Advances in osteoarthritis genetics

Kalliope Panoutsopoulou, Eleftheria Zeggini

ABSTRACT

Osteoarthritis (OA), the most common form of arthritis, is a highly debilitating disease of the joints and can lead to severe pain and disability. There is no cure for OA. Current treatments often fail to alleviate its symptoms leading to an increased demand for joint replacement surgery. Previous epidemiological and genetic research has established that OA is a multifactorial disease with both environmental and genetic components. Over the past 6 years, a candidate gene study and several genome-wide association scans (GWAS) in populations of Asian and European descent have collectively established 15 loci associated with knee or hip OA that have been replicated with genome-wide significance, shedding some light on the aetiology of the disease. All OA associated variants to date are common in frequency and appear to confer moderate to small effect sizes. Some of the associated variants are found within or near genes with clear roles in OA pathogenesis, whereas others point to unsuspected, less characterised pathways. These studies have also provided further evidence in support of the existence of ethnic, sex, and joint specific effects in OA and have highlighted the importance of expanded and more homogeneous phenotype definitions in genetic studies of OA.

INTRODUCTION

Osteoarthritis (OA) is a set of disorders of the musculoskeletal system characterised by degradation and loss of articular cartilage in synovial joints most commonly of the knee, hip, hand, foot, and spine. OA development appears to be a result of a complex set of interactions between mechanical, biological, biochemical, and molecular factors that destabilise the normal coupling of degradation and synthesis of articular cartilage chondrocytes and extracellular matrix, and subchondral bone. Although OA invariably involves articular cartilage, it affects all tissues of the joint; loss of articular cartilage is accompanied by subchondral bone remodelling with sclerosis and in many instances cysts, osteophyte formation at joint margins, ligamentous contractures and relaxation, muscle atrophy and spasms, and at clinical stages of the disease inflammation of the synovial membrane.

The health and socioeconomic burden posed by OA is substantial. The main symptom of OA is pain and loss of physical function leading to impaired mobility and impaired quality of life. Current regimens for OA management are multimodal in nature—that is, a combination of pharmacologic and non-pharmacologic treatments. However, these are often ineffective in targeting the main disease symptom leading to an increased demand for total joint replacement (TJR). OA is the most prevalent form of arthritis affecting over 40% of people over the age of 70, and its incidence is on the rise. In the USA alone 27 million adults had clinical evidence of OA in 2005, a rise of nearly 30% from the estimate of 21 million in 1995. With longer life expectancies and the obesity pandemic—with age and obesity/overweight being well established risk factors for disease development and progression—the prevalence of OA is expected to increase continuously and sharply.

Although the aetiology of OA is not fully understood it has been well established that the disease is caused by complex interplay between environmental and genetic factors. Age is the strongest risk factor for all types of OA whereas obesity appears to confer the greatest risk in knee OA, particularly among women. Epidemiological research also suggests that occupational physical workload, high sporting activity, joint injuries and being female may increase the risk of developing OA at particular joints (reviewed in Altman and Bierma-Zeinstra and Koes).

GENETIC STUDIES IN OA

The pre-genome-wide association scans era

Twin pair, sibling risk and segregation studies conducted in Europe and the USA have demonstrated a substantial genetic component for OA that is transmitted in a non-Mendelian manner, which is typical of multifactorial diseases. Heritability estimates range between 40–65%, with precise estimates varying depending on gender, affected joint, and severity of the disease, but overall appear stronger for hand and hip OA than for knee OA. Familial aggregation studies in the UK have estimated that the sibling recurrence risk (λs)—which indicates the disease risk of a sibling to an individual with OA compared to the disease prevalence in the general population—is ∼5. The notion that OA is simply a wear-and-tear disease of old age was largely superseded and these epidemiological studies provided a firm foundation for considerable genetic research aimed at identifying genetic loci responsible for OA susceptibility.

To date five genome-wide linkage scans performed on individuals collected in the UK, Finland, Iceland, and the USA have been published for OA but had limited success. Gene centric association studies, commonly known as candidate gene studies, have been extensively applied in populations of European and Asian ancestry to survey variants across genes believed to be implicated in OA based on prior biological knowledge. The majority of reported associations, however, have been either false positives—due to small sample sizes, lack of replication and lack of stringency in the reporting of significant results based on observed p values—or have yielded only suggestive evidence for association; that is, replication in at least one other study but not meeting genome-wide significance defined...
as $p<5\times10^{-8}$ (for examples, see Valdes and Spector\textsuperscript{10}). A notable exception of the success of the candidate gene sequencing approach in OA is the robust and reproducible association of rs143383 in the growth differentiation factor 5 (GDF5) gene,\textsuperscript{11–13} discussed in more detail below.

