in former soccer players. 286 players compared with matched controls. Acta Orthop. Scand. *64*, 165–167.

56. Anderson, J.J., and Felson, D.T. (1988). Factors associated with osteoarthritis of the knee in the first national Health and Nutrition Examination Survey (HANES I). Evidence for an association with overweight, race, and physical demands of work. Am. J. Epidemiol. *128*, 179–189.

57. Axmacher, B., and Lindberg, H. (1993). Coxarthrosis in farmers. Clin. Orthop., 82–86.

58. Croft, P., Cooper, C., Wickham, C., and Coggon, D. (1992). Osteoarthritis of the hip and occupational activity. Scand. J. Work. Environ. Health *18*, 59–63.

59. Heliövaara, M., Mäkelä, M., Impivaara, O., Knekt, P., Aromaa, A., and Sievers, K. (1993). Association of overweight, trauma and workload with coxarthrosis. A health survey of 7,217 persons. Acta Orthop. Scand. *64*, 513–518.

60. Lawrence, J.S. (1955). RHEUMATISM IN COAL MINERS: PART III: OCCUPATIONAL FACTORS. Br. J. Ind. Med. *12*, 249–261.

61. Partridge, R.E., and Duthie, J.J. (1968). Rheumatism in dockers and civil servants. A comparison of heavy manual and sedentary workers. Ann. Rheum. Dis. *27*, 559–568.

62. Takahashi, K., Kubo, T., Kobayashi, K., Imanishi, J., Takigawa, M., Arai, Y., and Hirasawa, Y. (1997). Hydrostatic pressure influences mRNA expression of transforming growth factor-beta 1 and heat shock protein 70 in chondrocyte-like cell line. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *15*, 150–158.

63. Takahashi, K., Kubo, T., Arai, Y., Kitajima, I., Takigawa, M., Imanishi, J., and Hirasawa, Y. (1998). Hydrostatic pressure induces expression of interleukin 6 and tumour necrosis factor α mRNAs in a chondrocyte-like cell line. Ann. Rheum. Dis. *57*, 231–236.

64. Mohtai, M., Gupta, M.K., Donlon, B., Ellison, B., Cooke, J., Gibbons, G., Schurman, D.J., and Smith, R.L. (1996). Expression of interleukin-6 in osteoarthritic chondrocytes and effects of fluid-induced shear on this expression in normal human chondrocytes in vitro. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *14*, 67–73.

65. Bougault, C., Paumier, A., Aubert-Foucher, E., and Mallein-Gerin, F. (2009). Investigating conversion of mechanical force into biochemical signaling in three-dimensional chondrocyte cultures. Nat. Protoc. *4*, 928–938.

66. Zhang, R.-K., Li, G.-W., Zeng, C., Lin, C.-X., Huang, L.-S., Huang, G.-X., Zhao, C., Feng, S.-Y., and Fang, H. (2018). Mechanical stress contributes to osteoarthritis development through the activation of transforming growth factor beta 1 (TGF-β1). Bone Jt. Res. *7*, 587–594.

67. Kawakita, K., Nishiyama, T., Fujishiro, T., Hayashi, S., Kanzaki, N., Hashimoto, S., Takebe, K., Iwasa, K., Sakata, S., Nishida, K., *et al.* (2012). Akt phosphorylation in human chondrocytes is regulated by p53R2 in response to mechanical stress. Osteoarthritis Cartilage *20*, 1603–1609.

68. Pichler, K., Herbert, V., Schmidt, B., Fischerauer, E.E., Leithner, A., and Weinberg, A.-M. (2013). Expression of matrix metalloproteinases in human growth plate chondrocytes is enhanced at high levels of mechanical loading: A possible explanation for overuse injuries in children. Bone Jt. J. *95-B*, 568–573.

69. Honda, K., Ohno, S., Tanimoto, K., Ijuin, C., Tanaka, N., Doi, T., Kato, Y., and Tanne, K. (2000). The effects of high magnitude cyclic tensile load on cartilage matrix metabolism in cultured chondrocytes. Eur. J. Cell Biol. *79*, 601–609.

70. Fujisawa, T., Hattori, T., Takahashi, K., Kuboki, T., Yamashita, A., and Takigawa, M. (1999). Cyclic Mechanical Stress Induces Extracellular Matrix Degradation in Cultured Chondrocytes via Gene Expression of Matrix Metalloproteinases and Interleukin-11.

