RESEARCH ARTICLE

REVISED Characterizing gene expression in an *in vitro* biomechanical strain model of joint health [version 2; peer review: 1 approved, 1 not approved]

Anthony Hung[1], Genevieve Housman[2], Emilie A. Briscoe[2], Claudia Cuevas[2], Yoav Gilad[2,3]

[1]Medical Scientist Training Program, University of Chicago, Chicago, IL, 60637, USA
[2]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, 60637, USA
[3]Department of Human Genetics, University of Chicago, Chicago, IL, 60637, USA

Abstract

Background: 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.

Methods: 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.

Results: 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.

Conclusions: Expanding this system to include iPSC-chondrocytes from a larger number of individuals will allow us to characterize and better understand gene-by-environment interactions related to joint health.

Keywords
Genomics, joint health, biomechanical stress, iPSC-derived cells, gene regulation

Open Peer Review

Approval Status ✔ ✘

	1	2
version 2 (revision) 27 Sep 2022	✔ view	✘ view
version 1 10 Mar 2022	? view	✘ view

1. **Alex Pollen** (ID), University of California San Francisco, San Francisco, USA
   Jenelle Wallace, University of California San Francisco, San Francisco, USA

2. **Shireen Lamande** (ID), Murdoch Children's Research Institute, Parkville, Australia
   The University of Melbourne, Parkville, Australia

Any reports and responses or comments on the article can be found at the end of the article.
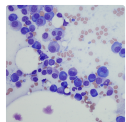
This article is included in the Cell & Molecular Biology gateway.

This article is included in the Genomics and Genetics gateway.

Corresponding author: Yoav Gilad (gilad@uchicago.edu)

> **REVISED** **Amendments from Version 1**
>
> In the revised version of the manuscript, we make minor additions and changes to address the feedback and suggestions of two reviewers. We add additional extended data figures that reflect further morphologic and gene expression characterization of iPSC-chondrocytes. We also expand the Discussion section to address sources of gene expression variation in the data and their implications for quantitative genetic studies of gene regulation in this system.
>
> **Any further responses from the reviewers can be found at the end of the article**

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.[1] OA is a major cause of disability in older adults and impacts approximately 300 million people worldwide.[2] 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.[3] 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.[4,5] 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.[6] This 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,[7,8] 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.[9] Thus, numerous studies have begun identifying dynamic regulatory expression quantitative trait loci (eQTLs) in various cell types and contexts, including drug-induced cardiotoxicity,[10] cardiomyocyte differentiation,[11] vitamin D exposure,[12] and response to infection.[13–17] 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.[18–22] Further, biomechanical factors may impact gene expression regulation in joint tissues and may interact with genetic factors to impact risk for joint diseases.[23] 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.[24–27] 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 between three individuals. 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.

Methods

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).[28] These LCLs were originally derived from individuals collected as part of the HapMap project.[29] Undifferentiated iPSCs were cultured on Matrigel-coated plates (Corning 356230) in Essential 8 (E8) medium at 37°C, 5% CO_2, and atmospheric O_2 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,[7] consisting of high glucose DMEM, 100 mg/mL Penicillin/Streptomycin, 50mg/mL L-Proline, 200 mM GlutaMax, 50 mg/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.

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.[30] Stained iPSC-chondrocytes and matched iPSC-MSCs from the same individuals were imaged using an Olympus dissecting microscope.

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),[7] 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). The IRB granted a waiver of informed consent for these samples as they were deidentified. 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.

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).[24–27] Plates were loaded onto the Flexercell baseplate (located in an incubator at 37°C, 5% CO_2, and atmospheric O_2), 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.

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 sequences of primers used for each marker gene are shown in Extended Data Table 9.[82] 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 $E^{(-\Delta\Delta Ct)}$, using *GUSB* as a housekeeping gene in all cases.[31]

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.[32] 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.

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).[33] The software *demuxlet* was used to deconvolute sample identity of individual cell droplets and detect multiplets in multiplexed samples with default parameters.[34] Previously collected and imputed genotype data for the three Yoruba individuals from the HapMap and 1000 Genomes Project were used as input for *demuxlet*.[29,35]

Processed gene count per cell barcode matrices were imported into R using the Seurat package (v3.2.0, RRID: SCR_007322).[36,37] 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 (Extended Data Figure 11[82]).

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.[38] 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,[39] for chondrogenic gene markers.

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,[40] iPSC-chondrocytes from an iPSC-chondrogenic pellet time-course,[41] primary human chondrocytes,[42,43] 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.[44] 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.[40] 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).[41] 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,[42,43] 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).[45] 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.

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.

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

Reads were mapped to the hg38 genome using STAR (version 2.6.1b).[33] Gene expression levels were quantified using the *featureCounts* function in Subread (v1.6.5 RRID:SCR_009803) using standard parameters.[46] All downstream processing and analysis steps were performed in R (v3.6.1, RRID:SCR_001905) unless otherwise stated.

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).[47] 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.

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)[48] 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.

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[38] 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.

$$Y \sim \beta_0 + \beta_1 * treatment + \beta_2 * individual + \beta_3 * sex + \beta_4 * replicate + \beta_5 * RUV_{W1} + \beta_6 * RUV_{W2} + \beta_7 * RIN + \varepsilon \quad (1)$$

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,[3,5] 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.

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.[49] 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[48] 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.

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.[50] 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.

Reanalysis of previous dynamic eQTL studies

We used summary statistics from eQTL mapping in three prior dynamic eQTL studies[13,50,51] 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.[38] 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.

Analysis code

The analysis code used in this study is available on GitHub and is archived with Zenodo.[83]

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.

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.[28] We differentiated the iPSCs along the chondrogenic lineage with an intermediate differentiation step into mesenchymal stem cells (MSCs; Figure 1a) using previously established protocols.[7] iPSC-derived MSCs (iPSC-MSCs) exhibited phenotypes and cell surface marker expression patterns characteristic of primary MSCs[52] (Extended Data Figure 1[82]). iPSC-chondrocytes showed a modest increase in proteoglycan side chain extracellular matrix (ECM) production as compared to matched iPSC-MSCs (Figure 1b; Extended Data Figure 2a[82]). Additionally, immunostaining for COL2A1 of iPSC-chondrocytes from one individual demonstrated increased expression compared to matched iPSC-MSCs (Extended Data Figure 2b[82]).

We treated iPSC-chondrocytes from each individual with 24 hours of CTS, which is known to induce a hypertrophic phenotype in cartilage[24–27]; 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.
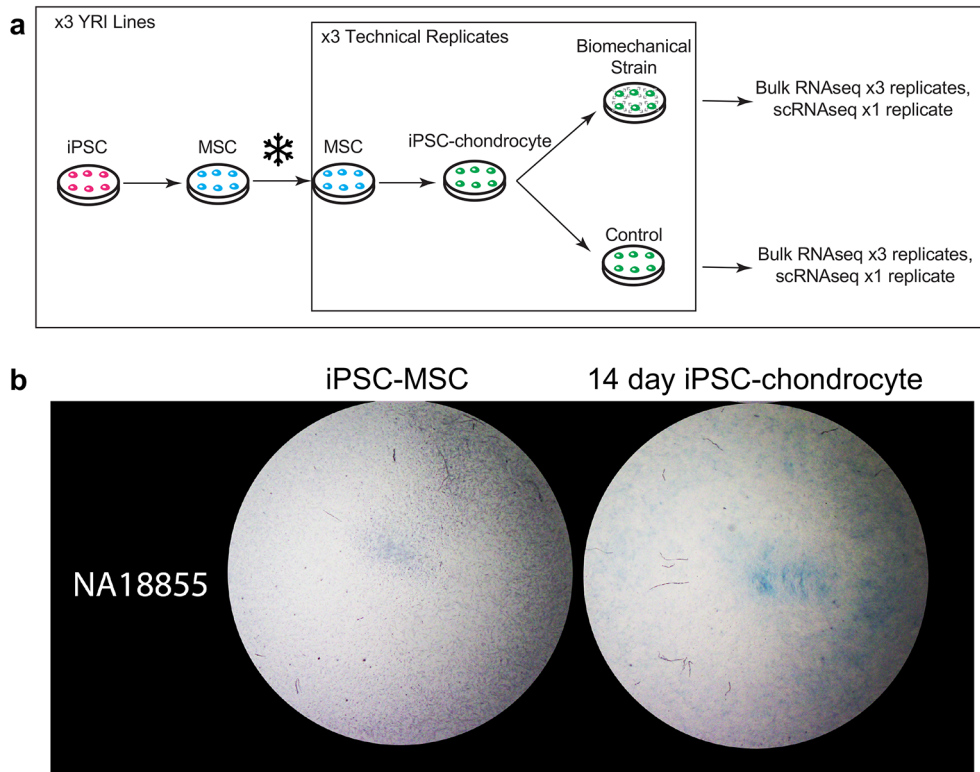
Figure 1. Description of *in vitro* biomechanical strain study design. (a) In our study design, iPSCs generated from three Yoruba individuals were first differentiated along the chondrogenic lineage with an intermediate differentiation step into mesenchymal stem cells (MSCs). iPSC-derived MSCs from each 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 Extended Data Figure 2.

