TITLE

Integrated molecular phenotyping identifies genes and pathways disrupted in osteoarthritis

AUTHORS

Graham R. S. Ritchie[1,2,3,4], Theodoros I. Roumeliotis[1], Julia Steinberg[1], Abbie L. A. Binch[5], Rachael Coyle[1], Mercedes Pardo[1], Christine L. Le Maitre[5], Jyoti S. Choudhary[1], J. Mark Wilkinson[6#], Eleftheria Zeggini[1#]

AFFILIATIONS

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

[2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

[3]Usher Institute of Population Health Sciences & Informatics, University of Edinburgh, Edinburgh, EH16 4UX, UK

[4]MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

[5]Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, S1 1WB, UK

[6]Department of Oncology and Metabolism, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, UK

[#]Correspondence to Eleftheria@sanger.ac.uk or j.m.wilkinson@sheffield.ac.uk

ABSTRACT

Osteoarthritis (OA) is a degenerative joint disease with substantial global health economic burden and no curative therapy. Here we investigate genes and pathways underpinning disease progression, combining methylation typing, RNA sequencing and quantitative proteomics in chondrocytes from matched damaged and healthy articular cartilage samples from OA patients undergoing knee replacement surgery. Our data highlight 49 genes differentially regulated at multiple levels, identifying 19 novel genes with a potential role in OA progression. An integrated pathway analysis identifies established and emerging biological processes. We perform an *in silico* search for drugs predicted to target the differentially regulated factors and identify several established therapeutic compounds which now warrant further investigation in OA. Overall this work provides a first integrated view of the molecular landscape of human primary chondrocytes and offers insights into the mechanisms of cartilage degeneration. The results point to new therapeutic avenues, highlighting the translational potential of integrated functional genomics. All data from these experiments are freely available for access in the appropriate repositories.

INTRODUCTION

Osteoarthritis (OA) affects over 40% of individuals over the age of 70 (1), and is a leading cause of pain and loss of physical function (2). There is no treatment for OA; instead, disease management targets symptom control and culminates in joint replacement surgery. OA is a complex disease, with both heritable and environmental factors contributing to susceptibility (3). Despite the increasing prevalence, morbidity, and economic impact of the disease (1,2,4), the underlying molecular mechanisms of OA pathogenesis and progression remain incompletely characterized and the mainstay of treatment for advanced disease is still total joint replacement. The reasons for the limited success in unraveling the pathogenesis of OA is due, in part, to reductionist approaches applied in models that do not accurately recapitulate the clinical disease suffered by patients (5).

Emergent high throughput technologies and bioinformatics analyses of clinical tissues offer the promise of novel functional approaches to disease characterization and therapeutic target and biomarker discovery in complex diseases like OA. In recent years, studies individually examining gene expression, DNA CpG methylation and proteomics have expanded our understanding of OA pathogenesis, reviewed in (6,7). OA is primarily characterized by cartilage degeneration, thought to be brought about by an imbalance between anabolic and catabolic processes through a complex network of proteins including proteases and cytokines, reviewed in (8-10). Here we report the first application of integrated omics, including DNA methylation, RNA sequencing and quantitative proteomics from joint tissue to obtain a comprehensive molecular portrait of diseased versus healthy knee cartilage in OA patients (Figure 1a). Our data highlight both familiar and less well-established disease processes with involvement across multiple omics levels, and reveal novel candidate molecular players with therapeutic or diagnostic potential.

RESULTS

We extracted cartilage and subsequently isolated chondrocytes from the knee joints of 12 OA patients undergoing total knee replacement surgery. We obtained two cartilage samples from each patient, scored using the OARSI cartilage classification system (11,12): one 'high-grade degenerate' sample which we classified as diseased, and one 'normal' or 'low-grade degenerate' sample which we classified as healthy (Supplementary Figure 1). We compared the healthy and diseased tissue across patient-matched samples.

Quantitative proteomics

We used isobaric labeling liquid chromatography-mass spectrometry (LC-MS) to quantify the relative abundance of 6540 proteins that mapped to unique genes. We identified 209 proteins with evidence of differential abundance (Supplementary Table 1); ninety were found at higher abundance in the diseased samples, and 119 were found at lower abundance. For two representative patients we also used an orthogonal label-free approach to confirm the protein quantification data (Supplementary Figure 2). This is the most comprehensive differential proteomics study on OA samples to date. One of the most strongly down-regulated proteins is hyaluronan and proteoglycan link protein 1 (HAPLN1) , which binds hyaluronic acid in the extracellular matrix. HAPLN1 was found to be more abundantly released to the culture media by OA cartilage explants compared to healthy control tissue (13). Intra-articular injection of hyaluronic acid is one of the few current targeted treatments for OA pain (14). Based on UniProt annotation (15) we found a number of proteoglycans, cartilage and chondrocyte function-related proteins as well as basement membrane proteins consistently at lower abundances in the diseased samples (Supplementary Table 2). This is potentially a reflection of the increased cartilage catabolism that occurs in OA. As is the case for HAPLN1, several of these are found increasingly in the media released by OA tissue or synovial fluid from OA patients (13,16).

We validated the higher levels in diseased cartilage of four of the differentially abundant proteins (ANPEP, AQP1, TGFBI and WNT5B) by Western blotting (Supplementary Figure 6). ANPEP (aminopeptidase E) is a broad specificity aminopeptidase that has previously been detected in the synovial fluid of OA patients (17), and therefore has potential as a novel OA biomarker. TGFBI/BGH3 is an adhesion protein induced by TGF-β that binds to ECM proteins, integrins and to periostin, another up-regulated protein in OA which promotes cartilage degeneration through WNT signaling (18,19). TGFBI protein is released by OA tissue explants (13) and is also found in the synovial fluid of OA patients, but is more abundant in that of rheumatoid arthritis patients (16). WNT5B, a ligand for frizzled receptors in the WNT signaling pathway, has previously been reported to be differentially transcribed in osteoarthritic bone (20).

RNA sequencing

We sequenced total RNA from all samples and identified 349 genes differentially expressed at a 5% false discovery rate (FDR) (Supplementary Figure 3). Of these, 296 and 53 genes demonstrated higher transcription levels in the diseased and healthy samples, respectively (Supplementary Table 3). Most of these genes are annotated as protein-coding, but the shortlist also included several species of non-coding genes including 7 long intergenic non-coding RNA genes, 3 microRNA genes, and 7 annotated pseudogenes. One of the most strongly down-regulated genes in chondrocytes isolated from damaged sites was *CHRDL2*, the presence of which was further confirmed by immunohistochemistry (IHC) (Supplementary Figure 4). This gene was also found at lower abundance in the proteomics data, and is a bone morphogenetic protein (BMP) inhibitor that has previously been reported to be lost from chondrocytes of the superficial zone and shifted to the middle zone

in OA cartilage in a targeted study (21). We also identify *BMP1* as up-regulated at the RNA level, suggesting dysregulation of bone morphogenetic protein pathways in OA.

Among the 209 proteins with evidence of differential abundance in the proteomics data, 31 were also included in the set of genes identified as differentially expressed at the RNA level (hypergeometric p=5.3e-7, Figure 1b). Of these, 26 genes showed concordant directions of effect between diseased and healthy samples (binomial p=0.0002), while the direction differed for 5 genes (*COL4A2*, *CXCL12*, *FGF10*, *HTRA3* and *WNT5B*). In all five cases the gene was found to be over-expressed at the RNA level and less abundant at the protein level in the diseased tissue. Using annotations from the Human Protein Atlas (22) we found that all five proteins encoded by these genes are annotated as predicted secreted proteins, suggesting these proteins might be increasingly released or secreted in OA. In agreement with this, several collagens are more abundantly released into the culture media from diseased tissue than from healthy tissue (13). These proteins have potential value as biomarkers of OA.

DNA methylation

We used the Illumina 450k methylation array to assay ~480k CpG sites across the genome. We first performed a probe level analysis and identified 9,896 differentially methylated probes (DMPs) at 5% FDR (Supplementary Table 4). CpG methylation is regionally correlated and we also sought to identify differentially methylated regions (DMRs) that may have more biological relevance. This analysis yielded 271 DMRs, associated with 296 unique overlapping genes (Supplementary Table 5). Sixteen of the genes with an associated DMR were also among the list of differentially transcribed genes, including 2 members of the ADAMTS family: *ADAMTS2* and *ADAMTS4*. Both genes are up-regulated in diseased samples based on the RNA sequencing data, and were consistently associated with hypo-methylated DMRs.

We also identified several other members of this family to be either transcriptionally up-regulated (*ADAMTS12* and *ADAMTS14*) or associated with a hypo-methylated DMR (*ADAMTS17*). We found no evidence for differential abundance of the associated proteins, possibly because they would have been secreted into the matrix. These genes encode peptidases that catabolise components of the extracellular matrix, including aggrecan, which was the most down-regulated protein in diseased samples. Previous studies have implicated this gene family in OA (23) and have suggested that *ADAMTS4*-mediated aggrecan degradation may be an important process (24). Accordingly, aggrecan is more abundant in OA synovial fluid (16).

Integrative analyses

To investigate the concordance between the transcriptome and proteome assays, we computed the global correlation in all samples irrespective of tissue status, comparing the RNA fragments per kilobase of transcript per million fragments mapped (FPKM) (25) to normalised peptide spectral counts (Supplementary Figure 5a). We found a significant positive correlation (Spearman's rho=0.29, $p<2.2e-16$) between RNA expression levels and protein abundance. To establish if there were also concordant differences in RNA and protein abundance according to tissue grade, we computed the correlations between RNA and protein ratios between diseased and healthy samples (Figure 2a). When considering all genes with data obtained from both assays, we identified a significant, but small, positive correlation (Pearson's r=0.17, $p<2.2e-16$). The magnitude of correlation became substantially stronger when we only considered the 31 genes that were expressed differentially in both datasets (Pearson's r=0.43, p=0.01).