71. Lin, Y.-Y., Tanaka, N., Ohkuma, S., Iwabuchi, Y., Tanne, Y., Kamiya, T., Kunimatsu, R., Huang, Y.-C., Yoshioka, M., Mitsuyoshi, T., *et al.* (2010). Applying an excessive

mechanical stress alters the effect of subchondral osteoblasts on chondrocytes in a co-culture system. Eur. J. Oral Sci. *118*, 151–158.

72. Farahat, M.N., Yanni, G., Poston, R., and Panayi, G.S. (1993). Cytokine expression in synovial membranes of patients with rheumatoid arthritis and osteoarthritis. Ann. Rheum. Dis. *52*, 870–875.

73. Smith, M.D., Triantafillou, S., Parker, A., Youssef, P.P., and Coleman, M. (1997). Synovial membrane inflammation and cytokine production in patients with early osteoarthritis. J. Rheumatol. *24*, 365–371.

74. Wang, Y., Teichtahl, A.J., Pelletier, J.-P., Abram, F., Wluka, A.E., Hussain, S.M., Martel-Pelletier, J., and Cicuttini, F.M. (2019). Knee effusion volume assessed by magnetic resonance imaging and progression of knee osteoarthritis: data from the Osteoarthritis Initiative. Rheumatol. Oxf. Engl. *58*, 246–253.

75. Felson, D.T., Niu, J., Neogi, T., Goggins, J., Nevitt, M.C., Roemer, F., Torner, J., Lewis, C.E., Guermazi, A., and MOST Investigators Group (2016). Synovitis and the risk of knee osteoarthritis: the MOST Study. Osteoarthritis Cartilage *24*, 458–464.

76. Fang, T., Zhou, X., Jin, M., Nie, J., and Li, Xi. (2021). Molecular mechanisms of mechanical load-induced osteoarthritis. Int. Orthop. *45*, 1125–1136.

77. Stöve, J., Huch, K., Günther, K.P., and Scharf, H.P. (2000). Interleukin-1beta induces different gene expression of stromelysin, aggrecan and tumor-necrosis-factor-stimulated gene 6 in human osteoarthritic chondrocytes in vitro. Pathobiol. J. Immunopathol. Mol. Cell. Biol. *68*, 144–149.

78. van de Loo, F.A., Joosten, L.A., van Lent, P.L., Arntz, O.J., and van den Berg, W.B. (1995). Role of interleukin-1, tumor necrosis factor alpha, and interleukin-6 in cartilage proteoglycan metabolism and destruction. Effect of in situ blocking in murine antigen- and zymosan-induced arthritis. Arthritis Rheum. *38*, 164–172.

79. Attur, M.G., Patel, I.R., Patel, R.N., Abramson, S.B., and Amin, A.R. (1998). Autocrine production of IL-1 beta by human osteoarthritis-affected cartilage and differential regulation of endogenous nitric oxide, IL-6, prostaglandin E2, and IL-8. Proc. Assoc. Am. Physicians *110*, 65–72.

80. Martel-Pelletier, J., Battista, J.D., and Lajeunesse, D. (1999). Biochemical Factors in Joint Articular Tissue Degradation in Osteoarthritis. In Osteoarthritis: Clinical and Experimental Aspects, J.-Y. Reginster, J.-P. Pelletier, J. Martel-Pelletier, Y. Henrotin, and L. Crasborn, eds. (Berlin, Heidelberg: Springer), pp. 156–187. Available at: https://doi.org/10.1007/978-3-642-60026-5_9 [Accessed June 22, 2022].

81. Visser, A.W., Ioan-Facsinay, A., de Mutsert, R., Widya, R.L., Loef, M., de Roos, A., le Cessie, S., den Heijer, M., Rosendaal, F.R., and Kloppenburg, M. (2014). Adiposity and

hand osteoarthritis: the Netherlands Epidemiology of Obesity study. Arthritis Res. Ther. *16*, R19.

82. Harasymowicz, N.S., Clement, N.D., Azfer, A., Burnett, R., Salter, D.M., and Simpson, A.H.W.R. (2017). Regional Differences Between Perisynovial and Infrapatellar Adipose Tissue Depots and Their Response to Class II and Class III Obesity in Patients With Osteoarthritis. Arthritis Rheumatol. Hoboken NJ *69*, 1396–1406.

83. Jia, Y., He, W., Zhang, H., He, L., Wang, Y., Zhang, T., Peng, J., Sun, P., and Qian, Y. (2020). Morusin Ameliorates IL-1β-Induced Chondrocyte Inflammation and Osteoarthritis via NF-κB Signal Pathway . Drug Des. Devel. Ther. *14*, 1227–1240.