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 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,[25,53] while gene marker *TIMP2* did not change in expression, in line with previous reports (Extended Data Figure 3, Extended Data Table 1[82]).[25]

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 (Methods), we demonstrated that our cells produce glycosaminoglycan ECM, a hallmark of chondrogenesis (Figure 1b; Extended Data Figure 2a[82]). Our cells also produce the collagen COL2A1, a protein that is almost exclusive to cartilage tissues[54,55] (Extended Data Figure 2b[82]). 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
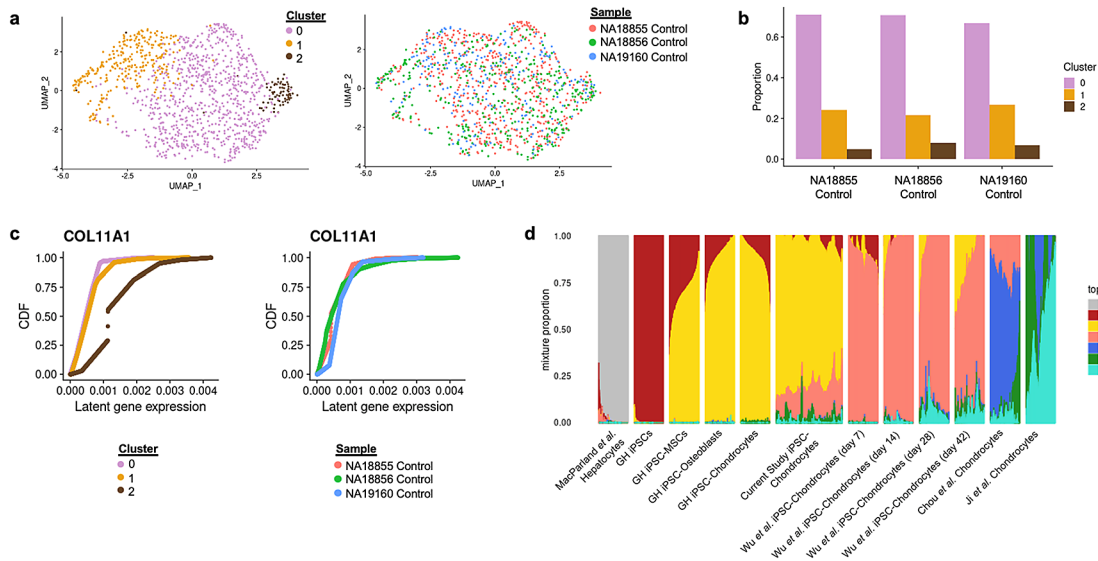
Figure 2. Characterization of cell type composition in iPSC-chondrocyte cultures. (a) Uniform Manifold Approximation and Projection (UMAP) of normalized and integrated single cell RNA sequencing data from control samples from all three individuals included in the study. (Left) UMAP colored by sample 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.

assess potential differences in differentiation efficiency among the three individuals. Following standardization and normalization (Methods), unsupervised clustering of the single cell data revealed three clusters of cells in our untreated control samples (Figure 2a). The proportion of cell membership in each cluster is comparable across individuals (Figure 2b). Five percent of cells from individual NA18855 fall into cluster 2, along with 8% of cells from NA18856 and 7% of 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[56] (Figure 2c; Methods; additional genes shown in Extended Data Figure 4[82]). We found no substantial difference in *COL11A1* expression between individuals. Thus, based on staining and gene expression patterns, iPSC-chondrocytes are undergoing chondrogenesis, and importantly, 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 (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,[57] 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). Individual topics in the model display high word probabilities of gene markers of hepatocytes (*ALB*), chondrocytes (*COL2A1*, *ACAN*, *SOX9*, *SOX5*, *SOX6*, and *COL9A1*), iPSCs (*POU5F1*, *SOX2*, and *NANOG*), and MSCs (*THY1*, *NT5E*, *ENG*) (Extended Data Figure 5a[82]). 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 2d). Topics 4-7 display relatively high levels of expression of several markers of chondrogenesis or cartilage fate, including *SOX9, SOX5, SOX6,*[58] and *COL9A1*[59] (Extended Data Figure 5a[82]). Furthermore, differential expression analyses across topics identified type IX collagen gene *COL9A3*[59] as highly occurring in topic 7 relative to all other topics (Extended Data Figure 5b[82]). Similarly, in topics 4 and 6, the chondrogenic marker *COMP*[60] 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.

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 (Extended Data Table 2[82]). We note the mapped reads from this sample cluster as expected with other technical replicates from the same individual and treatment (Extended Data Figure 6[82]), 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 (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 (Methods). Using this model, we tested for differential expression between treated and untreated cultures. At an FDR of 0.05, 987 genes are significantly differentially expressed (DE) between treated and untreated chondrogenic cells (Figure 3a-b; Extended Data Table 3[82]).

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 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.[61,62]

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.[3] 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.[3] We found that DE genes in our study are significantly enriched within the set of 553 genes previously associated with OA (p = 0.002; Extended Data Table 4[82]).

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.[5] 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 (Extended Data Table 5). 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.[63] *LAMC1* has been identified as a blood-based biomarker for detecting mild knee OA, with lower RNA expression identifying mild OA.[64] 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.[65] 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 (Extended Data Table 6[82]). 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.
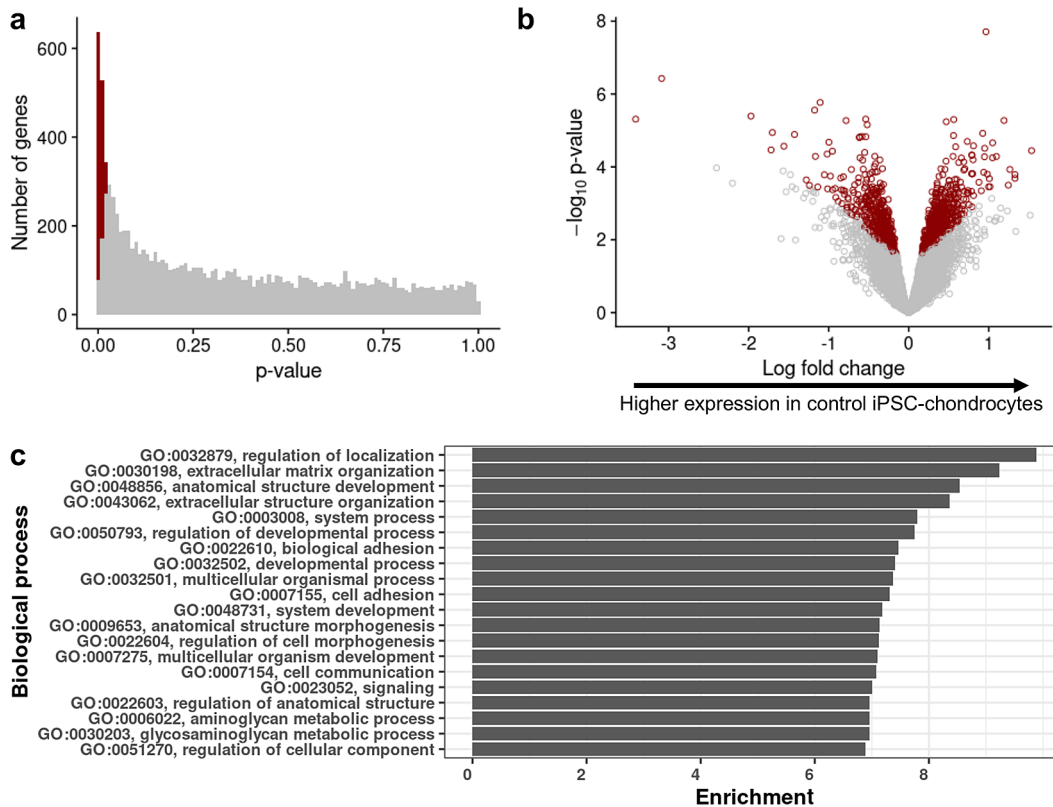
Figure 3. Results from linear mixed model differential expression (DE) analysis between treatment conditions. The linear mixed model used to conduct the analysis also accounted for the random effect of experimental batch and the fixed effects of individual cell line, sex, treatment status, two factors 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.

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 4).

For example, *MMP14* displays a different pattern of expression in each cell line before and after CTS (Figure 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.[66] In addition, *EXOSC8* and *COPG1* (Figure 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.

