

# SCIENTIFIC REPORTS



OPEN

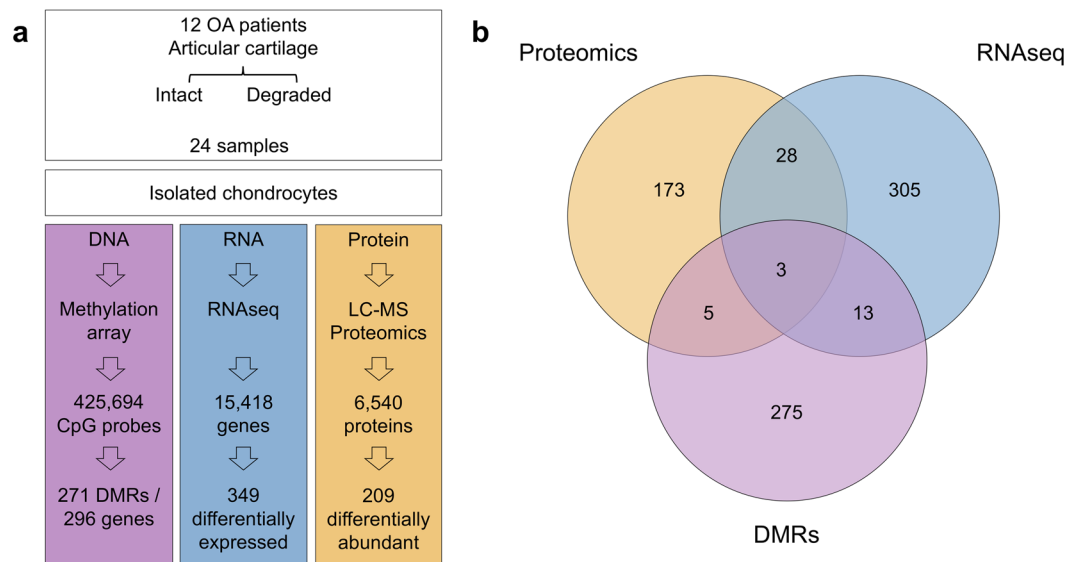
## Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis

Julia Steinberg<sup>1,10</sup>, Graham R. S. Ritchie<sup>1,2,3,4</sup>, Theodoros I. Roumeliotis<sup>1</sup>, Raveen L. Jayasuriya<sup>5</sup>, Matthew J. Clark<sup>5</sup>, Roger A. Brooks<sup>6</sup>, Abbie L. A. Binch<sup>7</sup>, Karan M. Shah<sup>5</sup>, Rachael Coyle<sup>1</sup>, Mercedes Pardo<sup>1</sup>, Christine L. Le Maitre<sup>7</sup>, Yolande F. M. Ramos<sup>8</sup>, Rob G. H. H. Nelissen<sup>9</sup>, Ingrid Meulenbelt<sup>8</sup>, Andrew W. McCaskie<sup>6</sup>, Jyoti S. Choudhary<sup>1</sup>, J. Mark Wilkinson<sup>5</sup> & Eleftheria Zeggini<sup>1</sup>

Osteoarthritis (OA) is a common disease characterized by cartilage degeneration and joint remodeling. The underlying molecular changes underpinning disease progression are incompletely understood. We investigated genes and pathways that mark OA progression in isolated primary chondrocytes taken from paired intact versus degraded articular cartilage samples across 38 patients undergoing joint replacement surgery (discovery cohort: 12 knee OA, replication cohorts: 17 knee OA, 9 hip OA patients). We combined genome-wide DNA methylation, RNA sequencing, and quantitative proteomics data. We identified 49 genes differentially regulated between intact and degraded cartilage in at least two –omics levels, 16 of which have not previously been implicated in OA progression. Integrated pathway analysis implicated the involvement of extracellular matrix degradation, collagen catabolism and angiogenesis in disease progression. Using independent replication datasets, we showed that the direction of change is consistent for over 90% of differentially expressed genes and differentially methylated CpG probes. *AQP1*, *COL1A1* and *CLEC3B* were significantly differentially regulated across all three –omics levels, confirming their differential expression in human disease. Through integration of genome-wide methylation, gene and protein expression data in human primary chondrocytes, we identified consistent molecular players in OA progression that replicated across independent datasets and that have translational potential.

Osteoarthritis (OA) affects in excess of 40% of individuals over the age of 70 years<sup>1</sup>, and is a leading cause of pain and loss of physical function<sup>2</sup>. The molecular mechanisms underlying OA remain incompletely understood and

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. <sup>2</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. <sup>3</sup>Usher Institute of Population Health Sciences & Informatics, University of Edinburgh, Edinburgh, EH16 4UX, UK. <sup>4</sup>MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK. <sup>5</sup>Department of Oncology and Metabolism, University of Sheffield, Beech Hill Road, Sheffield, S10 2RX, UK. <sup>6</sup>Division of Trauma & Orthopaedic Surgery, University of Cambridge, Box 180, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, UK. <sup>7</sup>Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, S1 1WB, UK. <sup>8</sup>Department of Medical Statistics and Bioinformatics, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, 2300RC, The Netherlands. <sup>9</sup>Department of Orthopedics, Leiden University Medical Center, Leiden, 2300RC, The Netherlands. <sup>10</sup>Present Address: Cancer Research Division, Cancer Council NSW, Sydney, NSW, 2011, Australia. Julia Steinberg, Graham R. S. Ritchie, J. Mark Wilkinson and Eleftheria Zeggini contributed equally to this work. Correspondence and requests for materials should be addressed to J.M.W. (email: [j.m.wilkinson@sheffield.ac.uk](mailto:j.m.wilkinson@sheffield.ac.uk)) or E.Z. (email: [Eleftheria@sanger.ac.uk](mailto:Eleftheria@sanger.ac.uk))



**Figure 1.** Overview of the genes identified in each -omics experiment and their overlap. **(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.

there is no curative therapy for OA; disease progression culminates in joint replacement surgery. OA is a complex disease, with both heritable and environmental factors contributing to susceptibility<sup>3</sup>.

Cartilage degeneration is one of the key features of OA. Cartilage tissue is readily accessible at joint replacement surgery, providing an opportunity to characterize the molecular processes underpinning disease development in the right tissue, both to improve our fundamental understanding of disease biology and to identify novel therapeutic opportunities. In recent years, -omics studies in OA have expanded our understanding of disease pathogenesis, reviewed in refs 4–7. Here we apply integrated multi-omics (DNA CpG methylation, RNA sequencing and quantitative proteomics) in knee OA tissue to obtain a comprehensive molecular portrait of cartilage degeneration (Fig. 1a). The fundamental question here addresses the biological processes underpinning disease progression within the OA joint, which is of direct clinical relevance to patients suffering from OA. To achieve this, we collected individually-matched pairs of cartilage tissue from patients undergoing joint replacement surgery, with one sample demonstrating advanced degenerative change and the other demonstrating little or no evidence of cartilage degeneration. The findings were then replicated in several independent populations of patients undergoing joint replacement. By integrating all three of methylation, gene expression, and protein abundance data, we discover disease processes with involvement across multiple levels, and reveal novel and robustly replicating molecular players with translational potential.