84. Pan, T., Chen, R., Wu, D., Cai, N., Shi, X., Li, B., and Pan, J. (2017). Alpha-Mangostin suppresses interleukin-1β-induced apoptosis in rat chondrocytes by inhibiting the NF-κB signaling pathway and delays the progression of osteoarthritis in a rat model. Int. Immunopharmacol. *52*, 156–162.

85. Wan, Z.-H., and Zhao, Q. (2017). Gypenoside inhibits interleukin-1β-induced inflammatory response in human osteoarthritis chondrocytes. J. Biochem. Mol. Toxicol. *31*.

86. Zhuang, Z., Ye, G., and Huang, B. (2017). Kaempferol Alleviates the Interleukin-1β-Induced Inflammation in Rat Osteoarthritis Chondrocytes via Suppression of NF-κB. Med. Sci. Monit. Int. Med. J. Exp. Clin. Res. *23*, 3925–3931.

87. Shakibaei, M., Schulze-Tanzil, G., John, T., and Mobasheri, A. (2005). Curcumin protects human chondrocytes from IL-l1beta-induced inhibition of collagen type II and beta1-integrin expression and activation of caspase-3: an immunomorphological study. Ann. Anat. Anat. Anz. Off. Organ Anat. Ges. *187*, 487–497.

88. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

89. Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

90. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al*. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917–1920.

91. Burrows, C.K., Banovich, N.E., Pavlovic, B.J., Patterson, K., Romero, I.G., Pritchard, J.K., and Gilad, Y. (2016). Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. PLOS Genet. *12*, e1005793.

92. Nejadnik, H., Diecke, S., Lenkov, O.D., Chapelin, F., Donig, J., Tong, X., Derugin, N., Chan, R.C.F., Gaur, A., Yang, F., *et al*. (2015). Improved Approach for Chondrogenic

Differentiation of Human Induced Pluripotent Stem Cells. Stem Cell Rev. Rep. *11*, 242–253.

93. Cuomo, A.S.E., Seaton, D.D., McCarthy, D.J., Martinez, I., Bonder, M.J., Garcia-Bernardo, J., Amatya, S., Madrigal, P., Isaacson, A., Buettner, F., *et al*. (2020). Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Nat. Commun. *11*.

94. Neavin, D., Nguyen, Q., Daniszewski, M.S., Liang, H.H., Chiu, H.S., Wee, Y.K., Senabouth, A., Lukowski, S.W., Crombie, D.E., Lidgerwood, G.E., *et al*. (2021). Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. Genome Biol. *22*, 76.

95. DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., Huang, H., Biggs, W., Sandoval, E., D'Antonio, M., *et al*. (2017). Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. Cell Stem Cell *20*, 533-546.e7.

96. Banovich, N.E., Li, Y.I., Raj, A., Ward, M.C., Greenside, P., Calderon, D., Tung, P.Y., Burnett, J.E., Myrthil, M., Thomas, S.M., *et al*. (2018). Impact of regulatory variation across human iPSCs and differentiated cells. Genome Res. *28*, 122–131.

97. Panopoulos, A.D., D'Antonio, M., Benaglio, P., Williams, R., Hashem, S.I., Schuldt, B.M., DeBoever, C., Arias, A.D., Garcia, M., Nelson, B.C., *et al*. (2017). iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. Stem Cell Rep. *8*, 1086–1100.

98. Aigner, T., and Schmitz, N. (2011). 173 - Pathogenesis and pathology of osteoarthritis.

99. Kloppenburg, M., and Berenbaum, F. (2020). Osteoarthritis year in review 2019: epidemiology and therapy. Osteoarthritis Cartilage *28*, 242–248.

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

101. Kariuki, S.N., Maranville, J.C., Baxter, S.S., Jeong, C., Nakagome, S., Hrusch, C.L., Witonsky, D.B., Sperling, A.I., and Rienzo, A.D. (2016). Mapping Variation in Cellular and Transcriptional Response to 1,25-Dihydroxyvitamin D3 in Peripheral Blood Mononuclear Cells. PLOS ONE *11*, e0159779.

102. Çalışkan, M., Baker, S.W., Gilad, Y., and Ober, C. (2015). Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. PLOS Genet. *11*, e1005111.

103. Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. *109*, 1204–1209.