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
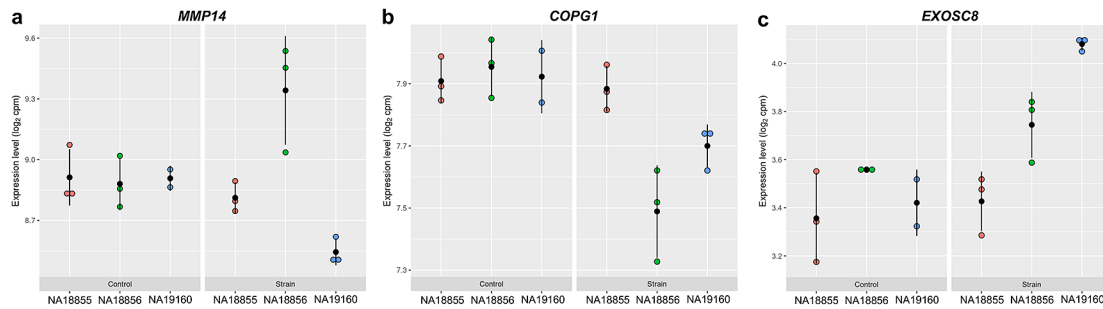
Figure 4. Examples of inter-individual differences in gene expression responses to cyclic tensile strain (CTS). (a-c) *MMP14, COPG1,* and *EXOSC8* each demonstrate inter-individual differences in gene expression in our dataset. Dot plots are the expression level (log2 cpm) of each gene in each sample, 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 down-regulation. 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.

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 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. As each technical replicate used MSCs from the same cryopreservation batch from each individual, we are not able to differentiate between variation introduced by individual level differences and variation introduced by differences in iPSC-MSC differentiation. Thus, these two sources of variation are captured together by the single variable of "individual" in our analysis.

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; Extended Data Figures 7 and 8; Extended Data Table 7[82]). Our results indicate the primary source of gene expression variation is individual (regression of PC1 by individual, $p = 1.34 \times 10^{-7}$, regression of PC2 by individual, $p = 4.61 \times 10^{-4}$). The second largest source of variation in the data is sex (regression of PC2 by sex, $p = 2.67 \times 10^{-4}$), which is unsurprising given our study included one female and two male cell lines. Although treatment shows a minor correlation with PC2 ($R^2 = 0.14$), PC3 ($R^2 = 0.14$), and PC4 ($R^2 = 0.3$), 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[48]; Methods). Following this we observed the top three sources of gene expression variation are individual (regression of PC1 by individual, $p = 7.34 \times 10^{-8}$, regression of PC2 by individual, $p = 3.98 \times 10^{-4}$), sex (regression of PC2 by sex, $p = 1.79 \times 10^{-4}$), and treatment effect (regression of PC3 by treatment, $p = 3.92 \times 10^{-2}$), all three of which are significant (Figure 5a-b, Extended Data Figure 9[82]).

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 (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[48] (Methods). We then included the filtered and normalized gene expression data and this single factor of unwanted variation in the model (Methods; Figure 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
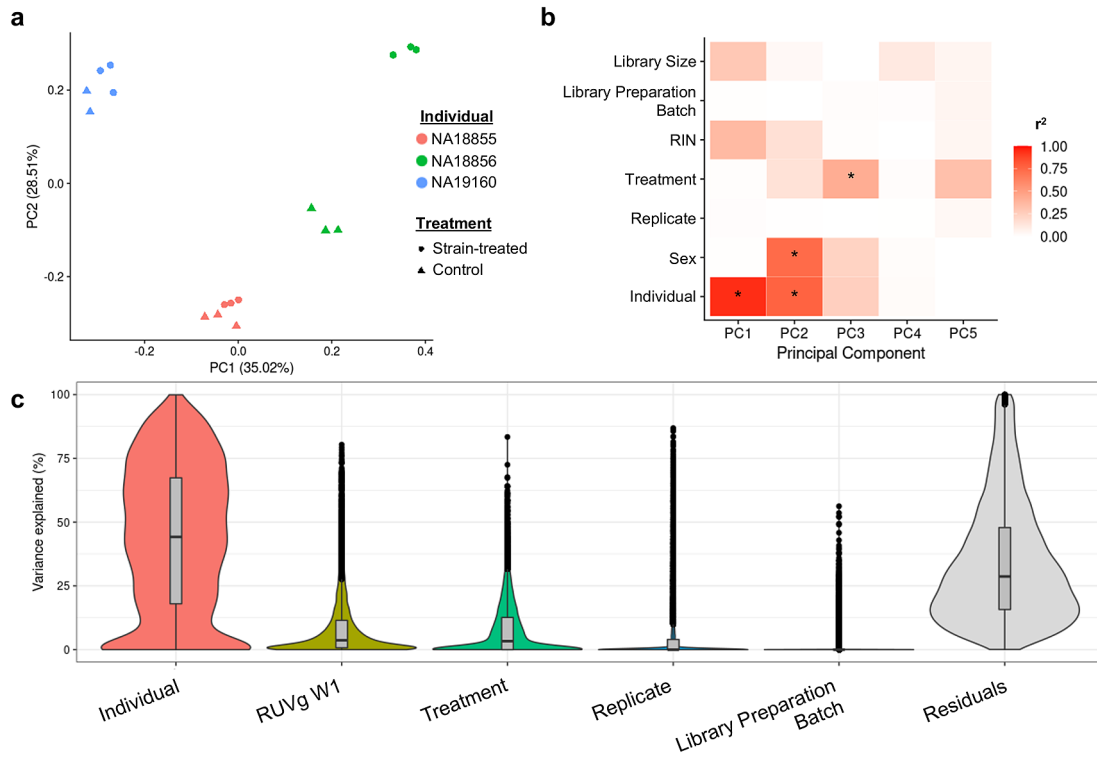
Figure 5. Characterization of sources of variation in bulk RNA sequencing data. (a) Principal components (PC) plot of normalized and RUVs-corrected bulk RNA sequencing samples colored by individual and shaped by treatment condition. Samples largely separate by Individual, with strain-treated samples 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.

negligible amount of variance (median of 8.7×10^{-7} % and 3.5×10^{-7} % variance explained, respectively). We observed similar results when running a model that did not include the factor of unwanted variation (Extended Data Figure 10[82]). 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.

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 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; 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. At this effect size, the power to detect eQTLs comes with a correspondingly low FDR (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[13,50,51] (Figure 6; Extended Data Table 8[82]). 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, the fact these estimates are conservative and come from eQTL studies in three different stimulus conditions demonstrates the potential effectiveness of this approach.
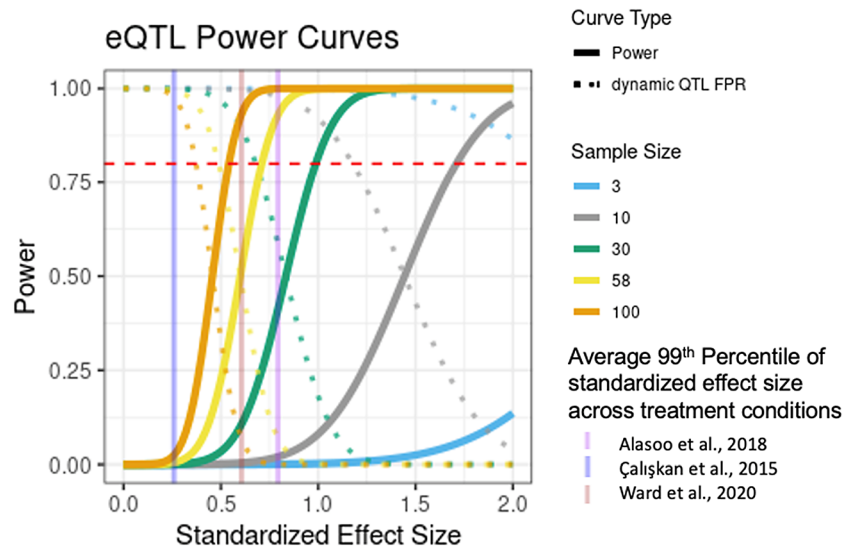
Figure 6. Power analysis for eQTL and dynamic eQTL study with two conditions. Power curves are derived under the assumption of a simple linear regression for expression quantitative trait locus (eQTL) mapping and plotted over standardized effect sizes (effect sizes divided by the phenotype standard deviation) 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.

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.[67] 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.[11]

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.[68–70] 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.[24–27,71–75] 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.[24–27] 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. Furthermore, previous multi-species studies of regulatory sequence variation in cartilage have found a high degree of constraint in these sequences, but that this conservation is violated in human-specific changes, which may bring about OA.[76] The presence of associations between sequence variation in human non-coding genomic regions and OA further supports the importance of gene regulation in this disease.[3] Therefore, eQTL and dynamic eQTL studies, including those carried out in iPSC-chondrocytes, may be fruitful for explaining the connection between human genetic variation and OA or joint health.