### The genome-wide association scans era

In the last decade, the Human Genome and International HapMap Projects have revolutionised the field of common complex disease genetics by providing an extensive catalogue of genome sequence variation and linkage disequilibrium (LD) patterns between common variants. This has enabled the selection of tag single nucleotide polymorphisms (tag SNPs)—a set of informative, non-redundant markers capturing the majority of common variations across the genome—which led to the development of high throughput genotyping platforms in which hundreds of thousands of SNPs can be concurrently examined for association with disease. In recent years, this hypothesis-free approach of interrogating common variation in a genome-wide manner dominated the field of human genetics and led to the identification of numerous novel associations with several common complex diseases and traits.\textsuperscript{1–4} OA was relatively late to enter the genome-wide association scans (GWAS) era but the returns were substantial: two novel associations from studies in individuals of Asian origin,\textsuperscript{15, 16} and 12 novel associations from scans performed in individuals of European ancestry,\textsuperscript{17–22} were detected with genome-wide significance bringing the total of established OA loci to 15 (table 1). This review considers all associations with OA that have surpassed or have approached closely the stringent threshold of genome-wide significance following replication in at least one independent dataset. Extensive fine mapping and functional studies are required to identify the causal variants and precise genes involved in OA pathogenesis.

### GENETIC ARCHITECTURE OF OA

In line with other common complex disorders the genetic architecture of OA appears to be highly polygenic with multiple variants across the full allele frequency spectrum contributing modest and small effects. The theory of a polygenic inheritance model for OA was first tested by the arcOGEN Consortium in a GWAS of 3177 cases and 4984 population based controls from the UK.\textsuperscript{23} Using analytical approaches previously applied to test the polygenic inheritance of schizophrenia and bipolar disorder,\textsuperscript{24} a set of independent associated SNPs was derived from a subset of the data (90% of arcOGEN samples); this score allele set was then used to evaluate the proportion of case-control status accounted for in the remaining samples (10% of arcOGEN samples). These analyses revealed a substantial genetic component to OA comprising multiple contributing variants with small effect sizes.

### OA ESTABLISHED LOCI

#### GDF5

In the only candidate gene study that yielded a strong reproducible association with OA thus far, Miyamoto et al.,\textsuperscript{11} searched for sequence variations in the exons and flanking regions of the GDF5 gene and identified the rs143383 polymorphism—a $T$ to $C$ transition located in the 5’ untranslated region (5’UTR) of the gene—to be significantly associated with hip OA. Combined evidence for association in two independent Japanese populations reached genome-wide significance with $p=2\times10^{-12}$ and allelic odds ratio (OR) of 1.79 (95% CI 1.53 to 2.09) (table 1).

### Table 1: Genetic associations with osteoarthritis established with genome-wise significance following replication in at least one independent dataset