104. Manry, J., Nédélec, Y., Fava, V.M., Cobat, A., Orlova, M., Thuc, N.V., Thai, V.H., Laval, G., Barreiro, L.B., and Schurr, E. (2017). Deciphering the genetic control of gene expression following Mycobacterium leprae antigen stimulation. PLOS Genet. *13*, e1006952.

105. Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., *et al*. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat. Commun. *8*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5559603/ [Accessed December 1, 2019].

106. Guilak, F. (2011). Biomechanical factors in osteoarthritis. Best Pract. Res. Clin. Rheumatol. *25*, 815–823.

107. Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F.C., Krause, D.S., Deans, R.J., Keating, A., Prockop, D.J., and Horwitz, E.M. (2006). Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy *8*, 315–317.

108. Tetsunaga, T., Nishida, K., Furumatsu, T., Naruse, K., Hirohata, S., Yoshida, A., Saito, T., and Ozaki, T. (2011). Regulation of mechanical stress-induced MMP-13 and ADAMTS-5 expression by RUNX-2 transcriptional factor in SW1353 chondrocyte-like cells. Osteoarthritis Cartilage *19*, 222–232.

109. Khetarpal, U., Robertson, N.G., Yoo, T.J., and Morton, C.C. (1994). Expression and localization of COL2A1 mRNA and type II collagen in human fetal cochlea. Hear. Res. *79*, 59–73.

110. COL2A1 protein expression summary - The Human Protein Atlas Available at: https://www.proteinatlas.org/ENSG00000139219-COL2A1 [Accessed November 10, 2021].

111. Li, Y., Lacerda, D.A., Warman, M.L., Beier, D.R., Yoshioka, H., Ninomiya, Y., Oxford, J.T., Morris, N.P., Andrikopoulos, K., Ramirez, F., *et al*. (1995). A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. Cell *80*, 423–430.

112. Dey, K.K., Hsiao, C.J., and Stephens, M. (2017). Visualizing the structure of RNA-seq expression data using grade of membership models. PLOS Genet. *13*, e1006599.

113. Liu, C.-F., and Lefebvre, V. (2015). The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. Nucleic Acids Res. *43*, 8183–8203.

114. Luo, Y., Sinkeviciute, D., He, Y., Karsdal, M., Henrotin, Y., Mobasheri, A., Önnerfjord, P., and Bay-Jensen, A. (2017). The minor collagens in articular cartilage. Protein Cell *8*, 560–572.

115. Yi, S.W., Kim, H.J., Oh, H.J., Shin, H., Lee, J.S., Park, J.S., and Park, K.-H. (2018). Gene expression profiling of chondrogenic differentiation by dexamethasone-conjugated polyethyleneimine with SOX trio genes in stem cells. Stem Cell Res. Ther. *9*, 341.

116. Rahmati, M., Nalesso, G., Mobasheri, A., and Mozafari, M. (2017). Aging and osteoarthritis: Central role of the extracellular matrix. Ageing Res. Rev. *40*, 20–30.

117. Wang, W., Rigueur, D., and Lyons, K.M. (2014). TGFβ Signaling in Cartilage Development and Maintenance. Birth Defects Res. Part C Embryo Today Rev. *102*, 37–51.

118. Marshall, K.W., Zhang, H., Yager, T.D., Nossova, N., Dempsey, A., Zheng, R., Han, M., Tang, H., Chao, S., and Liew, C.C. (2005). Blood-based biomarkers for detecting mild osteoarthritis in the human knee. Osteoarthritis Cartilage *13*, 861–871.

119. Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., *et al*. (2016). A survey of best practices for RNA-seq data analysis. Genome Biol. *17*, 13.

120. Chubinskaya, S., Kuettner, K.E., and Cole, A.A. (1999). Expression of matrix metalloproteinases in normal and damaged articular cartilage from human knee and ankle joints. Lab. Investig. J. Tech. Methods Pathol. *79*, 1669–1677.

121. Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

122. Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. *50*, 424–431.

123. Ward, M.C., Banovich, N.E., Sarkar, A., Stephens, M., and Gilad, Y. (2021). Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. eLife *10*, e57345.

124. Yoon, H.H., Bhang, S.H., Shin, J.-Y., Shin, J., and Kim, B.-S. (2012). Enhanced Cartilage Formation via Three-Dimensional Cell Engineering of Human Adipose-Derived Stem Cells. Tissue Eng. Part A *18*, 1949–1956.

125. Gurdasani, D., Barroso, I., Zeggini, E., and Sandhu, M.S. (2019). Genomics of disease risk in globally diverse populations. Nat. Rev. Genet. *20*, 520–535.