We identified specific differences in the gene expression response to CTS between individuals in this study. Additionally, one individual (NA18856) demonstrates a stronger transcriptional response to CTS than that found in the other two individuals in this study, and this stronger response is consistent across three replicates. We anticipate that there will be similar levels of heterogeneity in the severity of response to CTS in other individuals not included in this study. We further expect that these stronger responses represent a difference in the severity, but not the nature, of the gene expression changes during CTS. Therefore, we do not expect this heterogeneity to limit the ability of this system to map eQTLs in future studies, especially it is driven by genotype. 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.[10,13–17,51,77] 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.[10,11,14,16] 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 study does not distinguish between gene expression variation introduced by individual level differences and variation introduced by differences in iPSC-MSC differentiation. However, prior work demonstrates that gene expression patterns between separate iPSC-MSC lines differentiated independently from the same individual are highly consistent, and this consistency persists across species and across variations of MSC differentiation media.[78] Therefore, we anticipate that iPSC-MSC differentiation does not contribute significantly to gene expression variation on the individual level.

Our iPSC-chondrocyte system also facilitates investigations beyond those involving only human cells. The existence of panels of human and nonhuman primate iPSCs[44] 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.[79–81]

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.

Data availability

Underlying data

NCBI GEO: Characterizing gene expression responses to biomechanical strain in an in vitro model of osteoarthritis. Accession number: GSE165874; https://identifiers.org/geo:GSE165874.

NCBI GEO: Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model. Accession number: GSE167240; https://identifiers.org/geo:GSE167240.

Extended data

Open Science Framework: Characterizing gene expression in an in vitro biomechanical strain model of joint health, DOI https://doi.org/10.17605/OSF.IO/YQRJM.[82]

This project contains the following underlying data:

- Extended Data Table 1. Raw RT-PCR amplification

- Extended Data Table 2. Bulk RNA sequencing library metadata and quality control metrics.

- Extended Data Table 3. Differential expression results from limma analysis between strain-treated and control samples. Only genes that passed filters for low expression were kept.

- Extended Data Table 4. Genes from Tachmazidou et al., 2019

- Extended Data Table 5. Genes with significant cross-omics differences between high-grade and low-grade cartilage in Steinberg et al. 2019.

- Extended Data Table 6. Results from 10 random permutations of sample treatment condition labels.

- Extended Data Table 7. Full results of linear regression analysis between experimental variables and top 5 principal components in the bulk RNA sequencing data.

- Extended Data Table 8. Results from reanalysis of prior dynamic eQTL studies for eQTL power analysis.

- Extended Data Table 9. Nucleotide sequences of the primers used for RT-PCR

- Extended Data Figures 1-11

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Analysis code

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

Archived analysis code at time of publication: https://doi.org/10.5281/zenodo.6095200.[83]

License: MIT License

Acknowledgements

References

1. Aigner T, Schmitz N: *173 - Pathogenesis and pathology of osteoarthritis.* 2011.
Publisher Full Text

2. Kloppenburg M, Berenbaum F: **Osteoarthritis year in review 2019: epidemiology and therapy.** *Osteoarthritis Cartilage.* 2020; **28**: 242–248.
PubMed Abstract | **Publisher Full Text**

3. Tachmazidou I, *et al.*: **Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank.** *Nat. Genet.* 2019; **51**: 230–236.
PubMed Abstract | **Publisher Full Text**

4. Li H, Yang HH, Sun ZG, *et al.*: **Whole-transcriptome sequencing of knee joint cartilage from osteoarthritis patients.** *Bone Jt. Res.* 2019; **8**: 290–303.
Publisher Full Text

5. Steinberg J, *et al.*: **Decoding the genomic basis of osteoarthritis.** *bioRxiv 835850 [Preprint].* 2019 [cited 2019 Nov 12].
Publisher Full Text

6. The GTEx Consortium: **The GTEx Consortium atlas of genetic regulatory effects across human tissues.** *Science.* 2020; **369**: 1318–1330.
PubMed Abstract | **Publisher Full Text**

7. Nejadnik H, *et al.*: **Improved Approach for Chondrogenic Differentiation of Human Induced Pluripotent Stem Cells.** *Stem Cell Rev. Rep.* 2015; **11**: 242–253.
PubMed Abstract | **Publisher Full Text**

8. Oldershaw RA, *et al.*: **Directed differentiation of human embryonic stem cells toward chondrocytes.** *Nat. Biotechnol.* 2010; **28**: 1187–1194.
PubMed Abstract | **Publisher Full Text**

9. Umans BD, Battle A, Gilad Y: **Where Are the Disease-Associated eQTLs?.** *Trends Genet.* 2020; 0.

10. Knowles DA, *et al.*: **Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes.** *eLife.* 2018; **7**.
PubMed Abstract | **Publisher Full Text**

11. Strober BJ, *et al.*: **Dynamic genetic regulation of gene expression during cellular differentiation.** *Science.* 2019; **364**: 1287–1290.
PubMed Abstract | **Publisher Full Text**

12. Kariuki SN, *et al.*: **Mapping Variation in Cellular and Transcriptional Response to 1,25-Dihydroxyvitamin D3 in Peripheral Blood Mononuclear Cells.** *PLOS ONE.* 2016; **11**: e0159779.
PubMed Abstract | **Publisher Full Text**

13. Çalışkan M, Baker SW, Gilad Y, *et al.*: **Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection.** *PLOS Genet.* 2015; **11**: e1005111.
PubMed Abstract | **Publisher Full Text**

14. Barreiro LB, *et al.*: **Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection.** *Proc. Natl. Acad. Sci.* 2012; **109**: 1204–1209.
PubMed Abstract | **Publisher Full Text**

15. Manry J, *et al.*: **Deciphering the genetic control of gene expression following Mycobacterium leprae antigen stimulation.** *PLOS Genet.* 2017; **13**: e1006952.
PubMed Abstract | **Publisher Full Text**

16. Kim-Hellmuth S, *et al.*: **Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations.** *Nat. Commun.* 2017; **8**: 266.
PubMed Abstract | **Publisher Full Text**

17. Nédélec Y, *et al.*: **Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens.** *Cell.* 2016; **167**: 657–669.e21.
PubMed Abstract | **Publisher Full Text**

18. Buckwalter JA: **Osteoarthritis and articular cartilage use, disuse, and abuse: experimental studies.** *J. Rheumatol. Suppl.* 1995; **43**: 13–15.
PubMed Abstract

19. Palmoski M, Perricone E, Brandt KD: **Development and reversal of a proteoglycan aggregation defect in normal canine knee cartilage after immobilization.** *Arthritis Rheum.* 1979; **22**: 508–517.

20. Palmoski MJ, Colyer RA, Brandt KD: **Joint motion in the absence of normal loading does not maintain normal articular cartilage.** *Arthritis Rheum.* 1980; **23**: 325–334.
Publisher Full Text

21. Vingård E: **Sport and the Development of Osteoarthrosis of the Hip.** *Sports Med.* 1994; **18**: 1–3.
Publisher Full Text

22. Vingård E, Alfredsson L, Malchau H: **Osteoarthrosis of the hip in women and its relation to physical load at work and in the home.** *Ann. Rheum. Dis.* 1997; **56**: 293–298.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Guilak F: **Biomechanical factors in osteoarthritis.** *Best Pract. Res. Clin. Rheumatol.* 2011; **25**: 815–823.
PubMed Abstract | **Publisher Full Text**

24. Fujisawa T, *et al.*: **Cyclic Mechanical Stress Induces Extracellular Matrix Degradation in Cultured Chondrocytes via Gene Expression of Matrix Metalloproteinases and Interleukin-11.** *J. Biochem.* 1999; **125**: 966–975.
PubMed Abstract | **Publisher Full Text**

25. Honda K, *et al.*: **The effects of high magnitude cyclic tensile load on cartilage matrix metabolism in cultured chondrocytes.** *Eur. J. Cell Biol.* 2000; **79**: 601–609.
PubMed Abstract | **Publisher Full Text**

26. Lin Y-Y, *et al.*: **Applying an excessive mechanical stress alters the effect of subchondral osteoblasts on chondrocytes in a co-culture system.** *Eur. J. Oral Sci.* 2010; **118**: 151–158.
PubMed Abstract | **Publisher Full Text**

27. Pichler K, *et al.*: **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.* 2013; **95-B**: 568–573.
PubMed Abstract | **Publisher Full Text**

28. Banovich NE, *et al.*: **Impact of regulatory variation across human iPSCs and differentiated cells.** *Genome Res.* 2018; **28**: 122–131.
PubMed Abstract | **Publisher Full Text**

29. Gibbs RA, *et al.*: **The International HapMap Project.** *Nature.* 2003; **426**: 789–796.
Publisher Full Text

30. Goldstein DJ, Horobin RW: **Surface staining of cartilage by Alcian Blue, with reference to the role of microscopic dye aggregates in histological staining.** *Histochem. J.* 1974; **6**: 175–184.
PubMed Abstract | **Publisher Full Text**

31. Rao X, Huang X, Zhou Z, *et al.*: **An improvement of the 2⁻(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis.** *Biostat. Bioinforma. Biomath.* 2013; **3**: 71–85.
PubMed Abstract

32. Zheng GXY, *et al.*: **Massively parallel digital transcriptional profiling of single cells.** *Nat. Commun.* 2017; **8**: 14049.
PubMed Abstract | **Publisher Full Text**

33. Dobin A, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013; **29**: 15–21.
PubMed Abstract | **Publisher Full Text**