## Methods

For full details of methods see Supplementary Methods.

**Patients and samples.** Osteochondral samples were collected from 38 patients undergoing total joint replacement for OA (discovery cohort: 12 knee OA patients; 2 separate replication cohorts: 17 knee OA patients, 9 hip OA patients). Cartilage was separated from bone and chondrocytes were extracted from each sample. DNA, RNA, and protein were extracted from the isolated chondrocytes of each sample. All patients provided full written informed consent prior to participation. We obtained two cartilage samples from each patient. Knee samples were scored using the OARSI cartilage classification system<sup>8,9</sup>: one sample with high OARSI grade signifying high-grade degeneration (“degraded sample”), and one sample with low OARSI grade signifying healthy tissue or low-grade degeneration (“intact sample”) (Supplementary Fig. S1). Hip samples were classified macroscopically and visually as high-grade (“degraded”) or low-grade (“intact”), details see Supplementary Methods.

**Proteomics (discovery cohort).** LC-MS analysis was performed on the Dionex Ultimate 3000 UHPLC system coupled with the high-resolution LTQ Orbitrap Velos mass spectrometer. Proteins were defined as differentially abundant if the absolute median abundance ratio between degraded and intact samples was  $\geq 0.75$ , the absolute median abundance ratio was greater than the standard deviation of abundance ratios, and the protein was detected in paired samples from at least 5 of 11 patients post-QC.

**RNA-seq (discovery cohort).** Multiplexed libraries were sequenced on the Illumina HiSeq 2000 (75 bp paired-end read length) and individual indexed library BAM files were produced. Reads that passed quality control (QC) were re-aligned using tophat2<sup>10</sup>, and reads aligning to each gene in each sample were counted using the HTSeq package<sup>11</sup>. Absolute transcript abundance was calculated using the FPKM (fragments per kilobase of transcript per million fragments mapped) measure across 15,418 genes post-QC. We used edgeR v3.0<sup>12</sup> to identify

differentially expressed genes, applying a generalized linear model for tissue status (degraded or intact) with individual ID as covariate, and defined significance at 5% false-discovery rate (FDR).

**Methylation (discovery cohort).** Methylation was assayed using the Illumina 450k BeadChip. Data were parsed and QCed using ChAMP<sup>13</sup>; the probe beta values were quantile-normalized using the ‘dasen’ method from the watermelon package<sup>14</sup>. Differential methylation on 425,694 probes post-QC was tested using the CpGassoc package<sup>15</sup>, applying a linear model for tissue status at each probe, with individual ID as covariate.

Following previous analyses<sup>16</sup>, differentially methylated regions (DMRs) were defined as regions of at least 3 probes with  $FDR \leq 5\%$  (DMPs) and no more than 3 non-significant probes, with no more than 1 kb between adjacent probes. We used bedtools<sup>17</sup> to identify genes overlapping each DMR, extending gene boundaries to include likely promoter regions.

For a promoter-level analysis, we used the probe annotations from ChAMP to assign promoter probes to each gene. After computing the mean beta value in the promoter of each gene in each sample, we used a paired t-test to identify genes with differential promoter-region methylation between degraded and intact samples. Significance was set at 5% FDR.

**Replication of gene expression and methylation changes.** Gene expression and DNA methylation in both replication cohorts were measured using the same procedure as for the discovery data. We considered the expression of 14,762 genes that passed QC across all datasets; this included 332 of 349 genes with  $FDR \leq 5\%$  in the knee discovery data.

We considered 416,437 probes that passed QC across all datasets; this included 9,723 of 9,867 probes with  $FDR \leq 5\%$  in the knee discovery data (“DMPs”).

*In silico replication.* We carried out an additional replication analysis for gene expression changes using an available published dataset from the ongoing Research Arthritis and Articular Cartilage (RAAK) study<sup>18</sup>. This dataset comprised a comparison of affected and macroscopically preserved OA cartilage from 33 patients undergoing total joint replacement surgery (22 with hip OA, 11 with knee OA), measured using the Illumina HumanHT-12 v3 microarrays.

**Gene set analyses.** We tested whether particular biological gene sets were enriched among significant genes from each of the RNA-seq, methylation, and proteomics datasets. We used KEGG<sup>19</sup>, Reactome<sup>20</sup>, and, separately, Gene Ontology (GO) annotations<sup>21</sup>. Genes annotated to the same term were treated as a “pathway”. Enrichment was assessed using a 1-sided hypergeometric test, with significance at 5% FDR. Empirical p-values for the enrichments were obtained from randomisations.

For each gene set, we asked whether the association across the three –omics datasets (calculated as geometric mean of the p-values) was higher than expected by chance, obtaining empirical p-values from 100,000 randomisations.

**Ethical approval information.** The study was approved by Oxford NHS REC C (10/H0606/20). Tissue samples were collected under National Research Ethics approval reference 15/SC/0132, South Yorkshire and North Derbyshire Musculoskeletal Biobank, University of Sheffield; and under National Research Ethics approval reference 11/EE/0011, Cambridge Biomedical Research Centre Human Research Tissue Bank, Cambridge University Hospitals, UK. All participants provided written informed consent. All methods were carried out in accordance with the relevant guidelines and regulations.

**Data availability.** All genome-wide summary statistics generated in the discovery component of this study are available as Supplementary Tables. All discovery RNA sequencing data are available in the EGA repository, <https://www.ebi.ac.uk/ega/studies/EGAS00001001203>. All discovery methylation data are available in the EGA repository, <https://www.ebi.ac.uk/ega/studies/EGAS00001001213>. The discovery proteomics data has been deposited in the PRIDE archive under study ID PXD002014 [publication in process].