126. Shi, H., Burch, K.S., Johnson, R., Freund, M.K., Kichaev, G., Mancuso, N., Manuel, A.M., Dong, N., and Pasaniuc, B. (2020). Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data. Am. J. Hum. Genet. *106*, 805–817.

127. Mogil, L.S., Andaleon, A., Badalamenti, A., Dickinson, S.P., Guo, X., Rotter, J.I., Johnson, W.C., Im, H.K., Liu, Y., and Wheeler, H.E. (2018). Genetic architecture of gene expression traits across diverse populations. PLOS Genet. *14*, e1007586.

128. Takano-Yamamoto, T., Soma, S., Nakagawa, K., Kobayashi, Y., Kawakami, M., and Sakuda, M. (1991). Comparison of the effects of hydrostatic compressive force on glycosaminoglycan synthesis and proliferation in rabbit chondrocytes from mandibular condylar cartilage, nasal septum, and spheno-occipital synchondrosis in vitro. Am. J. Orthod. Dentofac. Orthop. Off. Publ. Am. Assoc. Orthod. Its Const. Soc. Am. Board Orthod. *99*, 448–455.

129. Maranville, J.C., Luca, F., Richards, A.L., Wen, X., Witonsky, D.B., Baxter, S., Stephens, M., and Di Rienzo, A. (2011). Interactions between Glucocorticoid Treatment and Cis-Regulatory Polymorphisms Contribute to Cellular Response Phenotypes. PLoS Genet. *7*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3131293/ [Accessed September 7, 2020].

130. Gallego Romero, I., Pavlovic, B.J., Hernando-Herraez, I., Zhou, X., Ward, M.C., Banovich, N.E., Kagan, C.L., Burnett, J.E., Huang, C.H., Mitrano, A., *et al*. (2015). A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. eLife *4*, e07103.

131. Jurmain, R. (2000). Degenerative joint disease in African great apes: an evolutionary perspective. J. Hum. Evol. *39*, 185–203.

132. Videan, E.N., Lammey, M.L., and Lee, D.R. (2011). Diagnosis and Treatment of Degenerative Joint Disease in a Captive Male Chimpanzee (Pan troglodytes). J. Am. Assoc. Lab. Anim. Sci. JAALAS *50*, 263–266.

133. Lowenstine, L.J., McManamon, R., and Terio, K.A. (2016). Comparative Pathology of Aging Great Apes: Bonobos, Chimpanzees, Gorillas, and Orangutans. Vet. Pathol. *53*, 250–276.

134. Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., *et al*. (2003). The International HapMap Project. Nature *426*, 789–796.

135. Goldstein, D.J., and Horobin, R.W. (1974). Surface staining of cartilage by Alcian Blue, with reference to the role of microscopic dye aggregates in histological staining. Histochem. J. *6*, 175–184.

136. Rao, X., Huang, X., Zhou, Z., and Lin, X. (2013). An improvement of the 2^(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat. Bioinforma. Biomath. *3*, 71–85.

137. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., *et al*. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.

138. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

139. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., *et al*. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94.

140. Clarke, L., Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., Toneva, I., Vaughan, B., Preuss, D., Leinonen, R., Shumway, M., *et al*. (2012). The 1000 Genomes Project: data management and community access. Nat. Methods *9*, 459–462.

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

142. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

143. Stephens, M. (2017). False discovery rates: a new deal. Biostat. Oxf. Engl. *18*, 275–294.

144. Sarkar, A., and Stephens, M. (2020). Separating measurement and expression models clarifies confusion in single cell RNA-seq analysis. bioRxiv, 2020.04.07.030007.

145. MacParland, S.A., Liu, J.C., Ma, X.-Z., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J., Khuu, N., Echeverri, J., Linares, I., *et al*. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat. Commun. *9*, 4383.

146. Wu, C.-L., Dicks, A., Steward, N., Tang, R., Katz, D.B., Choi, Y.-R., and Guilak, F. (2021). Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis. Nat. Commun. *12*, 362.

147. Chou, C.-H., Jain, V., Gibson, J., Attarian, D.E., Haraden, C.A., Yohn, C.B., Laberge, R.-M., Gregory, S., and Kraus, V.B. (2020). Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis. Sci. Rep. *10*, 10868.

148. Ji, Q., Zheng, Y., Zhang, G., Hu, Y., Fan, X., Hou, Y., Wen, L., Li, L., Xu, Y., Wang, Y., *et al*. (2018). Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. Ann. Rheum. Dis., annrheumdis-2017-212863.