34. Kang HM, *et al.*: **Multiplexed droplet single-cell RNA-sequencing using natural genetic variation.** *Nat. Biotechnol.* 2018; **36**: 89–94.
PubMed Abstract | **Publisher Full Text**

35. Clarke L, *et al.*: **The 1000 Genomes Project: data management and community access.** *Nat. Methods.* 2012; **9**: 459–462.
PubMed Abstract | **Publisher Full Text**

36. Butler A, Hoffman P, Smibert P, *et al.*: **Integrating single-cell transcriptomic data across different conditions, technologies,**

and species. *Nat. Biotechnol.* 2018; **36**: 411–420.
PubMed Abstract | **Publisher Full Text**

37. Stuart T, *et al.*: **Comprehensive Integration of Single-Cell Data.** *Cell.* 2019; **177**: 1888–1902.e21.
PubMed Abstract | **Publisher Full Text**

38. Stephens M: **False discovery rates: a new deal.** *Biostat. Oxf. Engl.* 2017; **18**: 275–294.

39. Sarkar A, Stephens M: **Separating measurement and expression models clarifies confusion in single cell RNA-seq analysis.** *bioRxiv.* 2020.
Publisher Full Text

40. MacParland SA, *et al.*: **Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations.** *Nat. Commun.* 2018; **9**: 4383.
PubMed Abstract | **Publisher Full Text**

41. Wu C-L, *et al.*: **Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis.** *Nat. Commun.* 2021; **12**: 362.
PubMed Abstract | **Publisher Full Text**

42. Chou C-H, *et al.*: **Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis.** *Sci. Rep.* 2020; **10**: 10868.
PubMed Abstract | **Publisher Full Text**

43. Ji Q, *et al.*: **Single-cell RNA-seq analysis reveals the progression of human osteoarthritis.** *Ann. Rheum. Dis.* 2018; **78**: 100–110.
PubMed Abstract | **Publisher Full Text**annrheumdis-2017-212863.

44. Gallego Romero I, *et al.*: **A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics.** *eLife.* 2015; **4**: e07103.
PubMed Abstract | **Publisher Full Text**

45. Carbonetto P, Luo K, Dey K, *et al.*: **fastTopics: fast algorithms for fitting topic models and non-negative matrix factorizations to count data. R package version 0.4-11.** 2021.

46. Liao Y, Smyth GK, Shi W: **The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.** *Nucleic Acids Res.* 2013; **41**: e108–e108.
PubMed Abstract | **Publisher Full Text**

47. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics.* 2010; **26**: 139–140.
PubMed Abstract | **Publisher Full Text**

48. Risso D, Ngai J, Speed TP, *et al.*: **Normalization of RNA-seq data using factor analysis of control genes or samples.** *Nat. Biotechnol.* 2014; **32**: 896–902.
PubMed Abstract | **Publisher Full Text**

49. Hoffman GE, Schadt EE: **variancePartition: interpreting drivers of variation in complex gene expression studies.** *BMC Bioinformatics.* 2016; **17**: 483.
PubMed Abstract | **Publisher Full Text**

50. Ward MC, Banovich NE, Sarkar A, *et al.*: **Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes.** *eLife.* 2021; **10**: e57345.
PubMed Abstract | **Publisher Full Text**

51. Alasoo K, *et al.*: **Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response.** *Nat. Genet.* 2018; **50**: 424–431.
PubMed Abstract | **Publisher Full Text**

52. Dominici M, *et al.*: **Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement.** *Cytotherapy.* 2006; **8**: 315–317.
PubMed Abstract | **Publisher Full Text**

53. Tetsunaga T, *et al.*: **Regulation of mechanical stress-induced MMP-13 and ADAMTS-5 expression by RUNX-2 transcriptional factor in SW1353 chondrocyte-like cells.** *Osteoarthritis Cartilage.* 2011; **19**: 222–232.
PubMed Abstract | **Publisher Full Text**

54. Khetarpal U, Robertson NG, Yoo TJ, *et al.*: **Expression and localization of COL2A1 mRNA and type II collagen in human fetal cochlea.** *Hear. Res.* 1994; **79**: 59–73.
PubMed Abstract | **Publisher Full Text**

55. **COL2A1 protein expression summary - The Human Protein Atlas.**
Reference Source

56. Li Y, *et al.*: **A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis.** *Cell.* 1995; **80**: 423–430.
PubMed Abstract | **Publisher Full Text**

57. Dey KK, Hsiao CJ, Stephens M: **Visualizing the structure of RNA-seq expression data using grade of membership models.** *PLOS Genet.* 2017; **13**: e1006599.
PubMed Abstract | **Publisher Full Text**

58. Liu C-F, Lefebvre V: **The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis.** *Nucleic Acids Res.* 2015; **43**: 8183–8203.
PubMed Abstract | **Publisher Full Text**

59. Luo Y, *et al.*: **The minor collagens in articular cartilage.** *Protein Cell.* 2017; **8**: 560–572.
PubMed Abstract | **Publisher Full Text**

60. Yi SW, *et al.*: **Gene expression profiling of chondrogenic differentiation by dexamethasone-conjugated polyethyleneimine with SOX trio genes in stem cells.** *Stem Cell Res. Ther.* 2018; **9**: 341.
PubMed Abstract | **Publisher Full Text**

61. Lorenzo P, Bayliss MT, Heinegård D: **Altered patterns and synthesis of extracellular matrix macromolecules in early osteoarthritis.** *Matrix Biol. J. Int. Soc. Matrix Biol.* 2004; **23**: 381–391.
PubMed Abstract | **Publisher Full Text**

62. Rahmati M, Nalesso G, Mobasheri A, *et al.*: **Aging and osteoarthritis: Central role of the extracellular matrix.** *Ageing Res. Rev.* 2017; **40**: 20–30.
PubMed Abstract | **Publisher Full Text**

63. Wang W, Rigueur D, Lyons KM: **TGFβ Signaling in Cartilage Development and Maintenance.** *Birth Defects Res. Part C Embryo Today Rev.* 2014; **102**: 37–51.
PubMed Abstract | **Publisher Full Text**

64. Marshall KW, *et al.*: **Blood-based biomarkers for detecting mild osteoarthritis in the human knee.** *Osteoarthritis Cartilage.* 2005; **13**: 861–871.
PubMed Abstract | **Publisher Full Text**

65. Conesa A, *et al.*: **A survey of best practices for RNA-seq data analysis.** *Genome Biol.* 2016; **17**: 13.
PubMed Abstract | **Publisher Full Text**

66. Chubinskaya S, Kuettner KE, Cole AA: **Expression of matrix metalloproteinases in normal and damaged articular cartilage from human knee and ankle joints.** *Lab. Investig. J. Tech. Methods Pathol.* 1999; **79**: 1669–1677.

67. Yoon HH, Bhang SH, Shin J-Y, *et al.*: **Enhanced Cartilage Formation via Three-Dimensional Cell Engineering of Human Adipose-Derived Stem Cells.** *Tissue Eng. Part A.* 2012; **18**: 1949–1956.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

68. Gurdasani D, Barroso I, Zeggini E, *et al.*: **Genomics of disease risk in globally diverse populations.** *Nat. Rev. Genet.* 2019; **20**: 520–535.
PubMed Abstract | **Publisher Full Text**

69. Shi H, *et al.*: **Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data.** *Am. J. Hum. Genet.* 2020; **106**: 805–817.
PubMed Abstract | **Publisher Full Text**

70. Mogil LS, *et al.*: **Genetic architecture of gene expression traits across diverse populations.** *PLOS Genet.* 2018; **14**: e1007586.
PubMed Abstract | **Publisher Full Text**

71. Takahashi K, *et al.*: **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.* 1997; **15**: 150–158.
PubMed Abstract | **Publisher Full Text**

72. Takahashi K, *et al.*: **Hydrostatic pressure induces expression of interleukin 6 and tumour necrosis factor α mRNAs in a chondrocyte-like cell line.** *Ann. Rheum. Dis.* 1998; **57**: 231–236.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

73. Takano-Yamamoto T, *et al.*: **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.* 1991; **99**: 448–455.

74. Mohtai M, *et al.*: **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.* 1996; **14**: 67–73.
PubMed Abstract | **Publisher Full Text**

75. Bougault C, Paumier A, Aubert-Foucher E, *et al.*: **Investigating conversion of mechanical force into biochemical signaling in three-dimensional chondrocyte cultures.** *Nat. Protoc.* 2009; **4**: 928–938.
PubMed Abstract | **Publisher Full Text**

76. Richard D, Liu Z, Cao J, *et al.*: **Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk.** *Cell.* 2020; **181**: 362–381.e28.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

77. Maranville JC, *et al.*: **Interactions between Glucocorticoid Treatment and Cis-Regulatory Polymorphisms Contribute to**

Cellular Response Phenotypes. *PLoS Genet.* 2011; **7**: e1002162.
PubMed Abstract | **Publisher Full Text**