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest* gene(s)</th>
<th>EA</th>
<th>EAF</th>
<th>OR, 95% CI</th>
<th>p Value</th>
<th>Site</th>
<th>Sex</th>
<th>Ethnic group</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs143383t</td>
<td>GDF5</td>
<td>T</td>
<td>0.74</td>
<td>1.79,1.53 to 2.09</td>
<td>$2\times10^{-13}$</td>
<td>Hip</td>
<td>Both</td>
<td>Asian</td>
<td>11</td>
</tr>
<tr>
<td>rs143383t</td>
<td>GDF5</td>
<td>T</td>
<td>0.16</td>
<td>1.11 to 1.22</td>
<td>8.3$\times10^{-9}$</td>
<td>Knee</td>
<td>Both</td>
<td>European</td>
<td>13</td>
</tr>
<tr>
<td>rs7636618</td>
<td>DWA</td>
<td>G</td>
<td>0.63</td>
<td>1.43, 1.28 to 1.59</td>
<td>7.3$\times10^{-11}$</td>
<td>Knee</td>
<td>Both</td>
<td>Asian</td>
<td>15</td>
</tr>
<tr>
<td>rs7777228</td>
<td>HLA-DQB1</td>
<td>T</td>
<td>0.62</td>
<td>1.34, 1.21 to 1.49</td>
<td>2.4$\times10^{-8}$</td>
<td>Knee</td>
<td>Both</td>
<td>Asian</td>
<td>16</td>
</tr>
<tr>
<td>rs10947262</td>
<td>BTN2L</td>
<td>C</td>
<td>0.58</td>
<td>1.31, 1.20 to 1.44</td>
<td>5.1$\times10^{-9}$</td>
<td>Knee</td>
<td>Both</td>
<td>Asian and European</td>
<td>17</td>
</tr>
<tr>
<td>rs3815148q</td>
<td>COGS††</td>
<td>C</td>
<td>0.23</td>
<td>1.14, 1.09 to 1.19</td>
<td>8$\times10^{-8}$</td>
<td>Knee and hand</td>
<td>Both</td>
<td>European</td>
<td>18</td>
</tr>
<tr>
<td>rs4370250q</td>
<td>DUS4L¶</td>
<td>G</td>
<td>0.17</td>
<td>1.17, 1.11 to 1.24</td>
<td>9.2$\times10^{-9}$</td>
<td>Knee</td>
<td>Both</td>
<td>European</td>
<td>19</td>
</tr>
<tr>
<td>rs1184287q</td>
<td>MF2L</td>
<td>A</td>
<td>0.93</td>
<td>1.17, 1.11 to 1.23</td>
<td>2.1$\times10^{-8}$</td>
<td>Knee and hip</td>
<td>Both</td>
<td>European</td>
<td>20</td>
</tr>
<tr>
<td>rs6976**</td>
<td>GLB1D1††</td>
<td>T</td>
<td>0.37</td>
<td>1.12, 1.08 to 1.16</td>
<td>7.2$\times10^{-11}$</td>
<td>Knee and hand</td>
<td>Both</td>
<td>European</td>
<td>21</td>
</tr>
<tr>
<td>rs11177**</td>
<td>GNL3t</td>
<td>A</td>
<td>0.38</td>
<td>1.12, 1.08 to 1.16</td>
<td>1.3$\times10^{-10}$</td>
<td>Knee and hand</td>
<td>Both</td>
<td>European</td>
<td>22</td>
</tr>
<tr>
<td>rs4836732</td>
<td>ASTN2</td>
<td>C</td>
<td>0.47</td>
<td>1.2, 1.13 to 1.27</td>
<td>6.1$\times10^{-10}$</td>
<td>Knee</td>
<td>Females</td>
<td>European</td>
<td>23</td>
</tr>
<tr>
<td>rs9505091</td>
<td>FILIP1; SENP6</td>
<td>T</td>
<td>0.11</td>
<td>1.18, 1.12 to 1.25</td>
<td>2.4$\times10^{-9}$</td>
<td>Hip</td>
<td>Both</td>
<td>European</td>
<td>24</td>
</tr>
<tr>
<td>rs10492367</td>
<td>KLHDC5; PTHLH</td>
<td>T</td>
<td>0.19</td>
<td>1.14, 1.09 to 1.20</td>
<td>1.5$\times10^{-9}$</td>
<td>Hip</td>
<td>Both</td>
<td>European</td>
<td>25</td>
</tr>
<tr>
<td>rs853487</td>
<td>CHST11</td>
<td>G</td>
<td>0.34</td>
<td>1.13, 1.09 to 1.18</td>
<td>1.6$\times10^{-9}$</td>
<td>Hip</td>
<td>Both</td>
<td>European</td>
<td>26</td>
</tr>
<tr>
<td>rs12107036</td>
<td>TP63</td>
<td>G</td>
<td>0.52</td>
<td>1.21, 1.13 to 1.29</td>
<td>6.7$\times10^{-8}$</td>
<td>Knee</td>
<td>Females</td>
<td>European</td>
<td>27</td>
</tr>
<tr>
<td>rs8044697‡‡</td>
<td>FTO</td>
<td>C</td>
<td>0.5</td>
<td>1.11, 1.07 to 1.15</td>
<td>6.9$\times10^{-8}$</td>
<td>Knee and hand</td>
<td>Females</td>
<td>European</td>
<td>28</td>
</tr>
<tr>
<td>rs10948172</td>
<td>SPUTFI; CDCSL</td>
<td>G</td>
<td>0.29</td>
<td>1.14, 1.09 to 1.20</td>
<td>7.9$\times10^{-9}$</td>
<td>Knee and hand</td>
<td>Males</td>
<td>European</td>
<td>29</td>
</tr>
<tr>
<td>rs6094710</td>
<td>NCOD3</td>
<td>A</td>
<td>0.04</td>
<td>1.28, 1.18 to 1.39</td>
<td>7.9$\times10^{-9}$</td>
<td>Hip</td>
<td>Both</td>
<td>European</td>
<td>30</td>
</tr>
<tr>
<td>rs1298274±</td>
<td>DOTT1</td>
<td>C</td>
<td>NA</td>
<td>1.17, 1.11 to 1.23</td>
<td>7.8$\times10^{-9}$</td>
<td>Hip</td>
<td>Females</td>
<td>Males</td>
<td>31</td>
</tr>
</tbody>
</table>