To investigate the concordance between the methylome and transcriptome data, we compared the aggregate methylation status of promoter region CpG probes with

transcription levels in all samples irrespective of tissue grade (Methods, Supplementary Table 6) and found the expected negative correlation between promoter region methylation and gene expression (Spearman's rho=-0.43, p<2.2e-16, Supplementary Figure 5b). Comparison of the changes in RNA expression to the differences in promoter-region methylation values when comparing samples according to tissue grade demonstrated a small but significant correlation (Pearson's r=-0.08, p<2.2.e-16, Figure 2b). Taking only those genes called as significantly changing at both the methylome and transcriptome level, the correlation increased in magnitude (Pearson's r=-0.48, p=0.002).

We identified 49 genes with evidence of differential regulation in chondrocytes extracted from diseased vs healthy cartilage from at least two of the three omics analyses (Supplementary Table 7). We identify three genes consistently across all 3 approaches: *AQP1*, *COL1A1* and *CLEC3B* (Figure 1b). All 3 genes were identified as up-regulated in diseased tissue in both the RNA-seq and proteomics analyses (Figure 2a), and as having associated DMRs. *AQP1* and *COL1A1* showed a consistent decrease in methylation of all CpG probes in their associated DMRs, commensurate with an increase in transcription, while the DMR associated with *CLEC3B* showed evidence of increased methylation (Supplementary Table 5). Using IHC we independently confirmed the presence of all 3 proteins within the chondrocytes in cartilage samples (Supplementary Figure 4). These three genes have previously been implicated in OA. *AQP1*, encoding aquaporin-1, is a member of a family of proteins that facilitate water transport across biological membranes. Chondrocyte swelling and increased cartilage hydration has been suggested as an important mechanism in OA (26). Accordingly, *AQP1* has been observed as over-expressed in a rat model of knee OA (27), and there is one previous report of transcriptional over-expression in knee OA in humans (28). *CLEC3B* (also known as *TNA*) encodes the protein tetranectin, which binds human tissue plasminogen activator (tPA) (29). Previous studies have identified *CLEC3B* as up-regulated in

human OA (30,31), mediating extracellular matrix destruction in cartilage and bone (32), and a candidate gene association study found evidence of association of a coding variant (rs13963, Gly106Ser) in *CLEC3B* with OA (33) (although this association has not been replicated in subsequent studies (34)). *COL1A1* is one of several collagen proteins (including *COL1A2*, *COL3A1*, *COL4A2*, *COL5A1*, Figure 2a), which we identified as differentially regulated in the diseased samples using both RNA-seq and proteomics. Collagens are the main structural components of cartilage and several studies have highlighted the importance of collagen dysregulation in OA (10,35). A recent study also identified up-regulation of *COL1A1* and *COL5A1* in synovium from humans with end-stage OA, in the synovium of mice with induced OA, and in human fibroblasts stimulated with TGF-β (36).

Of the 49 genes with evidence of differential regulation on at least two molecular levels, 19 genes (39%) have not previously been implicated in OA (Supplementary Table 7). Novel genes with convergent evidence include *MAP1A* and *MAP1B*, encoding microtubule-associated proteins, both of which were up-regulated significantly at both the RNA and protein levels (Figure 2a). These proteins are expressed mostly in the brain and are involved in regulation of the neural cytoskeleton (37). Cytoskeletal regulation is thought to be an important process in OA (38) and accordingly, recent studies have implicated these proteins in bone formation (39). The *PXDN* gene was up-regulated at the RNA level and associated with 2 hypo-methylated DMRs. *PXDN* encodes peroxidasin, which is secreted into the extracellular matrix and catalyses collagen IV cross-linking (40). We also identify 2 sub-units of collagen IV, *COL4A1* and *COL4A2*, as differentially regulated at several levels (Supplementary Table 7). Although many collagens have been found to be differentially regulated in OA, *COL4A2* has not been reported previously in association with OA. Collagen IV is the major structural constituent of basement membranes, but it has also been identified at the articulating surface of normal and osteoarthritic articular cartilage (41).

HTRA3, up-regulated at the RNA level, but found at lower abundance at the protein level possibly due to increased secretion, is a member of the HTRA family of serine proteinases involved in cartilage degradation and tissue turnover, important processes in OA progression (42). Another member of this family, *HTRA1*, is transcriptionally up-regulated in OA cartilage (43) and has also been implicated in the alteration of chondrocyte metabolism by disrupting the pericellular matrix (44). Genes with relatively little characterization include *CRTAC1*, encoding cartilage acidic protein 1 which is secreted by chondrocytes (45), and *PODN*, encoding podocan, which binds collagen in the extracellular matrix (46) . Our combined epigenetic, transcriptomic and proteomic analysis has uncovered a substantial number of genes associated with OA progression, some of which have known connections to cartilage or bone-related processes, and others with no links, whose detailed characterization should bring new insights into the molecular mechanisms of OA pathogenesis.

Gene set analyses

We performed a gene set enrichment analysis on the shortlisted genes from each separate omics analysis and found that several common biological processes are highlighted at multiple levels (Supplementary Tables 8 & 9, Methods). To identify pathways jointly affected by genes identified at multiple molecular levels, we used the geometric mean of the three enrichment *p*-values from each analysis, and calculated an empirical *p*-value for this statistic with a permutation strategy accounting for the overlap (Methods). We identified 19 enriched pathways from KEGG & Reactome at a combined 5% FDR, and 30 GO annotations (Supplementary Tables 8 & 9). Of these enriched gene sets, 7 from KEGG & Reactome and 11 from GO contain at least five significant genes from each of at least two omics approaches (Figure 3, Supplementary Figure 7).

A common theme in the highlighted pathways is cartilage matrix regulation and degeneration, in agreement with the notion that increased ECM turnover is a crucial component in OA pathogenesis. Pathways including “extracellular matrix organisation” and “collagen formation” were affected by genes identified by all three omics analyses, although the component genes do not all overlap (Figure 3). Results from the 3 analyses converge on shared mechanisms, supporting the importance of utilising evidence from an integrated perspective. The GO term analysis uncovered consistent evidence from all three omics assays for genes annotated with the terms “extracellular matrix disassembly” and “collagen catabolic process”. In these pathways we also find suggestive evidence of a link to genetic OA risk loci. These signals would not have been identified directly from GWAS data (Supplementary Results), highlighting the importance of synthesizing data at multiple molecular levels to obtain a more powerful integrated view.

Further interesting pathways and biological processes enriched at multiple levels were “positive regulation of ERK1/2 cascade”, “heparin-binding”, “platelet activation”, all of which are interconnected through common genes. Several studies have linked the extracellular signal-regulated kinase (ERK) cascade to OA (8,47-50). Heparin-binding growth factors have also been shown to be involved in OA (51-54), some in particular through activation of the ERK signaling pathway. Injection of platelet-rich plasma in OA knees leads to significant clinical improvement (55,56) and there is evidence to suggest that this effect is mediated via the ERK cascade (57). Our findings provide strong evidence supporting a role for this pathway in OA pathogenesis.

We also found enrichment of genes involved in the regulation of angiogenesis at multiple levels. The growth of blood vessels and nerves are closely linked processes that share regulatory mechanisms, including the ERK cascade and heparin-binding proteins mentioned

above (58). Accordingly, pathways like NCAM signaling for neurite outgrowth and PDGF signaling that play a significant role in blood vessel and nervous system formation were highlighted by the pathway analysis. To investigate this further we used the Human Protein Atlas to annotate the protein-coding genes identified in the RNA-seq and proteomics experiments, and found a significant enrichment in plasma proteins (RNA-seq hypergeometric p=6.9e-11, proteomics p=1.8e-5). This supports a role for angiogenesis and nerve growth in OA progression (58,59). Indeed, histological examination within the samples we investigated showed greater blood vessel ingrowth in tissues with more advanced OA (Supplementary Figure 1). Results from the three molecular analyses converge on shared biological mechanisms that are relevant to the pathogenesis of OA, supporting the importance of utilizing evidence from an integrated perspective. These data should be useful in pinpointing candidate targets to help improve therapeutic intervention.

***In silico* screen for new OA modifying drugs**

The molecular signatures highlighted above provide novel investigative candidates as prognostic biomarkers for OA and point to novel therapeutic opportunities. To identify existing drugs that could be applied to OA, we searched Drugbank (60) using the 49 differentially regulated genes identified by at least two of the functional genomics approaches. We uncovered 29 compounds with investigational or established actions on the corresponding proteins. After filtering to include only agents with current Food and Drug Administration Marketing Authorization for use in humans, we identified ten agents with actions on nine of the dysregulated proteins (Table 1). These agents cover a broad range of mechanisms of action and represent novel investigational targets for 'first in disease' studies of OA progression. These drugs have established safety profiles and pharmacokinetic data for use in man, which would shorten the investigative pipeline to clinical use in OA. One of the identified group of ten agents, those active against prostacyclin synthase (NSAIDs),

already have marketing authorization for the symptomatic treatment of OA. Another drug identified in this search was phylloquinone (vitamin K_1), an agonist of osteocalcin (BGLAP). Periostin, a protein with elevated expression in OA, in this study and others (13,18,61), is a vitamin K-dependent protein that induces cartilage degeneration (62). Interestingly, a recent study has associated sub-clinical vitamin K deficiency with knee OA incidence (63), warranting further investigation of this compound as a disease-modifying agent in OA. Thus, our work could help prioritize the repurposing of existing drugs for the treatment of OA.

DISCUSSION

Although previous studies have individually investigated methylation (64-66), transcription (31,67), and protein expression (7,68) in OA tissue, these results provide the first integrated, systematic and hypothesis-free analysis of the biological changes involved in human OA progression at all three molecular levels. Using this multi-level functional genomics approach, we have provided a first integrated view of the molecular alterations that accompany cartilage changes resulting in debilitating joint disease. We also highlight the clinical translation implications for drug repurposing to slow OA progression. Here we have focused on OA and demonstrate the potential of multi-omics approaches using a relatively small sample set. Larger sample sizes will be required for a more powerful characterisation of the disease progression-related molecular landscape changes. The integrative functional genomics approach illustrated here offers an opportunity to identify molecular signatures in disease-relevant tissues, thereby gaining insights into disease mechanism, identifying potential biomarkers, and discovering druggable targets for intervention. A key future challenge will be the development of powerful statistical approaches for the integration of high-dimensional molecular traits in the context of complex diseases. All data arising from the experiments described here are freely available to researchers in the appropriate online repositories.