149. Carbonetto, P., Luo, K., Dey, K., Hsiao, J., and Stephens, M. (2021). fastTopics: fast algorithms for fitting topic models and non-negative matrix factorizations to count data. R package version 0.4-11. Available at: https://github.com/stephenslab/fastTopics.

150. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108–e108.

151. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

152. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics *17*, 483.

153. The Burden of Musculoskeletal Diseases in the United States, Fourth Edition | BMUS: The Burden of Musculoskeletal Diseases in the United States Available at: https://www.boneandjointburden.org/fourth-edition [Accessed July 1, 2022].

154. Loeser, R.F., Goldring, S.R., Scanzello, C.R., and Goldring, M.B. (2012). Osteoarthritis: a disease of the joint as an organ. Arthritis Rheum. *64*, 1697–1707.

155. Soul, J., Dunn, S.L., Anand, S., Serracino-Inglott, F., Schwartz, J.-M., Boot-Handford, R.P., and Hardingham, T.E. (2018). Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage. Ann. Rheum. Dis. *77*, 423–423.

156. Dunn, S.L., Soul, J., Anand, S., Schwartz, J.-M., Boot-Handford, R.P., and Hardingham, T.E. (2016). Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses. Osteoarthritis Cartilage *24*, 1431–1440.

157. Sun, W. (2012). A Statistical Framework for eQTL Mapping Using RNA-seq Data. Biometrics *68*, 1–11.

158. Ma, B., Leijten, J.C.H., Wu, L., Kip, M., van Blitterswijk, C.A., Post, J.N., and Karperien, M. (2013). Gene expression profiling of dedifferentiated human articular chondrocytes in monolayer culture. Osteoarthritis Cartilage *21*, 599–603.

159. Lu, T.-J., Chiu, F.-Y., Chiu, H.-Y., Chang, M.-C., and Hung, S.-C. (2017). Chondrogenic Differentiation of Mesenchymal Stem Cells in Three-Dimensional Chitosan Film Culture. Cell Transplant. *26*, 417–427.

160. Retting, K.N., Song, B., Yoon, B.S., and Lyons, K.M. (2009). BMP canonical Smad signaling through Smad1 and Smad5 is required for endochondral bone formation. Dev. Camb. Engl. *136*, 1093–1104.

161. Lee, W., Nims, R.J., Savadipour, A., Zhang, Q., Leddy, H.A., Liu, F., McNulty, A.L., Chen, Y., Guilak, F., and Liedtke, W.B. (2021). Inflammatory signaling sensitizes Piezo1 mechanotransduction in articular chondrocytes as a pathogenic feed-forward mechanism in osteoarthritis. Proc. Natl. Acad. Sci. U. S. A. *118*, e2001611118.

162. Steinecker-Frohnwieser, B., Kullich, W., Kratschmann, C., Cezanne, M., Toegel, S., and Weigl, L. (2020). Activation of the mechanosensitive ion channel PIEZO1/2 by YODA1 modulates cellular functions of human oa chondrocytes. Osteoarthritis Cartilage *28*, S101.

163. Lee, W., Leddy, H.A., Chen, Y., Lee, S.H., Zelenski, N.A., McNulty, A.L., Wu, J., Beicker, K.N., Coles, J., Zauscher, S., *et al*. (2014). Synergy between Piezo1 and Piezo2 channels confers high-strain mechanosensitivity to articular cartilage. Proc. Natl. Acad. Sci. U. S. A. *111*, E5114-5122.

164. Torzilli, P.A., Bhargava, M., and Chen, C.T. (2011). Mechanical Loading of Articular Cartilage Reduces IL-1-Induced Enzyme Expression. Cartilage *2*, 364–373.

165. Stevens, A.L., Wishnok, J.S., White, F.M., Grodzinsky, A.J., and Tannenbaum, S.R. (2009). Mechanical Injury and Cytokines Cause Loss of Cartilage Integrity and Upregulate Proteins Associated with Catabolism, Immunity, Inflammation, and Repair *. Mol. Cell. Proteomics *8*, 1475–1489.

166. Wu, Z., Xu, L., He, Y., Xu, K., Chen, Z., Moqbel, S.A.A., Ma, C., Jiang, L., Ran, J., Wu, L., *et al*. (2020). DUSP5 suppresses interleukin-1β-induced chondrocyte inflammation and ameliorates osteoarthritis in rats. Aging *12*, 26029–26046.