78. Housman G, Briscoe E, Gilad Y: **Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model.** *PLoS Genet.* 2022; **18**(3): e1010073.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

79. Jurmain R: **Degenerative joint disease in African great apes: an evolutionary perspective.** *J. Hum. Evol.* 2000; **39**: 185–203.
PubMed Abstract | **Publisher Full Text**

80. Videan EN, Lammey ML, Lee DR: **Diagnosis and Treatment of Degenerative Joint Disease in a Captive Male Chimpanzee (Pan troglodytes).** *J. Am. Assoc. Lab. Anim. Sci. JAALAS.* 2011; **50**:

263–266.
PubMed Abstract

81. Lowenstine LJ, McManamon R, Terio KA: **Comparative Pathology of Aging Great Apes: Bonobos, Chimpanzees, Gorillas, and Orangutans.** *Vet. Pathol.* 2016; **53**: 250–276.
PubMed Abstract | **Publisher Full Text**

82. Hung A: **Characterizing gene expression in an in vitro biomechanical strain model of joint health.** 2022, February 23.
Publisher Full Text

83. Hung A: **anthonyhung/invitrostrain_pilot_repository: Characterizing gene expression in an in vitro biomechanical strain model of joint health (v1.0.0).** *Zenodo.* 2022.
Publisher Full Text

![F1000 Research]

Open Peer Review

Current Peer Review Status: ✓ ✗

- -

Version 2

Reviewer Report 21 December 2022

https://doi.org/10.5256/f1000research.138770.r151592

✓ **Alex Pollen** [iD]

The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, USA

The authors have met my concerns and I would like to accept without reservation.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 05 October 2022

https://doi.org/10.5256/f1000research.138770.r151593

✗ **Shireen Lamande** [iD]

[1] Murdoch Children's Research Institute, Parkville, VIC, 3052, Australia
[2] The University of Melbourne, Parkville, VIC, 3052, Australia

In response to my first review the authors concede that they do not have gene expression for ACAN, COL2A1 or COL9A1, to quote "zero counts measured for these genes". If I was asked to identify three genes that define a chondrocyte, these are the three I would pick. They are key chondrocyte markers and if these genes are not expressed in the cells they are studying they are not studying chondrocytes. There may be a different story in their data but it is not one about biomechanical strain and joint health. This is the fundamental issue.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetic skeletal disorders, cartilage, bone, iPSC differentiation to chondrocytes, bulk RNAseq analysis and DE gene expression

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 19 Jan 2023

Anthony Hung

We appreciate the response about the identity of our iPSC-derived cell types and the concern about the relevance of our results to primary chondrocytes. While we do not claim to perfectly recapitulate primary cells, we believe that as a model our system is bound to be an imperfect reflection of *in vivo* biology and this does not necessarily invalidate it as a useful approach to studying gene expression in joint health and biomechanical stress. Furthermore, the similarity between the gene expression patterns of our chondrogenic cell lines and those found in Wu *et al*, 2021 iPSC-chondrocytes demonstrates that beyond marker genes, many gene expression patterns are shared between our chondrogenic cells and those generated using alternative methods. If necessary, we can further make more textually clear the differences between our chondrogenic cell lines and primary chondrocytes in an updated version of the text.

Competing Interests: No competing interests were disclosed.

Version 1

Reviewer Report 29 June 2022

https://doi.org/10.5256/f1000research.121122.r138892

✗ **Shireen Lamande** [iD]
¹ Murdoch Children's Research Institute, Parkville, VIC, 3052, Australia
² The University of Melbourne, Parkville, VIC, 3052, Australia

The Hung et al manuscript seeks to define the gene regulatory responses of three human iPSC lines that have been differentiated into 'chondrocytes' to cyclic tensile strain treatment. The premise is that measuring how chondrocytes from different individuals respond to mechanical stress could uncover gene-environment interactions that are relevant to human joint health. This

is a question worthy of study – understanding how individual differences influence joint health could identify new therapeutic targets to treat osteoarthritis, an increasingly common and debilitating disease.

The paper includes a lot of complex bioinformatic and statistical analyses of single cell RNAseq and bulk RNAseq data generated in this study and also includes new analyses of publicly available RNAseq data that, in the first instance, are designed to demonstrate that 'the process of chondrocyte differentiation is robust to individual differences.' It seems this complex analysis was undertaken because the cell populations being compared contain only a small proportion of chondrocytes.

To draw robust conclusions about how chondrocytes respond to mechanical forces it is critical to start the study with a well-defined chondrocyte population. Hence, my reservations about this manuscript stem from the cell population that was studied. The study uses iPSCs from three individuals, 2 male and 1 female. The iPSCs were first differentiated to mesenchymal stem cells (MSCs) and after several passages the MSCs were cryopreserved at passage 5-7. A single batch of cryopreserved MSCs from each iPSC line was used for multiple technical replicate experiments where the MSCs were differentiated in chondrogenic medium for 14 days.

The greatest source of variability in iPSC differentiation protocols are technical parameters (eg different differentiation batch) rather than the cell line (see Phipson et al (2018) and this means that comparisons must be made using samples from concurrent carefully matched differentiations.[1] It is not clear if concurrent differentiations to MSCs were part of the experimental design but the images in extended data figure 1 show cells at very different densities suggesting differences in cell growth or seeding density. The MSCs differentiated from the three lines are compared by FACS using 8 different markers (extended data Fig 1) and this shows considerable variation between the three lines. Given the importance of starting the chondrocyte differentiation from matched cell populations it is imperative to show that this first step is robust and reproducible and that the cell population is similar at the start of the chondrocyte differentiation step. This is not shown in the manuscript. The authors conclude that "individual cell line contributes the largest amount of variance to the data" but this is based on a single differentiation to MSCs then technical replicates of the chondrocyte differentiation stage and it is unclear how reproducible the MSC differentiation is. The variance could have been generated by technical variation in the early differentiation stages.

Chondrocyte differentiation. Figure 1b and extended data Figure 2a show alcian blue staining. Alcian blue stains proteoglycan side chains so the staining doesn't show a collagenous ECM as claimed in the results. The staining is localised and very weak and in one line, not visible at all in the mechanically unstimulated cells. Data from one individual shows slight collagen II staining (extended data figure 2b). This staining pattern suggests that the chondrocyte differentiation protocol has produced a small proportion of immature chondroprogenitors and the scRNAseq data confirms this. Between 5% and 8% of the cells from each line form a cluster that is 'most like chondrocytes'. The only gene expression data shown that supports the chondrocyte like nature of this cell cluster is COL11A1 in figure 2c and TIMP2 and TIMP3 in extended data figure 4. It would be much easier to judge this if this is a chondrocyte population if expression data for a range of genes characteristically expressed in chondrocytes was shown – for example SOX9, SOX5, SOX6, COL2A1, ACAN, COL9A1, MATN3, SPARC, COL11A1, COL10A1, PRG4 – a comparison with the scRNAseq data published in the Wu et al 2021 paper would also be helpful when judging what

stage of chondrocyte development has been induced.[2] To substantiate any conclusions about how individual differences influence how chondrocytes respond to mechanical forces the authors would need to compare chondrocytes rather than a mixed population of cells where immature chondrocyte progenitors are <10% of the total cell population. When >90% of the cell population are not chondrocytes it is difficult to see how the differential gene expression in mechanically stimulated cultures can be attributed to the way chondrocytes respond.

References

1. Phipson B, Er PX, Combes AN, Forbes TA, et al.: Evaluation of variability in human kidney organoids.*Nat Methods*. **16** (1): 79-87 PubMed Abstract | Publisher Full Text
2. Wu C, Dicks A, Steward N, Tang R, et al.: Single cell transcriptomic analysis of human pluripotent stem cell chondrogenesis. *Nature Communications*. 2021; **12** (1). Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genetic skeletal disorders, cartilage, bone, iPSC differentiation to chondrocytes, bulk RNAseq analysis and DE gene expression

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 16 Sep 2022

Anthony Hung

Reviewer comment 1: The Hung et al manuscript seeks to define the gene regulatory responses of three human iPSC lines that have been differentiated into 'chondrocytes' to cyclic tensile strain treatment. The premise is that measuring how chondrocytes from

different individuals respond to mechanical stress could uncover gene-environment interactions that are relevant to human joint health. This is a question worthy of study – understanding how individual differences influence joint health could identify new therapeutic targets to treat osteoarthritis, an increasingly common and debilitating disease.

The paper includes a lot of complex bioinformatic and statistical analyses of single cell RNAseq and bulk RNAseq data generated in this study and also includes new analyses of publicly available RNAseq data that, in the first instance, are designed to demonstrate that 'the process of chondrocyte differentiation is robust to individual differences.' It seems this complex analysis was undertaken because the cell populations being compared contain only a small proportion of chondrocytes.

To draw robust conclusions about how chondrocytes respond to mechanical forces it is critical to start the study with a well-defined chondrocyte population. Hence, my reservations about this manuscript stem from the cell population that was studied. The study uses iPSCs from three individuals, 2 male and 1 female. The iPSCs were first differentiated to mesenchymal stem cells (MSCs) and after several passages the MSCs were cryopreserved at passage 5-7. A single batch of cryopreserved MSCs from each iPSC line was used for multiple technical replicate experiments where the MSCs were differentiated in chondrogenic medium for 14 days.