METHODS

Patient consent & study approval

All subjects provided written, informed consent prior to participation in the study. Tissue samples were collected between October 2013 and February 2014 under Human Tissue Authority license 12182, Sheffield Musculoskeletal Biobank, University of Sheffield, UK. All samples were collected from patients undergoing total knee replacement for primary osteoarthritis. Patients with diagnosis other than osteoarthritis were excluded from the study. The study was approved by Oxford NHS REC C (10/H0606/20).

Sample processing

Extraction of chondrocytes from osteochondral tissue taken at knee replacement

Osteochondral samples were transported in Dulbecco's modified Eagle's medium (DMEM)/F-12 (1:1) (Life Technologies) supplemented with 2mM glutamine (Life Technologies), 100 U/ml penicillin, 100 µg/ml streptomycin (Life Technologies), 2.5 µg/ml amphotericin B (Sigma-Aldrich) and 50 µg/ml ascorbic acid (Sigma-Aldrich) (serum free media). Half of each sample was taken for chondrocyte extraction and the remaining tissue was fixed in 10% neutral buffered formalin, decalcified in surgipath decalcifier (Leica) and embedded to paraffin wax for histological and immunohistochemical analysis. Chondrocytes were directly extracted from each paired macroscopic control and OA grade cartilage in order to remove the extracellular matrix allowing a higher yield of cells to be loaded onto the Qiagen column.

Cartilage was removed from the bone, dissected and washed twice in 1xPBS. Tissue was digested in 3 mg/ml collagenase type I (Sigma-Aldrich) in serum free media overnight at 37°C on a flatbed shaker. The resulting cell suspension was passed through a 70 µm cell strainer (Fisher Scientific) and centrifuged at 400g for 10 minutes; the cell pellet was then washed twice in serum free media, followed by centrifugation at 400g for 10 minutes. The resulting

cell pellet was resuspended in serum free media. Cells were counted and the viability checked using trypan blue exclusion and the Countess cell counter (Invitrogen). The optimal cell number for Qiagen column extraction from cells is between $4x10^6$ and $1x10^7$. Cells were pelleted at 400g for 10 minutes and homogenized in Qiagen RLT buffer containing β-Mercaptoethanol and using the QIAshredder column and DNA, RNA and protein extractions were performed as outlined for tissue extraction. RNA, DNA and protein were quantified using a Nanodrop.

Histological examination

Four micron sections of paraffin-embedded cartilage tissue were mounted onto positively charged slides. Sections were dewaxed in Sub-X, rehydrated in IMS, washed in distilled water, stained in 1% w/v Alcian blue/glacial acetic acid (pH 2.4) for 15 minutes, counter stained in 1% w/v aqueous neutral red for 1 minute or stained with Masson Trichrome (Leica) according to the manufacturer's instructions. Sections were dehydrated and mounted. Cartilage tissue was graded using the Mankin Score (0-14) with additional scores for abnormal features (0-4) and cartilage thickness (0-4) based on the OARSI scoring system (11,12). The total scores were used to determine the overall grade of the cartilage as low-grade (median: 4.5; IOR: 3-5.5; n=12) which we define as 'healthy', or high grade degenerate (median: 14; IOR: 11.75-18; n=12), which we define as 'diseased'.

Proteomics

Protein Digestion and TMT Labeling

The protein content of each sample was precipitated by the addition of 30 µL TCA 8 M at 4 °C for 30 min. The protein pellets were washed twice with ice cold acetone and finally re-suspended in 40 µL 0.1 M triethylammonium bicarbonate, 0.05% SDS with pulsed probe sonication. Protein concentration was measured with Quick Start Bradford Protein Assay

(Bio-Rad) according to manufacturer's instructions. Aliquots containing 30 µg of total protein were prepared for trypsin digestion. Cysteine disulfide bonds were reduced by the addition of 2 µL 50 mM tris-2-carboxymethyl phosphine (TCEP) followed by 1 h incubation in heating block at 60 °C. Cysteine residues were blocked by the addition of 1 µL 200 mM freshly prepared Iodoacetamide (IAA) solution and 30 min incubation at room temperature in dark. Trypsin (Pierce, MS grade) solution was added at a final concentration 70 ng/µL to each sample for overnight digestion. After proteolysis the peptide samples were diluted up to 100 µL with 0.1 M TEAB buffer. A 41 µL volume of anhydrous acetonitrile was added to each TMT 6-plex reagent (Thermo Scientific) vial and after vortex mixing the content of each TMT vial was transferred to each sample tube. Labeling reaction was quenched with 8 µL 5% hydroxylamine for 15 min after 1 h incubation at room temperature. Samples were pooled and the mixture was dried with speedvac concentrator and stored at -20 °C until the high-pH Reverse Phase (RP) fractionation.

Peptide fractionation

Offline peptide fractionation based on high pH Reverse Phase (RP) chromatography was performed using the Waters, XBridge C18 column (2.1 x 150 mm, 3.5 µm, 120 Å) on a Dionex Ultimate 3000 HPLC system equipped with autosampler. Mobile phase (A) was composed of 0.1% ammonium hydroxide and mobile phase (B) was composed of 100% acetonitrile, 0.1% ammonium hydroxide. The TMT labelled peptide mixture was reconstituted in 100 µL mobile phase (A), centrifuged and injected for fractionation. The multi-step gradient elution method at 0.2 mL/min was as follows: for 5 minutes isocratic at 5% (B), for 35 min gradient to 35% (B), gradient to 80% (B) in 5 min, isocratic for 5 minutes and re-equilibration to 5% (B). Signal was recorded at 280 nm and fractions were collected in a time dependent manner every one minute. The collected fractions were dried with SpeedVac concentrator and stored at -20 °C until the LC-MS analysis.

LC-MS Analysis

LC-MS analysis was performed on the Dionex Ultimate 3000 UHPLC system coupled with the high-resolution LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). Each peptide fraction was reconstituted in 40 µL 0.1% formic acid and a volume of 10 µL was loaded to the Acclaim PepMap 100, 100 µm × 2 cm C18, 5 µm, 100 Å trapping column with a user modified injection method at 10 µL/min flow rate. The sample was then subjected to a multi-step gradient elution on the Acclaim PepMap RSLC (75 µm × 50 cm, 2 µm, 100 Å) C18 capillary column (Dionex) retrofitted to an electrospray emitter (New Objective, FS360-20-10-N-20-C12) at 45 °C. Mobile phase (A) was composed of 96% H_2O, 4% DMSO, 0.1% formic acid and mobile phase (B) was composed of 80% acetonitrile, 16% H_2O, 4% DMSO, 0.1% formic acid. The gradient separation method at flow rate 300 nL/min was as follows: for 95 min gradient to 45% B, for 5 min up to 95% B, for 8 min isocratic at 95% B, re-equilibration to 5% B in 2 min, for 10 min isocratic at 5% B.

The ten most abundant multiply charged precursors within 380 -1500 *m/z* were selected with FT mass resolution of 30,000 and isolated for HCD fragmentation with isolation width 1.2 Th. Normalized collision energy was set at 40 and the activation time was 0.1 ms for one microscan. Tandem mass spectra were acquired with FT resolution of 7,500 and targeted precursors were dynamically excluded for further isolation and activation for 40 seconds with 10 ppm mass tolerance. FT max ion time for full MS experiments was set at 200 ms and FT MSn max ion time was set at 100 ms. The AGC target vales were 3×10e6 for full FTMS and 1×10e5 for MSn FTMS. The DMSO signal at m/z 401.922718 was used as a lock mass.

Database Search and Protein Quantification

The acquired mass spectra were submitted to SequestHT search engine implemented on the Proteome Discoverer 1.4 software for protein identification and quantification. The precursor mass tolerance was set at 30 ppm and the fragment ion mass tolerance was set at 0.02 Da. TMT6plex at N-termimus, K and Carbamidomethyl at C were defined as static modifications. Dynamic modifications included oxidation of M and Deamidation of N,Q. Maximum two different dynamic modifications were allowed for each peptide with maximum two repetitions each. Peptide confidence was estimated with the Percolator node. Peptide FDR was set at 0.01 and validation was based on q-value and decoy database search. All spectra were searched against a UniProt fasta file containing 20,190 Human reviewed entries. The Reporter Ion Quantifier node included a custom TMT 6plex Quantification Method with integration window tolerance 20 ppm and integration method the Most Confident Centroid. For each identified protein a normalized spectral count value was calculated for each one of the 6-plex experiments by dividing the number of peptide spectrum matches (PSMs) of each protein with the total number of PSMs. Median normalized spectral counts per protein were computed across the different multiplex experiments.

Differential abundance

To identify those proteins with evidence of differential expression, we shortlisted proteins with absolute median abundance ratios between diseased and healthy samples >= 0.75, where the median abundance ratio was greater than the standard deviation in all samples with data, and with evidence from at least 5 patients. This analysis identified 209 proteins (Supplementary Table 1).

Western blotting

Sample pairs were adjusted to the same protein concentration. Twenty micrograms of protein per sample were electrophoresed on 4-12% Bis-Tris NuPAGE gels (Life Technologies) and transferred to nitrocellulose membranes. Primary antibodies used were as follows: ANPEP, ab108382; AQP1, ab168387; COL1A, ab14918; TGFB1, ab89062; WNT5B, ab124818 (Abcam); GAPDH, sc-25778 (Santa Cruz Biotechnologies). Chemiluminescence detection was carried out using ECL Prime (GE Healthcare) or ECL Ultra (Lumigen) and ImageQuant LAS1400 (GE Healthcare). Densitometry was performed with ImageQuant Tool Box (GE Healthcare). Intensity values were normalised to GAPDH loading control before ratio calculation.