*Nearest gene(s) only shown.
†Summary statistics of the same SNP in separate studies in Asians and Europeans, respectively.
‡SNPs in strong linkage disequilibrium.
¶SNPs in strong linkage disequilibrium.
§chrQ2 locus encompasses more genes than shown here, for full details see Kerkhof et al.\textsuperscript{17} and Day-Williams et al.\textsuperscript{19}
**SNPs in strong linkage disequilibrium.
††chr3p21.1 locus encompasses more genes than shown here, for full details see arcOGEN Consortium.\textsuperscript{20}
‡‡SNPs in strong linkage disequilibrium.
\textsuperscript{1}This signal was attenuated after BMI adjustment, suggesting that the FTO locus exerts its effect on OA through obesity, BMI, body mass index; EA, Effect allele; EAF, effect allele frequency; OA, osteoarthritis; SNP, single nucleotide polymorphism.
Evidence for association with knee OA was weaker in datasets from China (p=3×10^{-4}) and Japan (p=0.002). A large scale meta-analysis employing 4791 hip OA cases and 6006 controls, and 4367 knee OA cases and 6291 controls, showed that in samples of European descent there was less compelling evidence for association with hip OA (OR=1.07, 95% CI 1.01 to 1.14; p=0.034) and more compelling evidence for association with knee OA (OR=1.13, 95% CI 1.06 to 1.20; p=9×10^{-5}), but with a much weaker effect size than the East Asian set. These differences can be ascribed to allele frequency disparities between ethnic groups. The GDF5 SNP was eventually found to be genome-wide significantly associated with knee OA in Europeans in a subsequent meta-analysis across a total of 6861 knee OA cases and 10 103 controls (table 1). Genetic variation in the GDF5 locus has also been robustly associated with height variation, and linked with suggestive significance to lumbar disc degeneration, fracture risk, congenital dislocation of the hip, and Achilles tendon pathology, suggesting a pleiotropic effect from this gene.

GDF5, alternatively known as cartilage derived morphogenetic protein 1 or bone morphogenetic protein 14, is an extracellular signalling molecule, a member of the transforming growth factor (TGF-β) superfamily. Differential allelic expression analysis supported a functional role of the rs143383 polymorphism with the risk allele (T) mediating reduced GDF5 transcription relative to the C allele.

Further studies in joint tissues (cartilage, synovium, meniscus, ligament, tendon, and fat pad) obtained from individuals undergoing elective joint replacement for OA demonstrated a consistent allelic expression imbalance in all tissues tested, implying that the functional effect mediated by rs143383 on GDF5 expression is joint-wide. The role of GDF5 in the development, maintenance, and repair of bone, cartilage, and other tissues of the synovial joint has been extensively reviewed. Mutations in the GDF5 gene have been previously implicated in a broad spectrum of skeletal disorders in humans (for an overview see Cornellis et al) and mice.

**DVWA**

Through a GWAS interrogating ~100 000 SNPs, Miyamoto et al identified a previously unknown gene on chromosome 3p24.3, DVWA, to be associated with susceptibility to knee osteoarthritis in Japanese individuals. The association of rs7639618—a missense SNP—was replicated in additional Japanese and Han Chinese cohorts with p=7.3×10^{-11} and OR=1.43 (95% CI 1.28 to 1.59) (table 1).