167. Liu, Y., Chang, J.-C., Hon, C.-C., Fukui, N., Tanaka, N., Zhang, Z., Lee, M.T.M., and Minoda, A. (2018). Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. Sci. Rep. *8*, 15499.

168. Yeowell, H.N., and Steinmann, B. (1993). PLOD1-Related Kyphoscoliotic Ehlers-Danlos Syndrome. In GeneReviews®, M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, and A. Amemiya, eds. (Seattle (WA): University of Washington, Seattle). Available at: http://www.ncbi.nlm.nih.gov/books/NBK1462/ [Accessed July 20, 2022].

169. Eyre, D., Shao, P., Weis, M.A., and Steinmann, B. (2002). The kyphoscoliotic type of Ehlers-Danlos syndrome (type VI): differential effects on the hydroxylation of lysine in collagens I and II revealed by analysis of cross-linked telopeptides from urine. Mol. Genet. Metab. *76*, 211–216.

170. Steinberg, J., Southam, L., Roumeliotis, T.I., Clark, M.J., Jayasuriya, R.L., Swift, D., Shah, K.M., Butterfield, N.C., Brooks, R.A., McCaskie, A.W., *et al*. (2021). A molecular quantitative trait locus map for osteoarthritis. Nat. Commun. *12*, 1309.

171. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063.

172. Hung, A., Housman, G., Briscoe, E.A., Cuevas, C., and Gilad, Y. (2021). Characterizing gene expression responses to biomechanical strain in an in vitro model of osteoarthritis Available at: https://www.biorxiv.org/content/10.1101/2021.02.22.432314v1 [Accessed August 23, 2021].

173. Lee, H.-S., Lee, C.-H., Tsai, H.-C., and Salter, D.M. (2009). Inhibition of cyclooxygenase 2 expression by diallyl sulfide on joint inflammation induced by urate crystal and IL-1β. Osteoarthritis Cartilage *17*, 91–99.

174. Kim, J., Xu, M., Xo, R., Mates, A., Wilson, G.L., Pearsall, A.W., and Grishko, V. (2010). Mitochondrial DNA damage is involved in apoptosis caused by pro-inflammatory cytokines in human OA chondrocytes. Osteoarthritis Cartilage *18*, 424–432.

175. Zhou, P.-H., Liu, S.-Q., and Peng, H. (2008). The effect of hyaluronic acid on IL-1β-induced chondrocyte apoptosis in a rat model of osteoarthritis. J. Orthop. Res. *26*, 1643–1648.

176. Elshaier, A.M., Hakimiyan, A.A., Rappoport, L., Rueger, D.C., and Chubinskaya, S. (2009). EFFECT OF INTERLEUKIN-1β ON OSTEOGENIC PROTEIN-1 INDUCED SIGNALING IN ADULT HUMAN ARTICULAR CHONDROCYTES. Arthritis Rheum. *60*, 143–154.

177. Aida, Y., Maeno, M., Suzuki, N., Shiratsuchi, H., Motohashi, M., and Matsumura, H. (2005). The effect of IL-1β on the expression of matrix metalloproteinases and tissue inhibitors of matrix metalloproteinases in human chondrocytes. Life Sci. *77*, 3210–3221.

178. Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

179. Urbut, S.M., Wang, G., Carbonetto, P., and Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat. Genet. *51*, 187–195.

180. Zhabotynsky, V., Huang, L., Little, P., Hu, Y.-J., Pardo-Manuel de Villena, F., Zou, F., and Sun, W. (2022). eQTL mapping using allele-specific count data is computationally feasible, powerful, and provides individual-specific estimates of genetic effects. PLoS Genet. *18*, e1010076.

181. Davis, J.R., Fresard, L., Knowles, D.A., Pala, M., Bustamante, C.D., Battle, A., and Montgomery, S.B. (2016). An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. Am. J. Hum. Genet. *98*, 216–224.

182. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

183. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. *100*, 9440–9445.

184. Kuyinu, E.L., Narayanan, G., Nair, L.S., and Laurencin, C.T. (2016). Animal models of osteoarthritis: classification, update, and measurement of outcomes. J. Orthop. Surg. *11*, 19.

185. Tse, H.M., Gardner, G., Dominguez-Bendala, J., and Fraker, C.A. (2021). The Importance of Proper Oxygenation in 3D Culture. Front. Bioeng. Biotechnol. *9*, 634403.

186. Akhavani, M.A., Madden, L., Buysschaert, I., Sivakumar, B., Kang, N., and Paleolog, E.M. (2009). Hypoxia upregulates angiogenesis and synovial cell migration in rheumatoid arthritis. Arthritis Res. Ther. *11*, R64.