The greatest source of variability in iPSC differentiation protocols are technical parameters (eg different differentiation batch) rather than the cell line (see Phipson et al (2018) and this means that comparisons must be made using samples from concurrent carefully matched differentiations.[1] It is not clear if concurrent differentiations to MSCs were part of the experimental design but the images in extended data figure 1 show cells at very different densities suggesting differences in cell growth or seeding density. The MSCs differentiated from the three lines are compared by FACS using 8 different markers (extended data Fig 1) and this shows considerable variation between the three lines. Given the importance of starting the chondrocyte differentiation from matched cell populations it is imperative to show that this first step is robust and reproducible and that the cell population is similar at the start of the chondrocyte differentiation step. This is not shown in the manuscript. The authors conclude that "individual cell line contributes the largest amount of variance to the data" but this is based on a single differentiation to MSCs then technical replicates of the chondrocyte differentiation stage and it is unclear how reproducible the MSC differentiation is. The variance could have been generated by technical variation in the early differentiation stages.

Author Response to comment 1: The reviewer indicated that they believe that the study is not sound and that our conclusions are not supported by the data. We do not understand these statements considering the more specific concerns detailed by the reviewer. We agree with the reviewer that our cellular model is not optimal, and that the differentiated chondrocytes are not mature and represent a small proportion of the culture. These details, however, are known to the reviewer because we clearly reported them in our paper and explained our conclusions in relationship to these caveats. We are aware of the limitations of the study and believe we reported them accordingly. In our opinion, this is the requirement for a sound study – being aware of the limitations. We provide further details

in our responses below.

Cell heterogeneity is critically important to consider, and we appreciate the reviewer's request for further clarification. While it is true that iPSC differentiation batch contributes greatly to variation in many complex organoid differentiation protocols, our observations in monoculture differentiations of iPSCs have not found this to be as sizeable of an issue. Data from Housman, Briscoe, and Gilad 2022 demonstrate that gene expression patterns between iPSC-MSC lines differentiated from the same individual but in different batches and even using different formulations of MSC media are remarkably similar (97% correlation between matched iPSC-MSC lines). This high degree of correlation is observed between human (98% correlation) and chimpanzee (93% correlation) iPSC-MSC lines, demonstrating the robustness of this observation.

Furthermore, our gene expression measurements across technical replicates of the experiment allow us to measure the contribution of individual variables (encompassing both individual cell line differences and individual donor differences) to variation in the system. These technical replicates of the experiment also allow us to account for individual factors in downstream analyses, which demonstrate reliable differences between treatment groups. We have modified our textual descriptions of individual cell line contributions to gene expression variation to reflect the fact that this variable encompasses both cell line and donor variation in the Results section.

If any of the individual level gene expression variation observed in this study segregates with genotype, this will be detected in an eQTL mapping study. While much can and has been achieved through this small scale study to determine the feasibility of an eQTL approach in this system, to definitively establish this feasibility, a full scale experiment will need to be performed. Therefore, we believe this study has achieved its intended purpose.

In this study, we made the choice to pause the differentiation at the iPSC-MSC state for each individual by cryopreserving cells in order to facilitate the design of larger-scale experiments including many more individuals. During the differentiation of each individual to MSCs, we were careful to keep reagent lot numbers consistent to minimize technical differences between MSC lines. The differentiation from iPSC-MSC to iPSC-chondrocyte consistently occurs over a period of 14 days, and the differentiation from iPSC to iPSC-MSC occurs over a period of around 28 days but is variable between individuals. Cryopreservation of iPSC-MSC lines enables larger scale studies to more reliably synchronize the collection of data from multiple individuals simultaneously to reduce the impact of batch effects from multiple sample collections.

Images of MSCs in Extended Data Figure 1 were taken prior to seeding of cells for differentiation and were not standardized to the same confluency, though this was not made adequately clear in the previous figure description. In Extended Data Figure 1, we now include images of iPSC-chondrocytes from each of the lines after 14 days of differentiation, demonstrating equal confluency of cells between individuals, in an updated version of the text.

Reviewer comment 2: Chondrocyte differentiation. Figure 1b and extended data Figure 2a show alcian blue staining. Alcian blue stains proteoglycan side chains so the staining doesn't show a collagenous ECM as claimed in the results. The staining is localised and very weak and in one line, not visible at all in the mechanically unstimulated cells. Data from one individual shows slight collagen II staining (extended data figure 2b). This staining pattern suggests that the chondrocyte differentiation protocol has produced a small proportion of immature chondroprogenitors and the scRNAseq data confirms this. Between 5% and 8% of the cells from each line form a cluster that is 'most like chondrocytes'. The only gene expression data shown that supports the chondrocyte like nature of this cell cluster is COL11A1 in figure 2c and TIMP2 and TIMP3 in extended data figure 4. It would be much easier to judge this if this is a chondrocyte population if expression data for a range of genes characteristically expressed in chondrocytes was shown – for example SOX9, SOX5, SOX6, COL2A1, ACAN, COL9A1, MATN3, SPARC, COL11A1, COL10A1, PRG4 – a comparison with the scRNAseq data published in the Wu et al 2021 paper would also be helpful when judging what stage of chondrocyte development has been induced.[2] To substantiate any conclusions about how individual differences influence how chondrocytes respond to mechanical forces the authors would need to compare chondrocytes rather than a mixed population of cells where immature chondrocyte progenitors are <10% of the total cell population. When >90% of the cell population are not chondrocytes it is difficult to see how the differential gene expression in mechanically stimulated cultures can be attributed to the way chondrocytes respond.

Author Response to comment 2: We appreciate the note about Alcian blue staining indicating proteoglycan side chains rather than collagen production. This wording is changed in our revised manuscript. We now include expression plots for several other chondrogenic marker genes as suggested. However, we were not able to plot expression data for COL2A1, ACAN, or COL9A1 due to zero counts measured for these genes.

Based on the topic model, Wu et al, 2021 iPSC-chondrocytes display gene expression patterns that are shared with our chondrogenic cells and with primary chondrocytes from Chou et al, 2020. While the iPSC-chondrocytes in Wu et al, 2021 do not explicitly move through an hMSC state as in our differentiation, the authors do observe MSC-like cells in their cultures. The presence of MSCs is reflected in the representation of topic 1 (iPSC-MSCs) in these samples in the topic model.

The majority of cell types in the iPSC chondrogenic cell cultures are closer to a MSC state than to a chondrocyte state, but the responses of these chondrogenic cells are still relevant to joint health. DE genes between control and CTS-treated iPSC-derived cells are enriched for genes relevant to joint health as determined through prior differential expression studies of high and low grade OA cartilage and genes implicated in OA GWAS. Also, these relatively immature cells still hold potential utility for studies of genetic control of gene regulation. Transient forms of gene regulation present in these intermediate chondrogenic cells may be masked in mature cell types, yet may still play a role in OA development, either through regulating proper chondrogenesis or modifying joint growth trajectories in a way that predisposes a joint to developing OA (Pitsillides & Beier, 2011). Further evidence for this hypothesis is present in the implication of chondrogenic development genes, including *SMAD3, TGFB1, and GDF5,* in OA susceptibility through OA GWAS (Boer et al, 2021). Thus, we

do agree with the fact that our model does not capture completely the responses of primary chondrocytes to mechanical strain, but that this does not necessarily invalidate it as a useful approach to study gene regulation in OA.

References:

Housman, G., Briscoe, E., & Gilad, Y. (2022). Evolutionary insights into primate skeletal gene regulation using a comparative cell culture model. PLoS Genetics, 18(3), e1010073. https://doi.org/10.1371/journal.pgen.1010073

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.
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.