Label free quantification of representative samples

For a selection of four representative control and disease samples, peptide aliquots of 500ng without TMT labelling were analysed on the Dionex Ultimate 3000 UHPLC system coupled with the Orbitrap Fusion (Thermo Scientific) mass spectrometer for label free quantification and validation. Tandem mass spectra were acquired over duplicate runs of 120 min with a top speed iontrap detection method and dynamic exclusion at 10 sec and MS R=120,000. Database search was performed on Proteome Discoverer 1.4 with the SequestHT engine and normalized spectral counts were computed based on the total number of peptide-spectrum matches attributed to each protein per sample divided by the maximum value along the different samples. With a minimum requirement of at least total 14 spectra per protein we found excellent agreement in the direction of change between isobaric labelling and label free quantification for at least 32 proteins which is approximately 90% of the common proteins between the TMT changing list and the label free identified list (Supplementary Figure 2).

RNA-seq

RNA sequencing

Using Illumina's TruSeq RNA Sample Prep v2 kits, poly-A tailed RNA (mRNA) was purified from total RNA using an oligo dT magnetic bead pull-down. The mRNA was then fragmented using metal ion-catalyzed hydrolysis. A random-primed cDNA library was then synthesised and this resulting double-strand cDNA was used as the input to a standard Illumina library prep: ends were repaired with a combination of fill-in reactions and exonuclease activitiy to produce blunt ends. A-tailing was performed, whereby an "A" base was added to the blunt ends followed by ligation to Illumina Paired-end Sequencing adapters containing unique index sequences, allowing samples to be pooled. The libraries then went through 10 cycles of PCR amplification using KAPA Hifi Polymerase rather than the kit-supplied Illumina PCR Polymerase due to better performance.

Samples were quantified and pooled based on a post-PCR Agilent Bioanalyzer, then the pool was size-selected using the LabChip XT Caliper. The multiplexed library was then sequenced on the Illumina HiSeq 2000, 75bp paired-end read length. Sequenced data was then analysed and quality controlled (QC and individual indexed library BAM files were produced.

Read alignment

The resulting reads that passed QC were realigned to the GRCh37 assembly of the human genome using a splice-aware aligner, bowtie version 2.2.3 (69), and using a reference transcriptome from Ensembl release 75 (70), using the –library-type option fr-firststrand to bowtie. We limited the alignments to uniquely mapping reads. We then counted the number of reads aligning to each gene in the reference transcriptome using htseq-count from the HTSeq package(71) separately for each sample to produce a read count matrix counting the number of reads mapping to each gene in the transcriptome for each sample. To quantify absolute transcript abundance we computed the fragments per kilobase of transcript per

million fragments mapped (FPKM) (25) for each gene using the total read counts from this matrix, and the exonic length of each gene calculated from gene models from Ensembl release 75. We obtained a mean of 49.3 million uniquely mapping reads from each sample (range: 39.2 - 71.4 million) with a mean of 84% of reads mapping to genes (range: 67.9% - 90.6%) which were used for the differential expression analysis.

Differential expression analysis

We used edgeR version 3.0 (72) to identify differentially expressed genes from the read count matrix. We restricted the analysis to 15,418 genes with >1 counts per million in at least 3 samples (similar to the protocol described by Anders *et al.* (73)). We followed the processing steps listed in the manual, using a generalised linear model with tissue status (diseased or healthy) and individual ID as covariates. 349 genes were differentially expressed between the diseased and healthy samples at 5% FDR (296 up-, 54 down-regulated in diseased tissue). The genes differentially expressed at 5% FDR had somewhat higher exonic length than the remaining genes (Wilcox-test p=0.00013; 4804 vs 4153 bases), hence we adjusted for gene length in the randomisations for gene set analyses.

Methylation

Illumina 450k BeadChip assay

Sample submission: samples were tested for quality and then quantified to 50ng/ul by the onsite sample management team prior to submission to the Illumina Genotyping pipeline. Before processing begins, manifests for submitted samples are uploaded to Illumina LIMS where each sample plate is assigned an identification batch so that it can be tracked throughout the whole process that follows.

Bisulfite Conversion: Before Pre-Amplification sample DNA requires bisulfite conversion using the Zymo EZ-96 DNA Methylation assay. This is completed manually as per Zymo SOP guidelines.

Pre-Amplification: Due to the differences in sample plates between the completed Zymo assay and the Illumina assay, pre-Amplification is performed manually following the Illumina MSA4 SOP. Once complete, sample and reagent barcodes are scanned through the Illumina LIMS tracking software. Four micro-litres (200ng) of sample is required (Illumina guidelines) for the pre-Amplification reaction – there is no quantification step after the completion of the Zymo assay.

Post-Amplification: Over three days, Post-Amplification (Fragmentation, Precipitation, Resupension, Hybrisation to beadchip and xStaining) processes are completed as per Illumina protocol using four Tecan Freedom Evos. Following the staining process, BeadChips are coated for protection and dried completely under vacuum before scanning commences on five Illumina iScans, four of which are paired with two Illumina Autloader 2.Xs.

Image Beadchip: The iScan Control software determines intensity values for each bead type on the BeadChip and creates data files for each channel (.idat). Genomestudio uses this data file in conjunction with the beadpool manifest (.bpm) to analysis the data from the assay.

QC: Prior to downstream analysis, all samples undergo an initial QC to establish how successful the assay has performed. Intensity graphs in Genomestudio's Control Dashboard identify sample performance by measuring dependent and non-dependent controls that are manufactured onto each BeadChip during production.

Probe-level analysis

The intensity files for each sample were processed using the ChAMP package (74). Probes mapping to chromosomes X & Y, and those with a detection p value > 0.01 (n=3,064) were excluded. The beta values for each probe were quantile-normalised, accounting for the design of the array, using the 'dasen' method from the wateRmelon package (75). We also excluded any probes with a common SNP (minor allele frequency > 5%) within 2 base pairs of the CpG site, and those predicted to map to multiple locations in the genome (76) (n=45,218), leaving a total of 425,694 probes for the probe-level differential methylation analysis. We annotated all probes with genomic position, gene and genic location information from the ChAMP package.

To identify probes with evidence of differential methylation we used the CpGassoc package (77) to fit a linear model at each probe, with tissue status and individual ID as covariates. This analysis yielded 9,867 differentially methylated probes (DMP) between diseased and healthy samples at 5% FDR.

To identify differentially methylated regions, we used custom software (available upon request) to identify regions containing at least 3 DMPs and no more than 3 non-significant probes with no more than 1kb between each constituent probe, following previous analyses (66). We used bedtools (78) to identify genes overlapping each DMR, using gene annotations from Ensembl release 75, and extending each gene's bound to include 1500 basepairs upstream of the transcription start site to include likely promoter regions. This analysis yielded 271 DMRs with a mean of 4.04 DMPs per region, and a mean length of 673 basepairs.

Gene-level analysis

We assigned probes in the promoter region of each gene using the probe annotations from the ChAMP package, and assigned to each gene any probe with the annotation "TSS1500", "TSS200", "5'UTR" and "1stExon" in order to capture probes in likely promoter regions. We then computed the mean normalised beta value of assigned probes for all genes with at least 5 associated probes for each sample separately, to produce a single methylation value for each gene in each sample. We used a paired t-test to identify genes with differential promoter-region methylation between diseased and healthy samples, and a 5% FDR cutoff to call a gene's promoter region as differentially methylated. Note that the paired t-test assumes an equivalent model to the linear model used for the probe-level analysis.

Immunohistochemistry

To identify whether native chondrocytes demonstrated expression of the key factors immunohistochemistry was deployed. Four micron sections were dewaxed, rehydrated, and endogenous peroxidase blocked using 3% hydrogen peroxide for 30 minutes. After washing sections with dH_2O, antigens were retrieved in 0.01% w/v chymotrypsin/$CaCl_2$ (Sigma, UK), for 30 minutes at 37˚C. Following TBS washing, nonspecific binding sites were blocked at room temperature for 2 hours with either 25% w/v goat serum or rabbit serum (Abcam, UK) in 1% w/v bovine serum albumin (Sigma, UK) in TBS. Sections were incubated overnight at 4˚C with either mouse monoclonal primary antibodies or rabbit polyclonal antibodies. Negative controls in which rabbit and mouse IgGs (Abcam, UK) replaced the primary antibody at an equal protein concentration were used. Slides were washed in TBS and a biotinylated secondary antibody was applied; either goat anti-rabbit or rabbit anti-mouse, both antibodies were applied at 1:400 dilution in 1% w/v BSA/TBS for 30 minutes at room temperature. Binding of the secondary antibody was disclosed with streptavidin-biotin complex (Vector Laboratories, UK) technique with 0.08% v/v hydrogen peroxide in 0.65 mg/mL 3,3'-diaminobenzidine tetrahydrochloride (Sigma, UK) in TBS. Sections were

counterstained with Mayer's haematoxylin (Leica, UK), dehydrated, cleared and mounted with Pertex (Leica, UK). All slides were visualised using an Olympus BX60 microscope and images captured using a digital camera and software program QCapture Pro v8.0 (MediaCybernetics, UK).

Protein atlas annotation

We downloaded annotations from Human Protein Atlas version 13, and annotated each protein-coding gene from the 3 experiments with the following terms taken from the annotation file: “Predicted secreted protein”, “Predicted membrane protein”, “Plasma protein”. The secreted and membrane protein predictions are based on a consensus call from multiple computational prediction algorithms, and the plasma protein annotations are taken from the Plasma Protein Database, as detailed in Uhlen *et al*. (22).

Identification of previously reported OA genes

In order to identify whether some of the genes we highlight have previously been reported as associated with OA we searched PubMed in June 2015. We used an “advanced” search of the form “(osteoarthritis) AND (<gene_name>)” where <gene_name> was set to each HGNC gene symbol and we report the number of citations returned for each search.