DVWA encodes a 276 amino-acid protein with two regions corresponding to the von Willebrand factor type A domain (VWA domain). DVWA expression studies in various human tissues revealed highest expression in cartilage tissues from both controls and individuals with OA suggesting that DVWA function is associated with cartilage. Mutations in the VWA domains of a different gene (MATN3) have been previously associated with hand OA in an Icelandic linkage scan, and with multiple epiphyseal dysplasia. DVWA protein binds to β-tubulin, and the binding is weakened when the risk alleles are taken forward for functional analysis. Immunohistochemistry experiments showed that the GPR22 protein was present in cartilage and osteophytes in OA mouse models but absent from normal cartilage, providing some indication that GPR22 could be the causal gene. Further gene expression studies using joint tissues from OA patients and control cartilage from patients who had neck of the femur fractures found significantly lower expression levels in OA cartilage compared with control cartilage for five genes in the region—the exception being GPR22.

**Chr7q22 locus**

The first novel locus for OA that reached genome-wide significance in Europeans was reported by a GWAS in Dutch individuals in a gene dense region on chromosome 7q22. Following large scale replication, allele C at rs3815148 was found to be associated with knee and/or hand OA with p=8×10^{-8} and OR=1.14 (95% CI 1.09 to 1.19) (table 1). The association with knee OA was further corroborated and reinforced by a meta-analysis across four other GWAS (deCODE, Rotterdam, Framingham, TwinsUK) performed under the auspices of the TreatOA Consortium.

The chr7q22 locus harbours six genes, PRKAR2B, HPB1, COG5, GPR22, DUS4L, and BCAP29, within a large linkage disequilibrium block making it difficult to pin down the culprit gene. Since the GPR22 gene encodes a G-protein coupled receptor which is an attractive, potential drug target, this gene was taken forward for functional analysis. Immunohistochemistry experiments showed that the GPR22 protein was present in cartilage and osteophytes in OA mouse models but absent from normal cartilage, providing some indication that GPR22 could be the causal gene. Further gene expression studies using joint tissues from OA patients and control cartilage from patients who had neck of the femur fractures found significantly lower expression levels in OA cartilage compared with control cartilage for five genes in the region—the exception being GPR22.
which was not detected.\textsuperscript{53} Carriers of the OA risk allele showed a significant reduction in expression of \textit{HBP1} (HMG-box transcription factor which encodes a transcriptional repressor) in cartilage and synovium and of \textit{DUS4L} (dihydrouridine synthase 4-Like) in fat pad.

**MCF2L**

Using 1000 Genomes Project based imputation in a GWAS for OA by the arcOGEN Consortium (3177 OA cases and 4894 controls), UK scientists were able to establish the third novel locus for OA in Europeans and the first common complex disease locus to be identified via 1000G imputation.\textsuperscript{19} Following large scale replication, rs11842874 in intron 4 of MCF2L (MCF2 cell line derived transforming sequence-like, encoding the rho-specific guanine nucleotide exchange factor) reached genome-wide significance with \( p=2.1×10^{-8} \) and OR=1.17 (95\% CI 1.11 to 1.23) (table 1).\textsuperscript{20}

MCF2L has been implicated in both skeletal and pain related outcomes of OA. MCF2L rat models of OA have shown that the protein is expressed in articular chondrocytes.\textsuperscript{34, 55} Another study in Mcf2L rat models found that its expression was highest in 5-week-old rat brain sections, and could be localised to neurons and \( \alpha \)-tanyctyes—bipolar cells in the hypothalamus bridging the cerebrospinal fluid to the portal capillaries—suggesting that Ost may participate in axonal transport in these specialised cells.\textsuperscript{56} In zebrashif, mcf2l was dynamically expressed in a range of cell types during development, including Kupffer’s vesicle.\textsuperscript{7} Diffuse expression of mcf2l was observed in the zebrashif brain throughout development—consistent with the strong expression seen in the brain in rat—and in the developing zebrashif jaw cartilages, which suggests that mcf2l could also have a function in cartilage development. In human cells, mcf2l has been shown to regulate neurotrophin-3-induced cell migration in Schwann cells.\textsuperscript{58} Neurotrophin-3 is a member of the nerve growth factor (NGF) family. Knee OA patients treated with a humanised monoclonal antibody that inhibits NGF have shown improvements in both joint function and pain outcomes.\textsuperscript{59, 60}