## Results

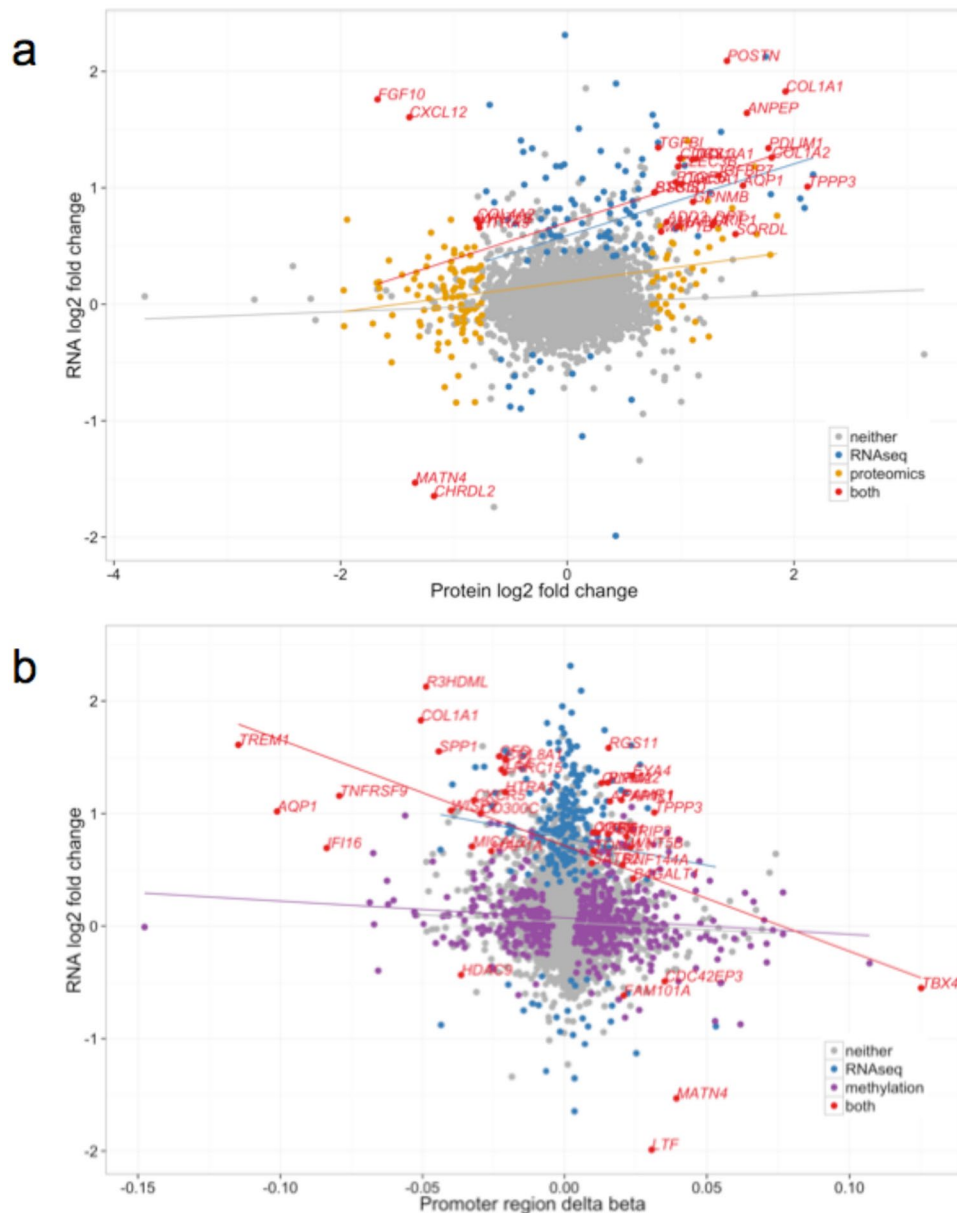
We compared the degraded and intact tissue across patient-matched samples.

We identified 209 proteins with evidence of differential abundance (Supplementary Table S1): 90 at higher and 119 at lower abundance in the degraded samples. We confirmed the quality of the data using an orthogonal label-free approach and, separately, Western blotting (see Supplementary Note).

We identified 349 differentially expressed genes: 296 genes showed higher and 53 lower expression levels in the degraded samples (Supplementary Table S2). One of the most strongly down-regulated genes in degraded cartilage was *CHRD2L*. The corresponding protein was also found at lower abundance in degraded cartilage and we confirmed its expression in cartilage by immunohistochemistry (Supplementary Fig. S2). *CHRD2L* is a bone morphogenetic protein (*BMP*) inhibitor that has been reported to be lost from chondrocytes of the superficial zone and shifted to the middle zone in OA cartilage in a targeted study<sup>22</sup>.

We identified 9,896 differentially methylated probes (DMPs) (Supplementary Table S3), and 271 differentially methylated regions (DMRs) composed of multiple differentially methylated CpG sites, overlapping 296 unique genes (Supplementary Table S4).

**Integration across multiple –omics levels.** *Methylation, RNA sequencing, and proteomics.* We identified 49 genes with evidence of differential regulation from at least two of the –omics analyses (Supplementary Table S5). Three genes demonstrated significant evidence for involvement in OA progression across all three levels: *AQP1*, *COL1A1* and *CLEC3B* (Fig. 1b). All three genes were up-regulated in degraded tissue in both the



**Figure 2.** Comparison of changes identified in the -omics experiments. **(a)** Comparison of the log-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 degraded samples. **(b)** Comparison of RNA-seq log-fold-change and mean promoter region methylation change. The trend lines are derived from a linear regression in each subset. Genes are coloured according to the results of the RNA-seq and the promoter-region analyses analogously to Fig. 2a.

RNA-seq and proteomics analyses (Fig. 2a). *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. Using IHC we independently confirmed the presence of *AQP1*, *COL1A1* and *CLEC3B* in articular cartilage chondrocytes (Supplementary Fig. S2). We also replicated the direction of gene expression change for all three genes in independent data (see below and Supplementary Table S6).

Of the 49 genes with evidence of differential regulation on at least two molecular levels, 33 add substantive evidence to genes previously reported and 16 genes (33%) have not previously been implicated in OA (Supplementary Table S5).

**Proteomics and RNA sequencing.** Among the 209 proteins with evidence of differential abundance in the proteomics data, 31 were also differentially expressed between degraded and intact samples at the RNA level

(hypergeometric  $p = 5.3 \times 10^{-7}$ , Fig. 1b). Twenty-six of these 31 genes showed concordant directions of changes at both the protein and RNA level (binomial  $p = 0.0002$ ). Five genes were over-expressed at the RNA level and less abundant at the protein level in the degraded tissue (*COL4A2*, *CXCL12*, *FGF10*, *HTRA3* and *WNT5B*); all five are annotated as secreted proteins in the Human Protein Atlas<sup>23</sup> (see Supplementary Note).

Using all samples irrespective of tissue status, we found a significant correlation between gene expression levels and protein abundance (Spearman's  $\rho = 0.29$ ,  $p < 2.2 \times 10^{-16}$ , Supplementary Note and Supplementary Fig. S3). Considering the RNA and protein changes in degraded compared to intact samples (Fig. 2a), we also identified a significant positive correlation (Pearson's  $r = 0.17$ ,  $p < 2.2 \times 10^{-16}$ ). This correlation became substantially higher when we only considered the 31 genes that were differentially expressed in both datasets (Pearson's  $r = 0.43$ ,  $p = 0.01$ ).

**Methylation and RNA sequencing.** Sixteen of the genes overlapping a DMR were also differentially transcribed (Fig. 1b). In the direct comparison of methylation and gene expression using all samples irrespective of tissue status, we found the expected negative correlation between promoter region methylation and gene expression (Spearman's  $\rho = -0.43$ ,  $p < 2.2 \times 10^{-16}$ , Supplementary Fig. S3). Based on the comparison of intact to degraded cartilage, the log-fold-changes in RNA expression and the differences in mean promoter region methylation values demonstrated a small but highly significant correlation (Pearson's  $r = -0.08$ ,  $p < 2.2 \times 10^{-16}$ , Fig. 2b, Supplementary Table S7). The correlation became substantially higher when we considered the 39 genes with significant differences at both the promoter methylation and transcription levels (Pearson's  $r = -0.48$ ,  $p = 0.002$ ).