Gene set analyses

Individual datasets

We aimed to test whether particular biological gene sets were enriched among the significant genes from each of the RNA-seq, methylation, and proteomics datasets. To this end, we downloaded KEGG (79) and Reactome (80) gene annotations from MSigDB (version 4) (81). We also downloaded Gene Ontology (GO) biological process and molecular function gene annotations from QuickGO (82) on 4 February 2015. For GO, we only considered

annotations with evidence codes IMP, IPI, IDA, IEP, and TAS. Genes annotated to the same term were treated as a "pathway". KEGG/Reactome and GO annotations were analysed separately and only pathways with 20 to 200 genes were considered (555 for KEGG/Reactome, 811 for GO). Enrichment was assessed using a 1-sided hypergeometric test and only considering genes with annotations from a particular resource. For example, among the 15418 genes with RNA sequencing data, 4787 genes had KEGG/Reactome annotations, and 65 genes were annotated to "extracellular matrix annotation" in KEGG. Of the 350 significantly differentially expressed genes, 134 had KEGG/Reactome annotations, and 12 genes were annotated to "extracellular matrix annotation". Consequently, the enrichment of "extracellular matrix annotation" genes among the differentially expressed genes was assessed by comparing 12 of 134 to 65 of 4787 genes. Multiple-testing was accounted for by using a 5% FDR (separately for KEGG/Reactome and GO, and for RNAseq, methylation, and protein expression data).

Empirical p-values for the enrichments were obtained from randomisations accounting for overlap of significant genes among the RNAseq, methylation, and protein expression datasets (see below).

Integrative gene set analyses

We aimed to integrate the gene sets analyses for the RNAseq, methylation, and protein expression datasets. For each gene set, we asked whether the association across the three datasets (calculated as geometric mean of the *p*-values) was higher than expected by chance. To this end, we obtained 1-sided empirical *p*-values from 100,000 sets of "random RNAseq genes, random methylation genes, and random protein expression genes". The "random" sets were chosen to conservatively match the overlap observed among the significant genes

as follows. We performed the randomisation separately for KEGG/Reactome and for GO, as we only considered genes with at least one annotation in the resource.

To jointly construct one set each of random RNAseq genes, random methylation genes, and random protein expression genes, we picked:

1) random genes for the overlap of RNAseq, methylation, and protein expression (KEGG/Reactome: 2; GO: 3);
2) additional random genes for the overlap of RNAseq and methylation (KEGG/Reactome: 4; GO: 10);
3) additional random genes for the overlap of RNAseq and protein expression (KEGG/Reactome: 13; GO: 21);
4) additional random genes for the overlap of methylation and protein expression (KEGG/Reactome: 2; GO: 5);
5) additional RNAseq random genes (KEGG/Reactome: 115; GO: 182);
6) additional methylation random genes (KEGG/Reactome: 73; GO: 102);
7) additional protein expression random genes (KEGG/Reactome: 62; GO: 113);

Random genes were picked to account for gene length as follows. In step 1, we subdivided all genes present in the RNAseq, methylation, and protein expression data into 100 bins by increasing exonic length. If the original significant genes in the overlap had g genes in a particular bin b, we picked g random genes from that same bin; this was done for all 100 bins. Steps 2 to 7 were done analogously.

We tested that 100 bins were enough: choosing 50 or 200 bins gave very similar results (Pearson correlation >0.99 for empirical p-values in all enrichment analyses). We also confirmed that 10,000 random gene sets were enough: repeating the analysis gave very

similar results (Spearman correlation >0.99 for empirical *p*-values in all enrichment analyses). To confirm the lower empirical *p*-values, we carried out 100,000 randomisations.

arcOGEN gene-set association analysis

We asked whether the 18 gene sets with strong evidence for association from the functional genomics data (Figure 4, Supplementary Figure 7) are also associated with OA based on GWAS. To this end, we used the arcOGEN GWAS, primarily the 3498 cases with knee OA and all 11009 controls. We assigned a SNP to a gene if it was located within the gene boundaries (Genome Assembly GRCh37). A SNP was assigned to a gene set if it was assigned to one of the genes in the given set.

First, we asked whether the average SNP *p*-value in a gene set was lower than expected by chance. We used the gene set test in plink, filtering independent SNPs at r^2=0.2, and 10000 case-control phenotype permutations to obtain empirical one-sided *p*-values. Five of the 18 gene sets had empirical *p*-values significant at 5% FDR (Supplementary Table 10). All of these five gene sets were also significantly associated at 5% FDR when considering all 7410 knee or/and hip OA cases and 11009 controls from arcOGEN, with similar results when filtering independent SNPs at r^2=0.5 (data not shown).

Second, we asked whether the results were confounded by population structure. To test this, we repeated the analysis accounting for population structure by using logistic regression with the 10 first principal components obtained from EIGENSTRAT when considering all 7410 cases and 11009 controls together with HapMap release 23a founder individuals. The results were as above (Supplementary Table 10).

Third, we asked whether the five gene sets with association in the first step were also associated compared to other gene sets, in particular, accounting for gene set size. We considered the 250 gene sets from GO, KEGG, and Reactome with the highest numbers of SNPs assigned. For each of the five highlighted gene sets, we chose 100 gene sets with the closest numbers of assigned SNPs. At least one in ten of the other gene sets had empirical *p*-values as low as the highlighted gene set (Supplementary Table 10).

arcOGEN hypothesis-free gene-set analysis

We also asked whether we would have identified the 18 gene sets if we had only considered the arcOGEN knee OA GWAS data in a hypothesis-free approach. We used two common methods – a gene-based overrepresentation test as analogue to the functional genomics work, and a direct set-based test as above.

First, we asked whether the gene sets highlighted from the functional genomics work are among the gene sets over-represented among the 25% genes with the lowest *p*-values. We calculated *p*-values for each gene using plink gene set analysis as above. No gene set enrichment was significant at 5% FDR when considering all GO gene sets, and, separately, all KEGG and Reactome gene sets. (When considering all arcOGEN cases and controls, one GO and five KEGG/Reactome gene sets were significant at 5% FDR; they do not overlap with any of the 18 gene sets highlighted from the functional genomics analyses.)

Second, we asked whether the gene sets highlighted from the functional genomics work would have been among the significant results when all gene sets are analysed for low average SNP *p*-values. Here, we used the plink gene set test and chi-squared SNP *p*-values as above. Of the GO gene sets, 26 were significant at 5% FDR, including “platelet activation” and “ECM disassembly”, two of the largest gene sets highlighted from the functional

genomics work. Of the KEGG/Reactome gene sets, 52 were significant at 5% FDR, including "signalling by PGDF". All of these three gene sets had *q*-values >0.03 and were thus not among the most significant gene sets identified.

AUTHOR CONTRIBUTIONS

Study design: EZ, JMW. Omics data analysis: GRSR, TIR, JS. Sample collection: JMW. RNA/DNA/Protein extraction: CLLM, ALAB. Proteomics: JSC, TIR. Histology and immunohistochemistry: CLLM, ALAB. Western blotting: MP, RC. Manuscript writing: GRSR, EZ, JS, TIR, JSC, MP, JMW, CLLM.

ACKNOWLEDGEMENTS

This work was funded by the Wellcome Trust (WT098051). The authors wish to thank Sara Dunn and Clive Buckle for contribution to extraction of RNA/DNA/protein from chondrocyte samples, Danielle Walker for research administration, and Pei-Chien Tsai and Jordana Bell for advice on the methylation analyses. The authors would like to acknowledge the contribution of the Wellcome Trust Sanger Institute Sample Management, Illumina Bespoke, and Genotyping teams to this work. GRSR was supported by the European Molecular Biology Laboratory and the Wellcome Trust Sanger Institute through an EBI-Sanger Postdoctoral Fellowship. This study utilized genotype data from arcOGEN (http://www.arcogen.org.uk/) funded by a special purpose grant from Arthritis Research UK (grant 18030). This study makes use of data generated by the Wellcome Trust Case-Control Consortium (the 1958 British Birth Cohort collection and the UK Blood Services Collection). A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

REFERENCES

1. Vos T, Flaxman AD, Naghavi M, AlMazroa MA, Memish ZA. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010 (vol 380, pg 2163, 2012). Lancet. 2013;381(9867):628–8.

2. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. Lancet. 2005;365(9463):965–73.

3. Valdes AM, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. Nat Rev Rheumatol. 2011 Jan;7(1):23–32.

4. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum. 1998 May;41(5):778–99.

5. Fang H, Beier F. Mouse models of osteoarthritis: modelling risk factors and assessing outcomes. Nat Rev Rheumatol. Nature Publishing Group; 2014 Jul 1;10(7):413–21.

6. Reynard LNL, Loughlin JJ. The genetics and functional analysis of primary osteoarthritis susceptibility. Expert Rev Mol Med. 2013 Jan 1;15:e2–e2.

7. Ruiz-Romero C, Fernández-Puente P, Calamia V, Blanco FJ. Lessons from the proteomic study of osteoarthritis. Expert Review of Proteomics. Informa Healthcare; 2015 Jul 7;:1–11.

8. Lee AS, Ellman MB, Yan D, Kroin JS, Cole BJ, van Wijnen AJ, et al. A current review of molecular mechanisms regarding osteoarthritis and pain. Gene. 2013 Sep 24;527(2):440–7.

9. Yuan XL, Meng HY, Wang YC, Peng J, Guo QY, Wang AY, et al. Bone–cartilage interface crosstalk in osteoarthritis: potential pathways and future therapeutic strategies. Osteoarthritis and Cartilage. 2014 Aug;22(8):1077–89.

10. Xia B, Di Chen, Zhang J, Hu S, Jin H, Tong P. Osteoarthritis pathogenesis: a review of molecular mechanisms. Calcif Tissue Int. 2014 Dec;95(6):495–505.

11. Mankin HJ, Dorfman H, Lippiello L, Zarins A. Biochemical and Metabolic Abnormalities in Articular Cartilage from Osteo-Arthritic Human Hips. The Journal of Bone & Joint Surgery. The Journal of Bone and Joint Surgery, Inc; 1971 Apr 1;53(3):523–37.