Chr3p21.1 locus

The following eight loci were discovered by the largest single GWAS for knee and/or hip OA to date, performed by the arcOGEN Consortium in 7410 cases and 11 009 population based controls from the UK, and confirmed in replication efforts including up to 7473 cases and 42 938 controls of European descent.\textsuperscript{20}

Two perfectly correlated SNPs in chr3p21.1 situated in an extended LD block comprising over 30 genes were associated with OA, and association was stronger in patients ascertained by thearcOGEN Consortium in 7410 cases and 4894 controls), UK scientists were able to establish the third novel locus for OA in Europeans and the first common complex disease locus to be identified via 1000G imputation.\textsuperscript{19} Following large scale replication, rs11842874 in intron 4 of MCF2L (MCF2 cell line derived transforming sequence-like, encoding the rho-specific guanine nucleotide exchange factor) reached genome-wide significance with \( p=2.1×10^{-8} \) and OR=1.17 (95\% CI 1.11 to 1.23) (table 1).\textsuperscript{20}

**ASTN2**

rs4836732 located within intron 18 of the ASTN2 gene was found to be most highly associated with female total hip replacement (THR) (\( p=6.11×10^{-10} \), OR=1.20, 95\% CI 1.13 to 1.27) (table 1).\textsuperscript{20} ASTN2 (astrotactin 2) is a membrane protein that regulates surface levels of ASTN1 during neuronal migration\textsuperscript{64} and is highly expressed in the developing and adult brain. An intronic SNP within ASTN2 has been shown to have some evidence of involvement with the pathogenesis of adult attention deficit hyperactivity disorder (ADHD).\textsuperscript{65} In rare CNV analysis it has been shown that exonic deletion and duplication in the ASTN2 locus is associated with schizophrenia.\textsuperscript{66}

**FILIP1; SENP6**

rs9350591 was found to be significantly associated with hip OA (\( p=2.42×10^{-9} \), OR=1.18, 95\% CI 1.12 to 1.25) (table 1).\textsuperscript{20} This variant is located 38 kb upstream of \textit{FILIP1} (filamin A interacting protein 1) and 70 kb upstream of SENP6 (sentrin specific peptidase 6). The role of these poorly characterised genes in OA has not been explored yet. However, \textit{COL12A1} (collagen, type XII, \( \alpha 1 \)) is found \( ∼326 \) kb away from the index SNP. Type XII collagen is expressed by osteoblasts and localises to the peristem—aactive area of bone formation. Col12a\( 1 \)/−/− mice exhibit several skeletal abnormalities and alterations in the organisation and polarisation of osteoblasts, suggesting a role for type XII collagen in osteoblast differentiation and bone matrix formation.\textsuperscript{67}

**KLHDC5; PTHLH**

rs10492367 reached genome-wide significance in the hip OA analysis (\( p=1.48×10^{-8} \), OR=1.14, 95\% CI 1.09 to 1.20) (table 1).\textsuperscript{20} This SNP is situated 59 kb downstream of \textit{KLHDC5} (kelch domain containing 5) and 96 kb downstream of \textit{PTHLH} (parathyroid hormone-like hormone). \textit{PTHLH} presents an excellent candidate gene for OA as this hormone is known to regulate endochondral ossification (ie, bone development) by inhibiting chondrocytes from hypertrophy (reviewed in Zhang et al\textsuperscript{68}). Parathyroid hormone related peptide expression is higher in chondrocytes from pathologic articular cartilage than from normal cartilage of humans.\textsuperscript{69} Pthrp\( 1/-/− \) mice that survived gestation have accelerated differentiation of chondrocytes in bone.\textsuperscript{70}