**Replication of gene expression and methylation changes.** We found good correlation between the global gene expression log-fold changes and the methylation log-fold changes estimated in the discovery and in the knee and, separately, hip replication data (Fig. 3a,b, Pearson  $r$  0.5–0.9,  $p < 10^{-15}$ ) and excellent replication rates for differentially expressed genes and probes with differential methylation (Fig. 3c,d, over 90% have same direction of change in replication data and at least 47% reach nominal significance, Supplementary Table S8).

We specifically considered the 49 genes with evidence from at least two -omics levels. Of these, 47 had gene expression data in the discovery and both replication datasets; the direction of change replicated at nominal significance for 36 genes in the knee and for 26 genes in the hip replication data (Supplementary Table S6). Notably, the direction of change replicated in at least one of the knee and hip replication datasets at nominal significance for 13 of the 16 genes that have not previously been associated with OA (Supplementary Table S6).

We additionally pursued replication in a further independent, published microarray gene expression dataset from the RAAK study<sup>18</sup>. This study used pooled knee and hip samples, as they found good agreement between the pooled and joint-stratified analyses<sup>18</sup>; to increase power, we also used the pooled data as *in silico* replication. Despite the difference in genomics technology (RNA-seq in discovery, microarray in RAAK), we again found good agreement of the log-fold-changes for the differentially expressed genes from the discovery data (over 60% have same direction of effect, Pearson  $r$  in 0.24–0.43 at  $p \leq 0.003$ , see Supplementary Note).

**Pathways involved in OA progression.** We identified biological processes with consistent evidence of involvement in OA progression at multiple levels (Fig. 4, Supplementary Fig. S4, Supplementary Tables S9–S10). A strong theme is cartilage matrix regulation, degeneration, and disassembly.

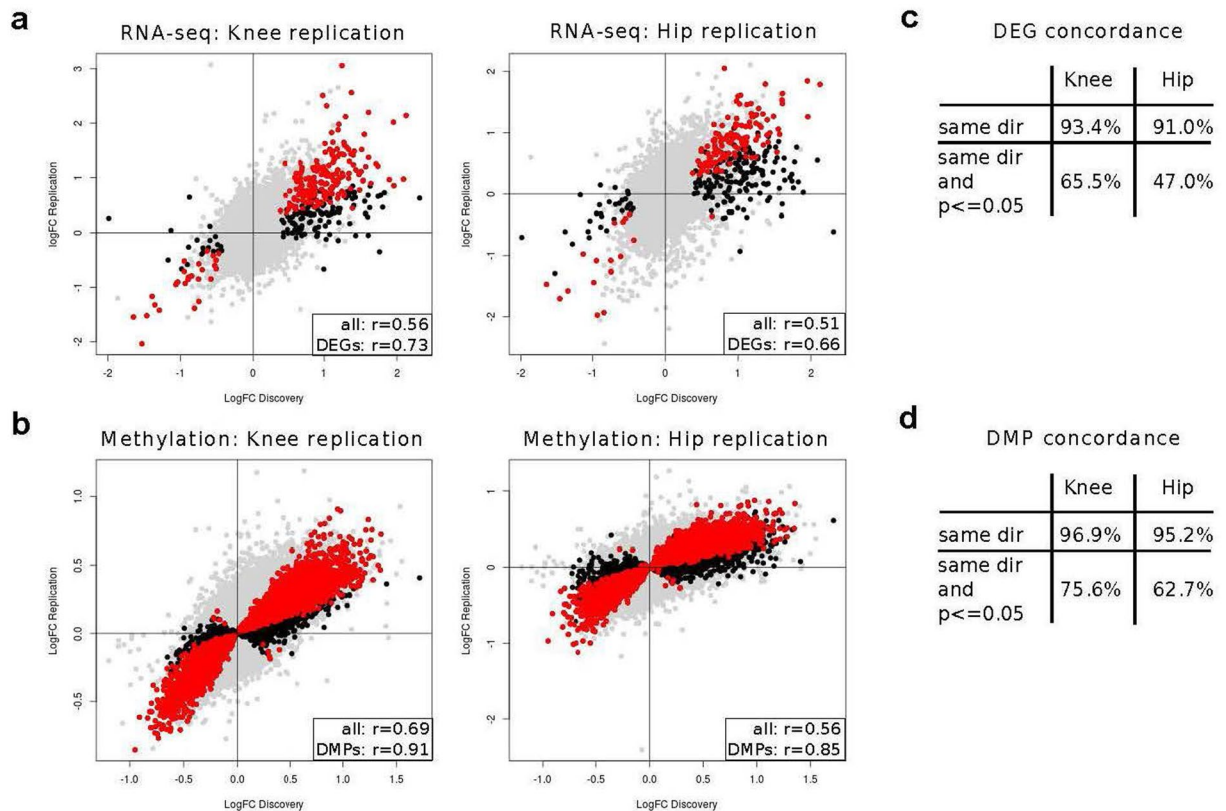
Positive regulation of ERK1/2 cascade, heparin-binding and platelet activation were also enriched at multiple molecular levels and are interconnected through common genes (Supplementary Note). We also found enrichment of genes involved in the regulation of angiogenesis at multiple levels, implicating pathways involved in blood vessel and nervous system formation (NCAM signaling for neurite outgrowth and PDGF signaling). The growth of blood vessels and nerves are closely linked processes that share regulatory mechanisms, including the ERK cascade and heparin-binding proteins<sup>24</sup>. Indeed, histological examination of the samples we investigated showed greater blood vessel ingrowth in tissues with more advanced OA (Supplementary Fig. S1).

**Drug repurposing potential.** To identify existing drugs that could be repurposed for OA, we searched Drugbank<sup>25</sup> for the 49 genes differentially regulated across at least two -omics levels. We identified ten agents with investigational or established actions on nine of the dysregulated proteins, which already have Food and Drug Administration Marketing Authorization for use in humans (Table 1). Among them are non-steroidal anti-inflammatory agents active against prostacyclin synthase (NSAIDs), which already have marketing authorisation for the symptomatic treatment of OA. Our search also identified phylloquinone (vitamin K<sub>1</sub>), an agonist of osteocalcin (*BGLAP* gene). Periostin (*POSTN* gene), a protein with elevated expression in OA in this study and others<sup>26–28</sup>, is a vitamin K-dependent protein that induces cartilage degeneration<sup>29</sup>. A recent study has associated sub-clinical vitamin K deficiency with knee OA incidence<sup>30</sup>, adding evidence that this compound could be a disease-modifying agent in OA.

## Discussion

Characterisation of the molecular landscape of OA is now increasingly feasible and catalyzed by technological advances in functional genomics, and the accessibility of the relevant tissue at joint replacement surgery. Indeed, the field of functional genomics in OA is now budding<sup>31</sup>. Previous studies of osteoarthritis have investigated methylation<sup>32,33</sup>, transcription<sup>18,34–37</sup>, or protein expression<sup>38,39</sup> separately, or a combination of up to two of these -omics assays<sup>16,40,41</sup>, some with the addition of genetic data<sup>42</sup>. By contrast, this study provides a systematic analysis of biological changes involved in OA across genome-wide methylation, gene and protein expression levels.