187. Quiñonez-Flores, C.M., González-Chávez, S.A., and Pacheco-Tena, C. (2016). Hypoxia and its implications in rheumatoid arthritis. J. Biomed. Sci. *23*, 62.

188. Yang, S., Kim, J., Ryu, J.-H., Oh, H., Chun, C.-H., Kim, B.J., Min, B.H., and Chun, J.-S. (2010). Hypoxia-inducible factor-2α is a catabolic regulator of osteoarthritic cartilage destruction. Nat. Med. *16*, 687–693.

189. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. PLOS Genet. *17*, e1009440.

190. Zou, Y., Carbonetto, P., Wang, G., and Stephens, M. (2022). Fine-mapping from summary data with the "Sum of Single Effects" model. PLoS Genet. *18*, e1010299.

191. Zhang, Z., Luo, K., Zou, Z., Qiu, M., Tian, J., Sieh, L., Shi, H., Zou, Y., Wang, G., Morrison, J., *et al*. (2020). Genetic analyses support the contribution of mRNA N6-methyladenosine (m6A) modification to human disease heritability. Nat. Genet. *52*, 939–949.

192. Brandt, M., Gokden, A., Ziosi, M., and Lappalainen, T. (2020). A polyclonal allelic expression assay for detecting regulatory effects of transcript variants. Genome Med. *12*, 79.

193. Cicuttini, F.M., and Wluka, A.E. (2014). Is OA a mechanical or systemic disease? Nat. Rev. Rheumatol. *10*, 515–516.

194. Visser, A.W., Mutsert, R. de, Cessie, S. le, Heijer, M. den, Rosendaal, F.R., and Kloppenburg, M. (2015). The relative contribution of mechanical stress and systemic processes in different types of osteoarthritis: the NEO study. Ann. Rheum. Dis. *74*, 1842–1847.

195. de Rezende, M.U., and de Campos, G.C. (2013). Is osteoarthritis a mechanical or inflammatory disease? Rev. Bras. Ortop. *48*, 471–474.

196. Benchek, P.H., and Morris, N.J. (2013). How meaningful are heritability estimates of liability? Hum. Genet. *132*, 10.1007/s00439-013-1334-z.

197. Pavlovic, B.J., Blake, L.E., Roux, J., Chavarria, C., and Gilad, Y. (2018). A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. Sci. Rep. *8*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193013/ [Accessed November 21, 2019].

198. Ward, M.C., and Gilad, Y. (2019). A generally conserved response to hypoxia in iPSC-derived cardiomyocytes from humans and chimpanzees. eLife *8*, e42374.

199. Housman, G., Briscoe, E., and Gilad, Y. (2022). Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model. PLOS Genet. *18*, e1010073.

200. Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. Cell *163*, 68–83.

201. Rhodes, K., Barr, K.A., Popp, J.M., Strober, B.J., Battle, A., and Gilad, Y. (2022). Human embryoid bodies as a novel system for genomic studies of functionally diverse cell types. eLife *11*, e71361.

202. Isern, J., Martín-Antonio, B., Ghazanfari, R., Martín, A.M., López, J.A., del Toro, R., Sánchez-Aguilera, A., Arranz, L., Martín-Pérez, D., Suárez-Lledó, M., *et al*. (2013). Self-renewing human bone marrow mesenspheres promote hematopoietic stem cell expansion. Cell Rep. *3*, 1714–1724.

203. Bartosh, T.J., Ylöstalo, J.H., Mohammadipoor, A., Bazhanov, N., Coble, K., Claypool, K., Lee, R.H., Choi, H., and Prockop, D.J. (2010). Aggregation of human mesenchymal stromal cells (MSCs) into 3D spheroids enhances their antiinflammatory properties. Proc. Natl. Acad. Sci. U. S. A. *107*, 13724–13729.

204. Wang, W., Itaka, K., Ohba, S., Nishiyama, N., Chung, U., Yamasaki, Y., and Kataoka, K. (2009). 3D spheroid culture system on micropatterned substrates for improved differentiation efficiency of multipotent mesenchymal stem cells. Biomaterials *30*, 2705–2715.

205. Lee, J.H., Han, Y.-S., and Lee, S.H. (2016). Long-Duration Three-Dimensional Spheroid Culture Promotes Angiogenic Activities of Adipose-Derived Mesenchymal Stem Cells. Biomol. Ther. *24*, 260–267.