12. Pearson RG, Kurien T, Shu KSS, Scammell BE. Histopathology grading systems for characterisation of human knee osteoarthritis--reproducibility, variability, reliability, correlation, and validity. Osteoarthr Cartil. 2011 Mar;19(3):324–31.

13. Lourido L, Calamia V, Mateos J, Fernández-Puente P, Fernández-Tajes J, Blanco FJ, et al. Quantitative Proteomic Profiling of Human Articular Cartilage Degradation in Osteoarthritis. J Proteome Res. American Chemical Society; 2014 Dec 5;13(12):6096–106.

14. Petrella RJ. Hyaluronic acid for the treatment of knee osteoarthritis: long-term outcomes from a naturalistic primary care experience. Am J Phys Med Rehabil. 2005

Apr;84(4):278–83–quiz284–293.

15. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014 Jan;42(Database issue):D191–8.

16. Mateos J, Lourido L, Fernández-Puente P, Calamia V, Fernández-López C, Oreiro N, et al. Differential protein profiling of synovial fluid from rheumatoid arthritis and osteoarthritis patients using LC–MALDI TOF/TOF. J Proteomics. 2012 Jun 5;75(10):2869–78.

17. Balakrishnan L, Nirujogi R, Ahmad S, Bhattacharjee M, Manda SS, Renuse S, et al. Proteomic analysis of human osteoarthritis synovial fluid. Clinical proteomics. 2014;11(1):6.

18. Attur M, Yang Q, Shimada K, Tachida Y, Nagase H, Mignatti P, et al. Elevated expression of periostin in human osteoarthritic cartilage and its potential role in matrix degradation via matrix metalloproteinase-13. The FASEB Journal. 2015 Oct 1;29(10):4107–21.

19. Kim BY, Olzmann JA, Choi SI, Ahn SY, Kim TI, Cho HS, et al. Corneal Dystrophy-associated R124H Mutation Disrupts TGFBI Interaction with Periostin and Causes Mislocalization to the Lysosome. J Biol Chem. 2009 Jul 10;284(29):19580–91.

20. Hopwood B, Tsykin A, Findlay DM, Fazzalari NL. Microarray gene expression profiling of osteoarthritic bone suggests altered bone remodelling, WNT and transforming growth factor-β/bone morphogenic protein signalling. Arthritis Res Ther. 2007;9(5):R100.

21. Nakayama N. A novel chordin-like BMP inhibitor, CHL2, expressed preferentially in chondrocytes of developing cartilage and osteoarthritic joint cartilage. Development. 2004 Jan 1;131(1):229–40.

22. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015 Jan 22;347(6220):1260419–9.

23. Verma P, Dalal K. ADAMTS-4 and ADAMTS-5: key enzymes in osteoarthritis. J Cell Biochem. 2011 Dec;112(12):3507–14.

24. Song R-H, Tortorella MD, Malfait A-M, Alston JT, Yang Z, Arner EC, et al. Aggrecan degradation in human articular cartilage explants is mediated by both ADAMTS-4 and ADAMTS-5. Arthritis Rheum. 2007 Feb;56(2):575–85.

25. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010 May 2;28(5):511–5.

26. Bush PG, Hall AC. The volume and morphology of chondrocytes within non-degenerate and degenerate human articular cartilage☆. Osteoarthritis and Cartilage. 2003 Apr;11(4):242–51.

27. Musumeci G, Leonardi R, Carnazza ML, Cardile V, Pichler K, Weinberg AM, et al.

Aquaporin 1 (AQP1) expression in experimentally induced osteoarthritic knee menisci: An in vivo and in vitro study. Tissue and Cell. 2013 Apr;45(2):145–52.

28. Geyer M, Grässel S, Straub RH, Schett G, Dinser R, Grifka J, et al. Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. Osteoarthr Cartil. 2009 Mar;17(3):328–35.

29. Westergaard UB, Andersen MH, Heegaard CW, Fedosov SN, Petersen TE. Tetranectin binds hepatocyte growth factor and tissue-type plasminogen activator. Eur J Biochem. 2003 Mar 31;270(8):1850–4.

30. Valdes AM, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, et al. Association study of candidate genes for the prevalence and progression of knee osteoarthritis. Arthritis Rheum. 2004;50(8):2497–507.

31. Karlsson C, Dehne T, Lindahl A, Brittberg M, Pruss A, Sittinger M, et al. Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. Osteoarthr Cartil. 2010 Apr;18(4):581–92.

32. Herren T, Swaisgood C, Plow EF. Regulation of plasminogen receptors. Front Biosci (Landmark Ed). 2003 Jan 1;8:d1–d8.

33. Valdes AM, Van Oene M, Hart DJ, Surdulescu GL, Loughlin J, Doherty M, et al. Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women. Arthritis Rheum. 2006;54(2):533–9.

34. Panoutsopoulou K, Zeggini E. Advances in osteoarthritis genetics. Journal of Medical Genetics. 2013 Jul 18.

35. Tchetina EV. Developmental Mechanisms in Articular Cartilage Degradation in Osteoarthritis. Arthritis. 2011;2011(7542):1–16.

36. Remst DFG, Blom AB, Vitters EL, Bank RA, van den Berg WB, Davidson ENB, et al. Gene expression analysis of murine and human osteoarthritis synovium reveals elevation of transforming growth factor β-responsive genes in osteoarthritis-related fibrosis. Arthritis Rheumatol. 2014 Mar 1;66(3):647–56.

37. Halpain S, Dehmelt L. The MAPI family of microtubule-associated proteins. Genome Biol. 2006;7(6):–224.

38. Blain EJ. Involvement of the cytoskeletal elements in articular cartilage homeostasis and pathology. Int J Exp Pathol. 2009 Feb;90(1):1–15.

39. Kanenari M, Zhao J, Abiko Y. Enhancement of microtubule-associated protein-1 Alpha gene expression in osteoblasts by low level laser irradiation. Laser Ther. 2011;20(1):47–51.

40. Péterfi Z, Geiszt M. Peroxidasins: novel players in tissue genesis. Trends Biochem Sci. 2014 Jul;39(7):305–7.

41. Foldager CB, Toh WS, Gomoll AH, Olsen BR, Spector M. Distribution of Basement Membrane Molecules, Laminin and Collagen Type IV, in Normal and Degenerated

Cartilage Tissues. Cartilage. 2014 Apr;5(2):123–32.

42. Wang Q, Rozelle AL, Lepus CM, Scanzello CR, Song JJ, Larsen DM, et al. Identification of a central role for complement in osteoarthritis. Nat Med. 2011 Dec;17(12):1674–9.

43. Hu SI. Human HtrA, an Evolutionarily Conserved Serine Protease Identified as a Differentially Expressed Gene Product in Osteoarthritic Cartilage. Journal of Biological Chemistry. 1998 Dec 18;273(51):34406–12.

44. Polur I, Lee PL, Servais JM, Xu L, Li Y. Role of HTRA1, a serine protease, in the progression of articular cartilage degeneration. Histol Histopathol. 2010 May;25(5):599–608.

45. Sokolove J, Lepus CM. Role of inflammation in the pathogenesis of osteoarthritis: latest findings and interpretations. Therapeutic Advances in Musculoskeletal Disease. 2013 Apr 26;5(2):77–94.

46. Shimizu-Hirota R, Sasamura H, Kuroda M, Kobayashi E, Saruta T. Functional characterization of podocan, a member of a new class in the small leucine-rich repeat protein family. FEBS Lett. 2004 Apr 9;563(1-3):69–74.

47. Otero M, Plumb DA, Tsuchimochi K, Dragomir CL, Hashimoto K, Peng H, et al. E74-like Factor 3 (ELF3) Impacts on Matrix Metalloproteinase 13 (MMP13) Transcriptional Control in Articular Chondrocytes under Proinflammatory Stress. J Biol Chem. 2012 Jan 27;287(5):3559–72.

48. Liu Z, Cai H, Zheng X, Zhang B, Xia C. The Involvement of Mutual Inhibition of ERK and mTOR in PLCγ1-Mediated MMP-13 Expression in Human Osteoarthritis Chondrocytes. IJMS. 2015 Aug;16(8):17857–69.

49. Prasadam I, Zhou Y, Shi W, Crawford R, Xiao Y. Role of dentin matrix protein 1 in cartilage redifferentiation and osteoarthritis. Rheumatology. 2014 Nov 21;53(12):2280–7.

50. Xu J, Yi Y, Li L, Zhang W, Wang J. Osteopontin induces vascular endothelial growth factor expression in articular cartilage through PI3K/AKT and ERK1/2 signaling. Mol Med Report. 2015 Jun 22.

51. Long DL, Ulici V, Chubinskaya S, Loeser RF. Heparin-binding epidermal growth factor-like growth factor (HB-EGF) is increased in osteoarthritis and regulates chondrocyte catabolic and anabolic activities. Osteoarthritis and Cartilage. 2015 Sep;23(9):1523–31.

52. Pufe T, Groth G, Goldring MB, Tillmann B, Mentlein R. Effects of pleiotrophin, a heparin-binding growth factor, on human primary and immortalized chondrocytes. Osteoarthritis and Cartilage. 2007 Feb;15(2):155–62.

53. Patil AS, Sable RB, Kothari RM. Occurrence, biochemical profile of vascular endothelial growth factor (VEGF) isoforms and their functions in endochondral ossification. J Cell Physiol. 2012 Jan 11;227(4):1298–308.

54. Chia S-L, Sawaji Y, Burleigh A, McLean C, Inglis J, Saklatvala J, et al. Fibroblast Growth Factor 2 Is an Intrinsic Chondroprotective Agent That Suppresses ADAMTS-5 and

Delays Cartilage Degradation in Murine Osteoarthritis. Arthritis Rheum. 2009 Jul;60(7):2019–27.

55. Marmotti A, Rossi R, Castoldi F, Roveda E, Michielon G, Peretti GM. PRP and Articular Cartilage: A Clinical Update. BioMed Research International. 2015;2015(11):1–19.