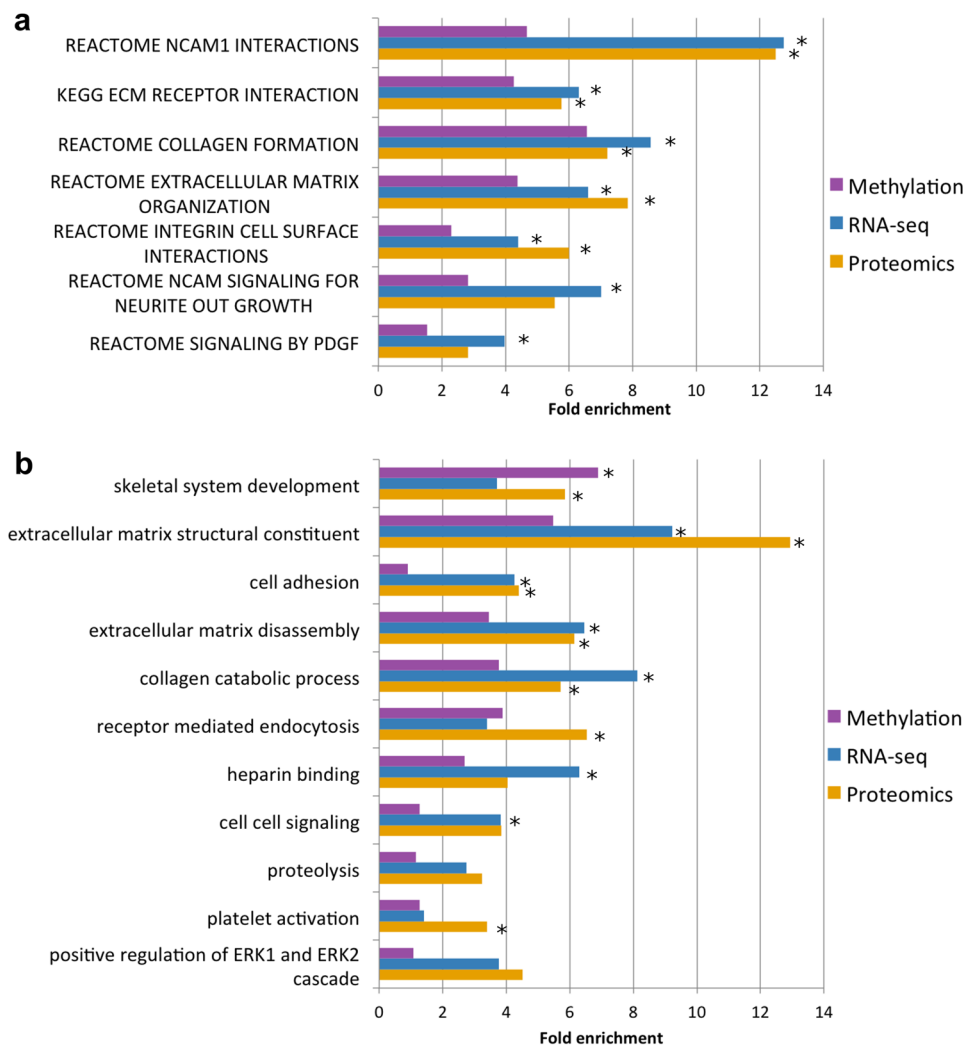
We identify 49 genes with evidence of changes between intact and degraded cartilage on multiple molecular levels, including some novel molecular players in OA. We provide robustly-replicating evidence with convergence of all three -omics levels of the involvement of *AQP1*, *COL1A1*, and *CLEC3B* in disease progression.



**Figure 3.** Replication of gene expression and methylation changes. **(a,b)** Replication of gene expression changes **(a)** and CpG methylation changes **(b)** in independent datasets of samples from patients with knee (left) and hip (right) OA. Differentially expressed genes **(a, DEGs)** and differentially methylated probes **(b, DMPs)** from the discovery data are marked in black; DEGs and DMPs that additionally show nominal significance in the replication data are marked in red. Inset: correlation between log-fold-changes in discovery and replication data. All: all 14,762 genes **(a)** or 416,437 probes **(b)** that passed QC in the knee discovery, knee replication, and hip replication data. DEGs: 332 of the 349 genes with  $FDR \leq 5\%$  in the knee discovery data that also pass QC in both replication datasets; DMPs: 9,723 of the 9,867 probes with  $FDR \leq 5\%$  in the knee discovery data that also pass QC in both replication datasets. All correlation values shown have  $p < 10^{-15}$ . **(c,d)** Directional concordance of changes between discovery and replication for the differentially expressed genes **(c, DEGs)** and differentially methylated probes **(d, DMPs)** from the discovery data. Same dir: proportion of DEGs or DMPs with same direction of change in the replication data. Same dir and  $p \leq 0.05$ : among DEGs or DMPs with same direction of change in the replication data, proportion of those that reach nominal significance in the replication dataset.

*AQP1* encodes aquaporin-1, a protein that facilitates water transport across biological membranes. Chondrocyte swelling and increased cartilage hydration has been suggested as an important mechanism in OA<sup>43</sup>. *AQP1* has been found to be over-expressed in the meniscus of a rat model of knee OA<sup>44</sup> and in degraded compared to intact articular cartilage in human patients with knee OA<sup>45</sup>. *CLEC3B* encodes tetranectin, which binds human tissue plasminogen activator (tPA)<sup>46</sup>. *CLEC3B* has previously been found to be up-regulated in OA<sup>34,47</sup>. *COL1A1* is one of several collagens that were differentially abundant at both the RNA and protein levels (see Supplementary Table S5). Collagens are the main structural components of cartilage and collagen dysregulation plays an important role in OA<sup>48,49</sup>. A recent study identified up-regulation of *COL1A1* in synovium from humans with end-stage OA, in the synovium of mice with induced OA and in human fibroblasts stimulated with TGF- $\beta$ <sup>50</sup>. Moreover, the products of *AQP1* and *CLEC3B* are targeted by existing drugs approved for human use: aquaporin-1 is inhibited by Acetazolamide, a carbonic anhydrase inhibitor diuretic agent, and tetranectin is bound by Tenecteplase, a thrombolytic agent (Table 1).

We further identify previously unreported genes with concordant molecular changes in the degraded tissue on multiple levels, and with robust replication in independent datasets. *MAPIA* and *MAPIB* were both significantly up-regulated in degraded cartilage at the RNA and protein levels (Fig. 2a). These proteins are mainly expressed in the brain and involved in regulating the neural cytoskeleton<sup>51</sup>. Cytoskeletal regulation is thought to be an important process in OA<sup>52</sup> and recently these proteins were implicated in bone formation<sup>53</sup>. *PXDN* was up-regulated at the RNA level, confirmed by replication, and overlapped two hypo-methylated DMRs. *PXDN* encodes peroxidasin, which is secreted into the extracellular matrix and catalyses collagen IV cross-linking<sup>54</sup>. The other replicated genes not previously implicated in OA (Supplementary Table S6) have relatively little characterisation.