Pitsillides, A., Beier, F. Cartilage biology in osteoarthritis—lessons from developmental biology. Nat Rev Rheumatol 7, 654–663 (2011). https://doi.org/10.1038/nrrheum.2011.129

Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, Wu TT, Zheng J, Hartley A, Teder-Laving M, Skogholt AH, Terao C, Zengini E, Alexiadis G, Barysenka A, Bjornsdottir G, Gabrielsen ME, Gilly A, Ingvarsson T, Johnsen MB, Jonsson H, Kloppenburg M, Luetge A, Lund SH, Mägi R, Mangino M, Nelissen RRGHH, Shivakumar M, Steinberg J, Takuwa H, Thomas LF, Tuerlings M; arcOGEN Consortium; HUNT All-In Pain; ARGO Consortium; Regeneron Genetics Center, Babis GC, Cheung JPY, Kang JH, Kraft P, Lietman SA, Samartzis D, Slagboom PE, Stefansson K, Thorsteinsdottir U, Tobias JH, Uitterlinden AG, Winsvold B, Zwart JA, Davey Smith G, Sham PC, Thorleifsson G, Gaunt TR, Morris AP, Valdes AM, Tsezou A, Cheah KSE, Ikegawa S, Hveem K, Esko T, Wilkinson JM, Meulenbelt I, Lee MTM, van Meurs JBJ, Styrkársdóttir U, Zeggini E. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021 Sep 2;184(18):4784-4818.e17. doi: 10.1016/j.cell.2021.07.038. Epub 2021 Aug 26. Erratum in: Cell. 2021 Nov 24;184(24):6003-6005. PMID: 34450027; PMCID: PMC8459317.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 April 2022

https://doi.org/10.5256/f1000research.121122.r127025

? **Alex Pollen** 🔟

The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, USA

Jenelle Wallace

University of California San Francisco, San Francisco, CA, USA

The manuscript by Hung et al seeks to build a foundation for measuring population-level gene expression responses to biomechanical stress in cartilage. This framework is important because recent studies indicate that genetic variation linked to physiological responses of relevant tissues relates better to disease traits than does baseline genetic variation. However, studying dynamic changes in gene expression in human tissue across many individuals is extremely challenging. This study uses pluripotent stem cells to generate chondrocytes and measure responses to biomechanical stress via cyclic tensile strain.

The major findings are that: 1) the chondrocyte differentiation protocol does produce a consistent, albeit low (5 – 8%) percentage of chondrocytes that share gene expression features with normal human chondrocytes and produce proteoglycans; 2) that cyclic tensile strain induces gene expression changes in relevant gene ontology categories; 3) that the induction of some genes varies; 4) given low technical variation the system is likely to enable dynamic eQTL mapping with reasonably high sensitivity in an available panel of 58 iPSC lines.

This feasibility study has a number of strengths – strong premise of studying dynamic responses to physiological stimuli, elegant topic modeling to examine the fidelity of iPS-derived chondrocytes to primary samples, and providing a careful feasibility rationale for larger-scale future QTL mapping studies.

Given that this represents a feasibility study for future QTL mapping, there are a few experiments/explanations that would strengthen the foundation:

1. What is the nature of the cells that make up the majority of the chondrocyte differentiation? Are these simply cells in the right lineage, but in a more immature state? If so, the foundation for future studies would benefit from experimenting with longer differentiation times, or strategies to accelerate differentiation/maturation, or enrichment strategies (e.g., FACS or MACS) to focus more specifically on the cell type of interest. Indeed, 5% - 8% are relatively similar, but still represent a 60% difference in efficiency. Developing antibody markers of the on-target lineage would also be beneficial to future studies for assessing/enriching samples prior to sequencing.

2. I am concerned that the NA18856 cells appear to be more dramatically changed by the cyclic tensile strain on the axes of PC1 and especially PC2 than the other cell lines. This is dismissed as "treatment shows a minor correlation with PC2 (R2 = 0.14), PC3 (R2 = 0.14), and PC4 (R2 = 0.3)", but I wonder if this result implies concerted individual variability across many genes in this system, rather than random variability in responses across individuals at specific loci. It would help to also show the samples plotted along PC3 and PC4 to evaluate if the response of this single individual to treatment is driving the correlations, and to discuss possible implication of a single individual having distinct concerted responses.

Minor comments:

1. This claim in the introduction is too strong "Through this study, we evaluated whether the

process of chondrocyte differentiation is robust to individual differences" given that only used 3 individuals were used in the present study.

2. The authors could cite analyses of human sequence variation in chondrocyte regulatory elements: e.g. Richard *et al.* (2020).[1]

3. For the topic modeling, this is discussed in the text, but it would be helpful to show an additional panel in the figure or supplement on the key genes in each topic and their preservation, especially for markers characteristic of adult chondrocytes.

4. The authors could consider presenting Figure 5 before Figure 3 since Figure 5 makes it clear why using a linear model to account for the effects of individual, sex, etc is a useful approach for discovering the effect of treatment in this dataset.

References
1. Richard D, Liu Z, Cao J, Kiapour A, et al.: Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. *Cell*. 2020; **181** (2): 362-381.e28 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: stem cell biology, gene regulation.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 16 Sep 2022

Anthony Hung

The point-by-point responses to comments:

Reviewer comment 1: What is the nature of the cells that make up the majority of the chondrocyte differentiation? Are these simply cells in the right lineage, but in a more immature state? If so, the foundation for future studies would benefit from experimenting with longer differentiation times, or strategies to accelerate differentiation/maturation, or enrichment strategies (e.g., FACS or MACS) to focus more specifically on the cell type of interest. Indeed, 5% - 8% are relatively similar, but still represent a 60% difference in efficiency. Developing antibody markers of the on-target lineage would also be beneficial to future studies for assessing/enriching samples prior to sequencing.

Author Response to comment 1: Based on the single-cell RNA sequencing data and phenotypic staining data collected from iPSC-chondrocytes, these cells represent an early differentiation stage distinct from MSCs and with some similarity to mature chondrocytes. We agree with the points listed here about the potential for differentiation of more mature chondrocytes in this system. However, the relative immaturity of iPSC-chondrocytes as they currently stand does not necessarily hamper their potential to yield useful insight into human joint health and disease. For instance, prior research has found that studying cells representing transient developmental timepoints can uncover forms of gene regulation that would be invisible to studies of terminal cell types (Strober et al 2019, Elorbany et al, 2022).

Reviewer comment 2: I am concerned that the NA18856 cells appear to be more dramatically changed by the cyclic tensile strain on the axes of PC1 and especially PC2 than the other cell lines. This is dismissed as "treatment shows a minor correlation with PC2 (R2 = 0.14), PC3 (R2 = 0.14), and PC4 (R2 = 0.3)", but I wonder if this result implies concerted individual variability across many genes in this system, rather than random variability in responses across individuals at specific loci. It would help to also show the samples plotted along PC3 and PC4 to evaluate if the response of this single individual to treatment is driving the correlations, and to discuss possible implication of a single individual having distinct concerted responses.

Author Response to comment 2: While the small scale of this study makes it difficult to generalize to other individuals not included in the study, we anticipate that there will be some degree of heterogeneity in the severity of response to CTS in a larger group of individuals. We also think it is more likely that NA18856 does not represent concerted individual variability across many genes specific to this one individual but rather a more severe response that would be found across other individuals. Therefore, we do not expect this heterogeneity to limit our ability to map eQTLs in this system, especially if many other individuals show a similar large response to the strain treatment. We present PCA plots of samples plotted along additional PCs beyond PC1 and 2 in the Extended Data figures (Extended Data Figures 7 and 9), and in an updated version of this text, we include more details about the implications of a subset of individuals having strong responses to CTS in the Discussion section.

Reviewer Minor comment 1: This claim in the introduction is too strong "Through this study, we evaluated whether the process of chondrocyte differentiation is robust to individual differences" given that only used 3 individuals were used in the present study.

Author Response to Minor comment 1: Thank you for the suggestion. We modified the sentence to read: "Through this study, we evaluated whether the process of chondrocyte differentiation is robust to individual differences between three individuals."

Reviewer Minor comment 2: The authors could cite analyses of human sequence variation in chondrocyte regulatory elements: e.g. Richard *et al.* (2020).[1]

Author Response to Minor comment 2: Thank you for the suggestion. We mention prior studies of known variation in chondrocyte regulatory regions in an updated version of the text.

Reviewer Minor comment 3: For the topic modeling, this is discussed in the text, but it would be helpful to show an additional panel in the figure or supplement on the key genes in each topic and their preservation, especially for markers characteristic of adult chondrocytes.

Author Response to Minor comment 3: We present a heatmap of word probabilities of selected marker genes for different cell types across the topic model and volcano plots for relative occurrence of genes in each topic compared to all other topics presented in the Extended Data figures (Extended Data Figure 5). We emphasize the availability of these figures in an updated version of the main text.

Reviewer Minor comment 4: The authors could consider presenting Figure 5 before Figure 3 since Figure 5 makes it clear why using a linear model to account for the effects of individual, sex, etc is a useful approach for discovering the effect of treatment in this dataset.

Author Response to Minor comment 4: Thank you for the suggestion. We agree that the discussion of the sources of variation in the data does help to justify the linear model. However, because we want to conclude our paper with an emphasis on the potential for this system for future studies of gene expression in human joint cells, we feel that concluding with this figure is more impactful than presenting it earlier in the paper.

References:
Strober BJ, Elorbany R, Rhodes K, Krishnan N, Tayeb K, Battle A, Gilad Y. Dynamic genetic regulation of gene expression during cellular differentiation. Science. 2019 Jun 28;364(6447):1287-1290. doi: 10.1126/science.aaw0040. PMID: 31249060; PMCID: PMC6623972.

Elorbany R, Popp JM, Rhodes K, Strober BJ, Barr K, Qi G, Gilad Y, Battle A. Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during

human cardiomyocyte differentiation. PLoS Genet. 2022 Jan 21;18(1):e1009666. doi: 10.1371/journal.pgen.1009666. PMID: 35061661; PMCID: PMC8809621.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com   F1000 Research