56. Meheux CJ, McCulloch PC, Lintner DM, Varner KE, Harris JD. Efficacy of Intra-articular Platelet-Rich Plasma Injections in Knee Osteoarthritis: A Systematic Review. Arthroscopy. 2015 Sep 29.

57. Zhou Q, Xu C, Cheng X, Liu Y, Yue M, Hu M, et al. Platelets promote cartilage repair and chondrocyte proliferation via ADP in a rodent model of osteoarthritis. Platelets. Informa Healthcare.

58. Mapp PI, Walsh DA. Mechanisms and targets of angiogenesis and nerve growth in osteoarthritis. Nat Rev Rheumatol. Nature Publishing Group; 2012 Jul 1;8(7):390–8.

59. Ashraf S, Walsh DA. Angiogenesis in osteoarthritis. Current Opinion in Rheumatology. 2008 Sep;20(5):573–80.

60. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2013 Dec 28;42(D1):D1091–7.

61. Chijimatsu R, Kunugiza Y, Taniyama Y, Nakamura N, Tomita T, Yoshikawa H. Expression and pathological effects of periostin in human osteoarthritis cartilage. BMC Musculoskelet Disord. 2015 Aug 21;16(1):95.

62. Coutu DL, Wu JH, Monette A, Rivard GE, Blostein MD, Galipeau J. Periostin, a Member of a Novel Family of Vitamin K-dependent Proteins, Is Expressed by Mesenchymal Stromal Cells. J Biol Chem. 2008 Jun 20;283(26):17991–8001.

63. Misra D, Booth SL, Tolstykh I, Felson DT, Nevitt MC, Lewis CE, et al. Vitamin K deficiency is associated with incident knee osteoarthritis. Am J Med. 2013 Mar;126(3):243–8.

64. Jeffries MA, Donica M, Baker LW, Stevenson ME, Annan AC, Humphrey MB, et al. Genome-Wide DNA Methylation Study Identifies Significant Epigenomic Changes in Osteoarthritic Cartilage. Arthritis & Rheumatology. 2014 Sep 26;66(10):2804–15.

65. Moazedi-Fuerst FC, Hofner M, Gruber G, Weinhaeusel A, Stradner MH, Angerer H, et al. Epigenetic differences in human cartilage between mild and severe OA. J Orthop Res. 2014 Dec;32(12):1636–45.

66. Hollander den W, Ramos YFM, Bos SD, Bomer N, van der Breggen R, Lakenberg N, et al. Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. Annals of the Rheumatic Diseases. 2014 Oct 30;73(12):2208–12.

67. Tew SR, McDermott BT, Fentem RB, Peffers MJ, Clegg PD. Transcriptome-wide analysis of messenger RNA decay in normal and osteoarthritic human articular chondrocytes. Arthritis Rheumatol. 2014 Nov 1;66(11):3052–61.

68. Stenberg J, Rüetschi U, Skiöldebrand E, Kärrholm J, Lindahl A. Quantitative

proteomics reveals regulatory differences in the chondrocyte secretome from human medial and lateral femoral condyles in osteoarthritic patients. Proteome Sci. 2013;11(1):43.

69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Publishing Group. Nature Publishing Group; 2012 Mar 4;9(4):357–9.

70. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic acids …. 2013.

71. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 8;31(2):166–9.

72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009 Dec 22;26(1):139–40.

73. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. Nature Publishing Group; 2013 Aug 22;8(9):1765–86.

74. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics. 2014 Jan 27;30(3):428–30.

75. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. BioMed Central Ltd; 2013 May 1;14(1):293.

76. Chen Y-A, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. Taylor & Francis.

77. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. Bioinformatics. 2012 Apr 26;28(9):1280–1.

78. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

79. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

80. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, et al. Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. 2009 Jan 5;37(D619-622).

81. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005 Oct 25;102(43):15545–50.

82. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. bioinformaticsoxfordjournalsorg.

FIGURES

Figure 1: (a) A schematic view of the 3 functional genomics experiments identifying the number of genes shortlisted for each. (b) Venn diagram identifying the number of overlapping shortlisted genes from each individual experiment.

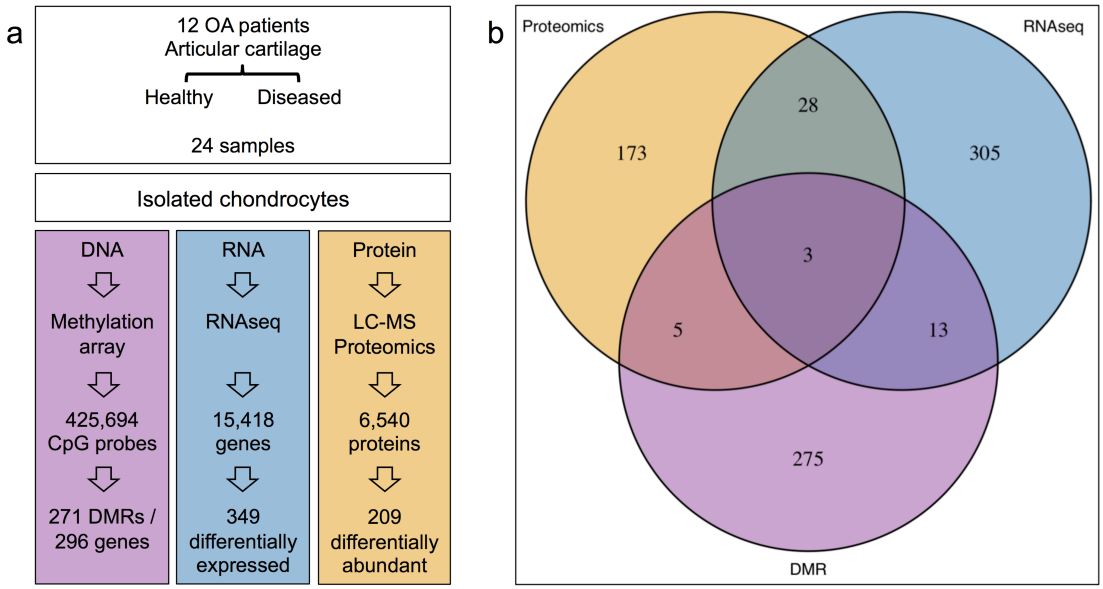


Figure 2: (a) A comparison of the fold changes between all genes identified in both the proteomics and RNA-seq experiments. Each gene is represented as a single point, and the colour corresponds to whether the gene is identified as differentially expressed using edgeR in the RNA-seq or proteomics experiments, or both. The trend lines are derived from a linear regression in each subset. Positive fold changes indicate increased expression in diseased samples. (b) Comparison of RNA-seq fold change, and promoter region methylation. The trend lines are derived from a linear regression in each subset.