**Figure 4.** Significantly enriched gene sets in the integrative analysis. Only gene sets that contain at least 5 genes from 2 experiments are shown. Asterisks indicate significant (5% FDR) enrichment for each individual gene-set in an individual -omics experiment. **(a)** Fold enrichment for KEGG & Reactome pathways significant at 5% FDR in the integrative analysis. **(b)** Fold enrichment for GO terms significant at 5% FDR in the integrative analysis.

*ANPEP* (aminopeptidase E) is a broad specificity aminopeptidase that has previously been detected in the synovial fluid of OA patients<sup>55</sup> and therefore has potential as a novel OA biomarker. *CHRD2* is a bone morphogenetic protein (*BMP*) inhibitor that has been reported to be lost from chondrocytes of the superficial zone and shifted to the middle zone in OA cartilage in a targeted study<sup>22</sup>. *WNT5B*, a ligand for frizzled receptors in the WNT signaling pathway, has been found to be differentially transcribed in osteoarthritic bone, consistent with current understanding that OA is a disease involving both cartilage and bone<sup>56</sup>.

We identify extracellular matrix organization, collagen catabolism and angiogenesis as biological pathways that have significant association with disease progression, in agreement with the consensus findings of previous methylation and gene expression studies of OA<sup>7</sup>. These results corroborate that increased ECM turnover is a crucial component in OA pathogenesis. We found suggestive evidence of a link between some of these pathways and genetic OA risk loci (see Supplementary Note, Supplementary Table S11). These signals would not have been identified directly from GWAS data, highlighting the importance of synthesizing information from multiple molecular levels to obtain a more powerful integrated view. Follow-up mechanistic studies will be required before causal relationships between the identified pathways and OA progression can be established. In the chondrocyte, we found little evidence of differential inflammatory pathway activity between the intact and degraded samples<sup>57,58</sup>. This is not surprising, as all of the patients studied here had a diagnosis of OA and clinically advanced disease in at least one location within the joint. Inflammatory mediators are soluble factors present throughout the joint, to which both the healthy and the diseased chondrocyte populations are potentially exposed, and for which the regulatory molecules may differ compared to the articular cartilage tissue studied here.

Based on our discovery data, we estimate that ~95% of significantly different genes identified in this study are true positives (Supplementary Fig. S5). The replication data confirm the high true positive rate and yield

| Gene symbol | Protein                             | Compound                                | FDA marketing status | Target specific mechanism     | Drug mechanism of action  | Reference  |
|-------------|-------------------------------------|---|----------------------|-------------------------------|---|--|
| ANPEP       | Aminopeptidase N                    | Ezetimibe                               | Prescription         | binds protein                 | Anti-hyperlipidemic medication. Used to lower cholesterol absorption in the small intestine.  | Kramer <i>et al.</i> (2005) PMID: 15494415             |
| ANPEP       | Aminopeptidase N                    | Icatibant                               | Prescription         | inhibits protein              | Synthetic peptidomimetic drug consisting of ten amino acids, 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 <i>et al.</i> (2006) PMID: 17026984            |
| AQP1        | Aquaporin-1                         | Acetazolamide                           | 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 <i>et al.</i> (2004) PMID: 15169637              |
| BGLAP       | Osteocalcin                         | Phylloquinone (vitamin K <sub>1</sub> ) | NA (vitamin)         | agonist, carboxylates protein | Fat-soluble vitamin necessary for post-translational modification of certain proteins, mostly required for blood coagulation.   | Schurgers <i>et al.</i> (2001) PMID: 11374034          |
| CLEC3B      | Tetranectin                         | Tenecteplase                            | Prescription         | binds protein                 | Tissue plasminogen activator (tPA). Used as a thrombolytic agent.   | Westergaard <i>et al.</i> <sup>46</sup> PMID: 12694198 |
| CXCL12      | Stromal cell-derived factor 1       | Tinzaparin                              | Prescription         | binds protein                 | Low molecular weight heparin (LMWH). Used in the treatment and prophylaxis of venous thrombo-embolism.  | Koo <i>et al.</i> (2008) PMID: 18991783                |
| FGFR2       | Fibroblast growth factor receptor 2 | Palifermin                              | Prescription         | binds protein                 | Recombinant human keratinocyte growth factor (KGF). Used to treat oral mucositis in patients undergoing cancer chemotherapy.  | Beaven <i>et al.</i> (2007) PMID: 17728847             |
| MAP1A       | Microtubule-associated protein 1A   | Estramustine                            | Prescription         | disrupts protein              | A nitrogen mustard linked to estradiol. Used in palliative care of prostatic neoplasms.   | Stearns <i>et al.</i> (1991) PMID: 1647395             |
| PTGIS       | Prostacyclin synthase               | Non-steroidal anti-inflammatory agents  | Prescription         | inhibits protein              | COX1 and COX2 inhibitors used in the symptomatic treatment of OA.   | Reed <i>et al.</i> (1985) PMID: 3917545                |
| S100A4      | Protein S100-A4                     | Trifluoperazine                         | Prescription         | inhibits protein function     | A phenothiazine with actions similar to chlorpromazine. Used as an antipsychotic and an antiemetic.   | Malashkevish <i>et al.</i> (2010) PMID: 20421509       |

**Table 1.** Results of Drugbank<sup>25</sup> ([www.drugbank.ca](http://www.drugbank.ca)) search for therapeutic compounds with current FDA marketing authorisation for a clinical indication and a potential role in OA treatment. The mechanisms of action and references are taken from Drugbank.

independent strong evidence for multiple novel OA genes. This shows that the gene expression and methylation changes identified are robust. Within the constraints of this study, the changes between intact and degraded OA cartilage are largely joint-independent. Further work will be needed to narrow down which individual gene-level differences between intact and degraded cartilage are joint-independent and which are specific to a joint. This work is complementary to previous studies which examined differences in lesion-distant cartilage between joints<sup>59</sup> or by pooling intact and degraded cartilage<sup>16</sup>, as neither of these studies examined differences between intact and degraded cartilage, and whether such differences are joint-independent.

Notably, this study is a proof-of-concept for integrative deep molecular phenotyping across methylation, gene expression, and protein abundance. As such, it was not powered to provide an exhaustive list of molecular targets and pathways. Indeed, we estimate that only ~10% of the true differentially expressed genes are statistically significant in this study (Supplementary Fig. S5). The sample size could also affect the degree of overlap and agreement between the methylation, gene expression, and protein abundance (this overlap could also be affected by difficulties in assigning the effects of methylation changes to genes: for example, it is possible that the expression of a gene is affected by methylation changes in a distal enhancer, or that a given gene contains an enhancer region for a different gene, and thus methylation of the first gene also affects the expression of the second). Larger sample sizes will be required for a more powerful characterisation of the molecular changes occurring with disease progression. Moreover, investigations of further OA-relevant cell types (including synoviocytes and adipocytes) will be necessary to identify disease-related changes in other tissues, and the biological mechanisms specific to such tissues.

In summary, the integrative functional genomics approach undertaken here has identified biological changes in disease-relevant tissues, highlighting three genes and several pathways that are involved on all three levels examined. Moreover, the approach identified nine genes with changes on multiple molecular levels that are already targeted by drugs approved for human use, highlighting the potential of discovering targets for intervention. These drugs have established safety profiles and pharmacokinetic data for use in humans, which would shorten the investigative pipeline to clinical use in OA. These agents cover a broad range of mechanisms of action and represent novel investigational targets for 'first in disease' studies of OA progression. Further studies will be necessary to comprehensively characterize the molecular signatures of OA.



## References

- Vos, T. *et al.* 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. *The Lancet* **380**, 2163–2196, doi:10.1016/s0140-6736(12)61729-2 (2012).
- Dieppe, P. A. & Lohmander, L. S. Pathogenesis and management of pain in osteoarthritis. *The Lancet* **365**, 965–973, doi:10.1016/s0140-6736(05)71086-2 (2005).
- Valdes, A. M. & Spector, T. D. Genetic epidemiology of hip and knee osteoarthritis. *Nat Rev Rheumatol* **7**, 23–32, doi:10.1038/nrrheum.2010.191 (2011).
- Reynard, L. N. & Loughlin, J. The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev. Mol. Med.* **15**, doi:10.1017/erm.2013.4 (2013).
- Ruiz-Romero, C., Fernández-Puente, P., Calamia, V. & Blanco, F. J. Lessons from the proteomic study of osteoarthritis. *Expert Review of Proteomics* **12**, 433–443, doi:10.1586/14789450.2015.1065182 (2015).
- Ramos, Y. F. & Meulenbelt, I. The role of epigenetics in osteoarthritis: current perspective. *Curr Opin Rheumatol*. doi:10.1097/bor.0000000000000355 (2016).
- Steinberg, J. & Zeggini, E. Functional genomics in osteoarthritis: Past, present, and future. *Journal of orthopaedic research: official publication of the Orthopaedic Research Society* **34**, 1105–1110, doi:10.1002/jor.23296 (2016).
- Mankin, H. J., Dorfman, H., Lippiello, L. & Zarins, A. Biochemical and metabolic abnormalities in articular cartilage from osteoarthritic human hips. II. Correlation of morphology with biochemical and metabolic data. *The Journal of bone and joint surgery. American volume* **53**, 523–537 (1971).
- Pearson, R. G., Kurien, T., Shu, K. S. S. & Scammell, B. E. Histopathology grading systems for characterisation of human knee osteoarthritis – reproducibility, variability, reliability, correlation, and validity. *Osteoarthritis and Cartilage* **19**, 324–331, doi:10.1016/j.joca.2010.12.005 (2011).
- Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, doi:10.1093/bioinformatics/btu638 (2014).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, doi:10.1093/bioinformatics/btp616 (2009).
- Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430, doi:10.1093/bioinformatics/btt684 (2014).
- Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293, doi:10.1186/1471-2164-14-293 (2013).
- Barfield, R. T., Kilaru, V., Smith, A. K. & Conneely, K. N. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics* **28**, 1280–1281, doi:10.1093/bioinformatics/bts124 (2012).
- den Hollander, W. *et al.* Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. *Annals of the Rheumatic Diseases* **73**, 2208–2212, doi:10.1136/annrheumdis-2014-205980 (2014).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, doi:10.1093/bioinformatics/btq033 (2010).
- Ramos, Y. F. M. *et al.* Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK Study. *PLoS ONE* **9**, e103056, doi:10.1371/journal.pone.0103056 (2014).
- Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30, doi:10.1093/nar/28.1.27 (2000).
- Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research* **37**, D619–D622, doi:10.1093/nar/gkn863 (2009).
- Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046, doi:10.1093/bioinformatics/btp536 (2009).
- Nakayama, N. A novel chordin-like BMP inhibitor, CHL2, expressed preferentially in chondrocytes of developing cartilage and osteoarthritic joint cartilage. *Development* **131**, 229–240, doi:10.1242/dev.00901 (2004).
- Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419–1260419, doi:10.1126/science.1260419 (2015).
- Mapp, P. I. & Walsh, D. A. Mechanisms and targets of angiogenesis and nerve growth in osteoarthritis. *Nat Rev Rheumatol* **8**, 390–398, doi:10.1038/nrrheum.2012.80 (2012).
- Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research* **42**, D1091–D1097, doi:10.1093/nar/gkt1068 (2013).
- Attur, M. *et al.* Elevated expression of periostin in human osteoarthritic cartilage and its potential role in matrix degradation via matrix metalloproteinase-13. *The FASEB Journal* **29**, 4107–4121, doi:10.1096/fj.15-272427 (2015).
- Chijimatsu, R. *et al.* Expression and pathological effects of periostin in human osteoarthritis cartilage. *BMC Musculoskelet Disord* **16**, 215, doi:10.1186/s12891-015-0682-3 (2015).
- Lourido, L. *et al.* Quantitative Proteomic Profiling of Human Articular Cartilage Degradation in Osteoarthritis. *J. Proteome Res.* **13**, 6096–6106, doi:10.1021/pr501024p (2014).
- Coutu, D. L. *et al.* Periostin, a Member of a Novel Family of Vitamin K-dependent Proteins, Is Expressed by Mesenchymal Stromal Cells. *Journal of Biological Chemistry* **283**, 17991–18001, doi:10.1074/jbc.m708029200 (2008).
- Misra, D. *et al.* Vitamin K Deficiency Is Associated with Incident Knee Osteoarthritis. *The American Journal of Medicine* **126**, 243–248, doi:10.1016/j.amjmed.2012.10.011 (2013).
- Meulenbelt, I. M. *et al.* The first international workshop on the epigenetics of osteoarthritis. *Connect Tissue Res* **58**, 37–48, doi:10.3109/03008207.2016.1168409 (2017).
- Jeffries, M. A. *et al.* Genome-Wide DNA Methylation Study Identifies Significant Epigenomic Changes in Osteoarthritic Cartilage. *Arthritis & Rheumatology* **66**, 2804–2815, doi:10.1002/art.38762 (2014).
- Moazedi-Fuerst, F. C. *et al.* Epigenetic differences in human cartilage between mild and severe OA. *Journal of Orthopaedic Research* **32**, 1636–1645, doi:10.1002/jor.22722 (2014).
- Karlsson, C. *et al.* Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis and Cartilage* **18**, 581–592, doi:10.1016/j.joca.2009.12.002 (2010).
- Tew, S. R., McDermott, B. T., Fentem, R. B., Peffers, M. J. & Clegg, P. D. Transcriptome-Wide Analysis of Messenger RNA Decay in Normal and Osteoarthritic Human Articular Chondrocytes. *Arthritis & Rheumatology* **66**, 3052–3061, doi:10.1002/art.38849 (2014).
- Snelling, S. *et al.* A gene expression study of normal and damaged cartilage in anteromedial gonarthrosis, a phenotype of osteoarthritis. *Osteoarthritis and Cartilage* **22**, 334–343, doi:10.1016/j.joca.2013.12.009 (2014).
- Dunn, S. L. *et al.* Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses. *Osteoarthritis and Cartilage* **24**, 1431–1440, doi:10.1016/j.joca.2016.03.007 (2016).
- 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* **11**, 43, doi:10.1186/1477-5956-11-43 (2013).
- Ruiz-Romero, C. *et al.* Proteomic analysis of human osteoarthritic chondrocytes reveals protein changes in stress and glycolysis. *Proteomics* **8**, 495–507, doi:10.1002/pmic.200700249 (2008).