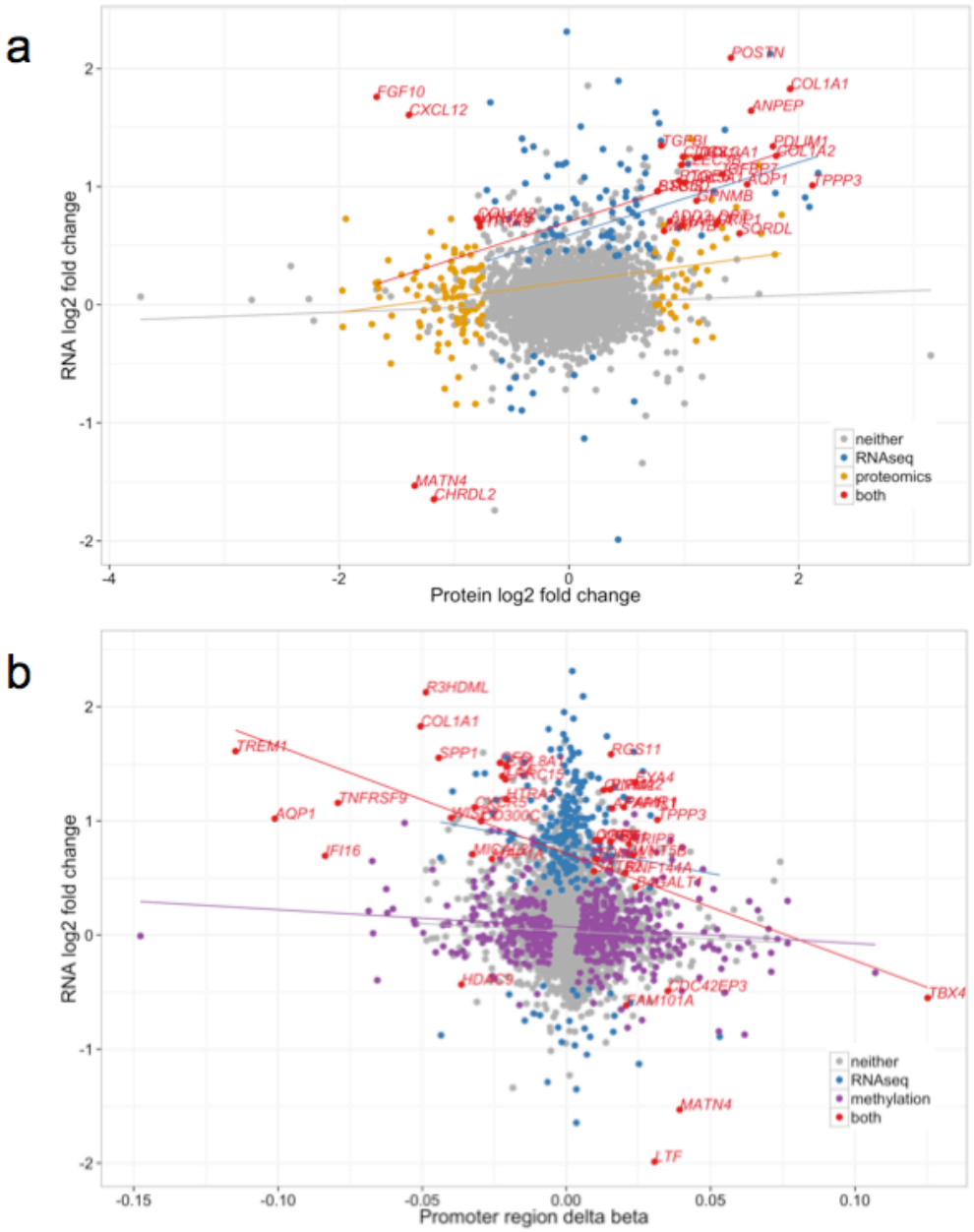
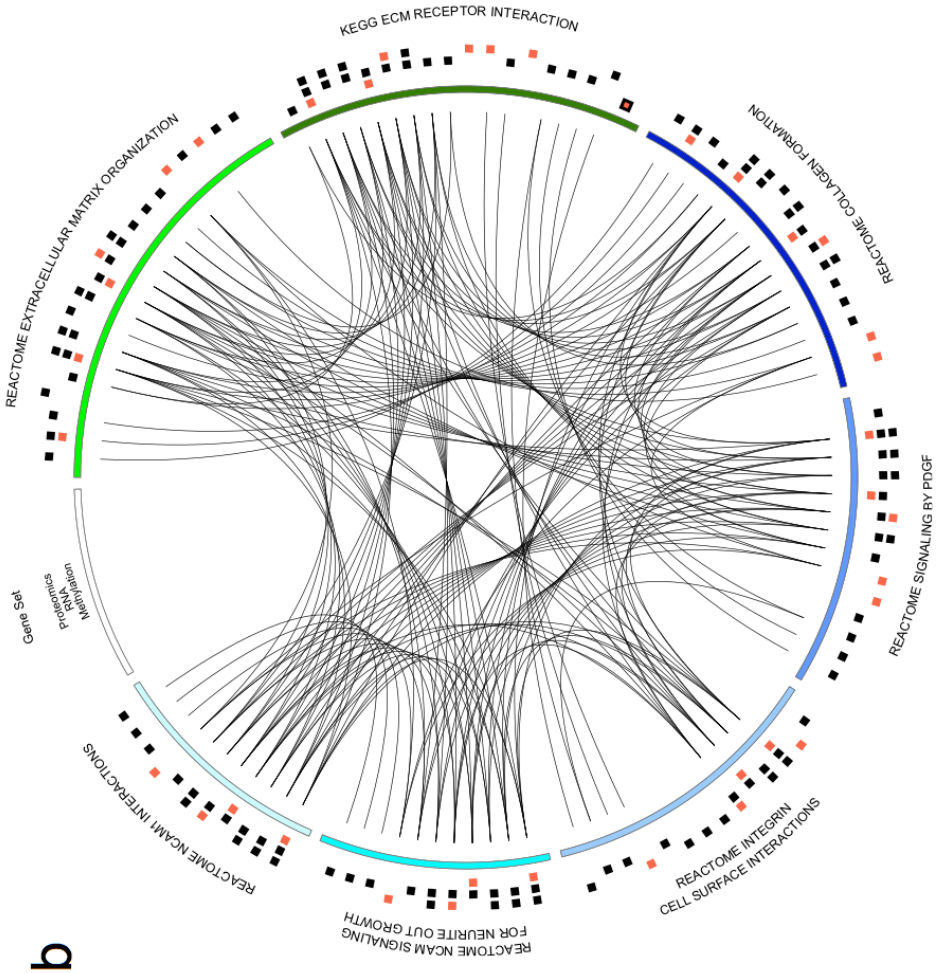
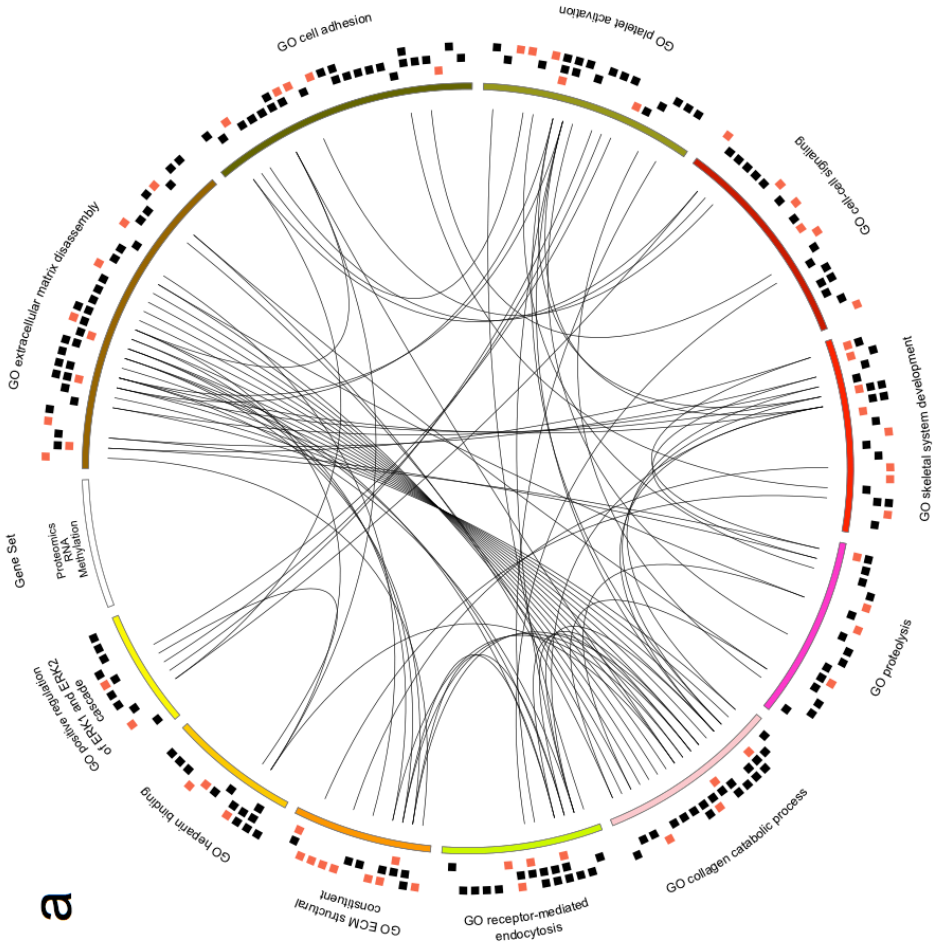


Figure 3: Significant gene set enrichments from KEGG/Reactome (a) and Gene Ontology (b). The circos plots show enriched gene sets, with genes differentially regulated in at least one of the methylation, RNA-seq, or proteomics experiments. Lines connect genes that occur in several gene sets. The three outside circles show boxes for genes with significantly higher (black) or lower (red) methylation, gene, or protein expression data. A red box with black border indicated a gene that overlaps hyper- as well as hypo-methylated DMRs.





TABLES

Table 1: Results of Drugbank[15] (www.drugbank.ca) search for therapeutic compounds with current FDA marketing authorization for a clinical indication and a potential role in OA treatment. The mechanisms of action and references are taken from Drugbank.

gene symbol	protein	compound	compound status	FDA marketing status	target specific mechanism	drug mechanism of action	reference
ANPEP	Aminopeptidase N	Ezetimibe	approved for human use	Prescription	binds protein	Anti-hyperlipidemic medication used to lower cholesterol absorption in the small intestine.	Kramer et al. (2005) PMID: 15494415
ANPEP	Aminopeptidase N	Icatibant	approved for human use	Prescription	inhibits protein	Synthetic peptidomimetic drug consisting of ten amino acids, and acts as a specific antagonist of bradykinin B2 receptors. Used in symptomatic treatment of acute attacks of hereditary angioedema in adults with C1-esterase-inhibitor deficiency.	Bawolak et al. (2006) PMID: 17026984
AQP1	Aquaporin-1	Acetazolamide	approved for human use	Prescription	inhibits protein	Carbonic anhydrase inhibitor diuretic agent. Used for the medical treatment of glaucoma, epileptic seizure, idiopathic intracranial hypertension, altitude sickness, cystinuria, periodic paralysis, central sleep apnea, and dural ectasia.	Xiang et al. (2004) PMID: 15169637
BGLAP	Osteocalcin	Phylloquinone (vitamin K_1)	approved for human use	NA (vitamin)	agonist, carboxylates protein	Fat-soluble vitamin necessary forposttranslational modification of certain proteins, mostly required for blood coagulation.	Schurgers et al. (2001) PMID: 11374034
CLEC3B	Tetranectin	Tenecteplase	approved for human use	Prescription	binds protein	Tissue plasminogen activator (tPA). Used as a thrombollytic agent.	Westergaard et al. (2003) PMID: 12694198
CXCL12	Stromal cell-derived factor 1	Tinzaparin	approved for human use	Prescription	binds protein	Tinzaparin is a low molecular weight heparin (LMWH). Used in the treatment and prophylaxis of venous thrombo-embolism.	Koo et al. (2008) PMID: 18991783
FGFR2	Fibroblast growth factor receptor 2	Palifermin	approved for human use	Prescription	binds protein	Recombinant human keratinocyte growth factor (KGF). Used to treat oral mucositis in patients undergoing cancer chemotherapy.	Beaven et al. (2007) PMID: 17728847
MAP1A	Microtubule-associated protein 1A	Estramustine	approved for human use	Prescription	disrupts protein	A nitrogen mustard linked to estradiol; used in palliative care of prostatic neoplasms.	Stearns et al. (1991_ PMID: 1647395
PTGIS	Prostacyclin synthase	Non-steroidal anti-inflammatory agents	approved for human use	Prescription	inhibits protein	COX1 and COX2 inhibitors used in the symptomatic treatment of OA.	Reed et al. (1985) PMID: 3917545
S100A4	Protein S100-A4	Trifluoperazine	approved for human use	Prescription	inhibits protein function	A phenothiazine with actions similar to chlorpromazine. It is used as an antipsychotic and an antiemetic	Malashkevish et al. (2010) PMID: 20421509