40. den Hollander, W. *et al.* Transcriptional Associations of Osteoarthritis-Mediated Loss of Epigenetic Control in Articular Cartilage. *Arthritis & Rheumatology* **67**, 2108–2116, doi:10.1002/art.39162 (2015).
41. Bonin, C. A. *et al.* Identification of Differentially Methylated Regions in New Genes Associated with Knee Osteoarthritis. *Gene* **576**, 312–318, doi:10.1016/j.gene.2015.10.037 (2016).
42. Rushton, M. D. *et al.* Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Human Molecular Genetics* **24**, 7432–7444, doi:10.1093/hmg/ddv433 (2015).
43. Bush, P. G. & Hall, A. C. The volume and morphology of chondrocytes within non-degenerate and degenerate human articular cartilage. *Osteoarthritis and Cartilage* **11**, 242–251, doi:10.1016/s1063-4584(02)00369-2 (2003).
44. Musumeci, G. *et al.* Aquaporin 1 (AQP1) expression in experimentally induced osteoarthritic knee menisci: An *in vivo* and *in vitro* study. *Tissue and Cell* **45**, 145–152, doi:10.1016/j.tice.2012.10.004 (2013).
45. Geyer, M. *et al.* Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. *Osteoarthritis and Cartilage* **17**, 328–335, doi:10.1016/j.joca.2008.07.010 (2009).
46. Westergaard, U. B., Andersen, M. H., Heegaard, C. W., Fedosov, S. N. & Petersen, T. E. Tetranectin binds hepatocyte growth factor and tissue-type plasminogen activator. *Eur J Biochem* **270**, 1850–1854, doi:10.1046/j.1432-1033.2003.03549.x (2003).
47. Valdes, A. M. *et al.* Association study of candidate genes for the prevalence and progression of knee osteoarthritis. *Arthritis & Rheumatism* **50**, 2497–2507, doi:10.1002/art.20443 (2004).
48. Tchetina, E. V. Developmental Mechanisms in Articular Cartilage Degradation in Osteoarthritis. *Arthritis* **2011**, 1–16, doi:10.1155/2011/683970 (2011).
49. Xia, B. *et al.* Osteoarthritis Pathogenesis: A Review of Molecular Mechanisms. *Calcif Tissue Int* **95**, 495–505, doi:10.1007/s00223-014-9917-9 (2014).
50. Remst, D. F. G. *et al.* Gene Expression Analysis of Murine and Human Osteoarthritis Synovium Reveals Elevation of Transforming Growth Factor  $\beta$ -Responsive Genes in Osteoarthritis-Related Fibrosis. *Arthritis & Rheumatology* **66**, 647–656, doi:10.1002/art.38266 (2014).
51. Halpain, S. & Dehmelt, L. The MAP1 family of microtubule-associated proteins. *Genome Biology* **7**, 224, doi:10.1186/gb-2006-7-6-224 (2006).
52. Blain, E. J. Involvement of the cytoskeletal elements in articular cartilage homeostasis and pathology. *International Journal of Experimental Pathology* **90**, 1–15, doi:10.1111/j.1365-2613.2008.00625.x (2009).
53. Kanenari, M., Zhao, J. & Abiko, Y. Enhancement of microtubule-associated protein-1 Alpha gene expression in osteoblasts by low level laser irradiation. *Laser Ther* **20**, 47–51 (2011).
54. Péterfi, Z. & Geiszt, M. Peroxidases: novel players in tissue genesis. *Trends in Biochemical Sciences* **39**, 305–307, doi:10.1016/j.tibs.2014.05.005 (2014).
55. Balakrishnan, L. *et al.* Proteomic analysis of human osteoarthritis synovial fluid. *Clin Proteomics* **11**, 6, doi:10.1186/1559-0275-11-6 (2014).
56. Hopwood, B., Tsykin, A., Findlay, D. M. & Fazzalari, N. L. Microarray gene expression profiling of osteoarthritic bone suggests altered bone remodelling, WNT and transforming growth factor- $\beta$ /bone morphogenic protein signalling. *Arthritis Res Ther* **9**, R100, doi:10.1186/ar2301 (2007).
57. Goldring, M. B. & Otero, M. Inflammation in osteoarthritis. *Current Opinion in Rheumatology* **23**, 471–478, doi:10.1097/bor.0b013e328349c2b1 (2011).
58. Wojdasiewicz, P., Poniatowski, L. A. & Szukiewicz, D. The Role of Inflammatory and Anti-Inflammatory Cytokines in the Pathogenesis of Osteoarthritis. *Mediators of Inflammation* **2014**, 1–19, doi:10.1155/2014/561459 (2014).
59. Rushton, M. D. *et al.* Characterization of the Cartilage DNA Methylome in Knee and Hip Osteoarthritis. *Arthritis & Rheumatology* **66**, 2450–2460, doi:10.1002/art.38713 (2014).

## Acknowledgements

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. 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](http://www.wtccc.org.uk). We thank all participants of the RAAK study. We would like to acknowledge the Cambridge Biomedical Research Center Human Research Tissue Bank for storing the hip gene expression and methylation replication samples. This work was funded by the Wellcome Trust (WT098051). This study utilized genotype data from arcOGEN (<http://www.arcogen.org.uk/>) funded by a special purpose grant from Arthritis Research UK (grant 18030). The RAAK study is supported by Leiden University Medical Center. Research leading to the RAAK study results has received funding from the Dutch Arthritis Association (DAA 2010\_017) and the IDEAL project (European Union's Seventh Framework Program (FP7/2007–2011) under grant agreement no. 259679). GRSR was supported by the European Molecular Biology Laboratory and the Wellcome Trust Sanger Institute through an EBI-Sanger Postdoctoral Fellowship. RAB received funding from the National Institute for Health Research (Cambridge Biomedical Research Centre).

## Author Contributions

Study design: E.Z., J.M.W. Omics data analysis: J.S., G.R.S.R., T.I.R. Discovery data sample collection: J.M.W. Knee replication data (non-RAAK) sample collection: R.L.J., M.J.C., K.M.S., J.M.W. Hip replication data (non-RAAK) sample collection (including protocol development): R.A.B., A.W.M. Proteomics: J.S.C., T.I.R. Histology and immunohistochemistry: C.L.L.M., A.L.A.B. Western blotting: M.P., R.C. Gene expression replication data from RAAK study: Y.F.M.R., R.G.H.H.N., I.M. Manuscript writing and editing: J.S., G.R.S.R., T.I.R., R.A.B., M.P., C.L.L.M., Y.F.M.R., I.M., A.W.M., J.S.C., J.M.W., E.Z.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-09335-6

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017