**CHST11**

rs835487 within intron 2 of \textit{CHST11} was found to be most significantly associated with THR (\( p=1.64×10^{-8} \), OR=1.13, 95\% CI 1.09 to 1.18) (table 1).\textsuperscript{20} \textit{CHST11} (carbohydrate sulfotransferase 11), also known as chondroitin-4-sulfotransferase-1 (\textit{C4ST-1}), encodes an enzyme specific for the transfer of sulfate to the 4-O position in chondroitin, with chondroitin sulfate (CS) being an important component of cartilage proteoglycans. \textit{CHST11} has significantly higher expression in osteoarthritic compared to normal articular cartilage.\textsuperscript{71} Proper 4-O sulfation of chondroitin by \textit{CHST11} plays a crucial role in skeletal development and signalling events and in human disease, including cancer (reviewed in Kluppel\textsuperscript{72}). Mice homozygous for a mutation in \textit{CHST11} die shortly after birth with severe chondrodysplasia, growth plate defects, and accelerated chondrocyte differentiation. The reduction in CS\( 4 \)S in mutant mice leads to changes in the spatial distribution of CS and altered patterns of TGF-\( \beta \) and bone morphogenetic protein signalling in the cartilage growth plate.\textsuperscript{73}
TP63
rs12107036 in intron 12 of TP63 (tumour protein p63) was associated with total knee replacement (TKR) in females with borderline genome-wide significance (p=6.71×10^{-8}, OR=1.21, 95% CI 1.13 to 1.29) (table 1). The role of p63 mutations in cancer is well established (reviewed in Muller and Vousden20). Recently a polymorphism in this gene has been robustly associated with facial morphology in Europeans.75 p63 null mice have major defects in their limb, craniofacial and epidermal development implying a role for this gene in skeletal function.76 77

FTO
rs8044769 within intron 1 of FTO was most highly associated with OA in females (p=6.85×10^{-8}, OR=1.11, 95% CI 1.07 to 1.15) (table 1).20 Variation in the FTO (fat mass and obesity associated) gene is known to play an important role in susceptibility to obesity,78 and rs8044769 is in partial LD (r2>0.6) with the reported most highly associated SNP for body mass index (BMI). Overweight/obesity is a well established risk factor for OA susceptibility and it is also a predictor for OA progression, especially of the knee joint and less of the hip joint.8 79 The signal was attenuated after BMI adjustment, suggesting that the FTO gene exerts its effect on OA through obesity.20

SUPT3H; CDC5L
The only signal from the arcOGEN GWAS that exhibited the highest association in the male OA stratum was from rs10948172 (p=7.92×10^{-8}, OR=1.14, 95% CI 1.09 to 1.20) situated between the CDC5L (CDC5 cell division cycle 5-like) and SUPT3H (suppressor of Ty3 homolog) genes with unclear roles in OA (table 1).20 However ~500 kb away but in the same LD block is the RUNX2 (runt related transcription factor 2) gene which codes for a multifunctional transcription factor essential for osteoblast development and proper bone formation.80 RUNX2 controls skeletal development by regulating the differentiation of chondrocytes and osteoblasts and the expression of many extracellular matrix protein genes during this process (reviewed in Komori81). Consistent with its role as a master organiser, alterations in RUNX2 expression levels have been associated with skeletal diseases in human and mice.82 83 RUNX2 has been suggested as a possible biomarker of bone metabolism in several forms of arthritis.84

NCOA3
The largest GWAS meta-analysis for OA to date by the TREAT-OA consortium (in 11 277 hip OA cases and 67 473 controls including follow-up studies) established an additional variant, rs6094710, located near NCOA3 with p=7.9×10^{-9} and OR=1.28 (95% CI 1.18 to 1.39) (table 1).22 This gene is expressed in articular cartilage and its expression is significantly reduced in OA affected cartilage compared to preserved cartilage from the same joint.22

The molecular mechanism by which NCOA3 could cause OA is rather unclear. NCOA3 is involved in the co-activation of different nuclear receptors, such as for steroids (GR and ER), retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR) and prostanoids (PPARs). Several of these hormones have been implicated in skeletal metabolism and osteoarthritis. NCOA3 knockout mice exhibit growth retardation and reduced adult body size, but the molecular mechanism responsible for this growth retardation remains largely unknown. In female mice the reproductive system showed abnormal development and oestrogen values were significantly lower than in the wild type indicating a putative role of NCOA3 in steroid regulation.85 Alternative hypotheses for a causal effect of the NCOA3 gene in OA are through regulation of the target tissue responses to thyroid hormone T3 or through transcriptional regulation in mechanotransduction.22

DOT1L
Prompted by the strong association of a variant in the DOT1L gene with minimum joint space width (minJSW) at the hip,86 the TreatOA consortium and other European studies performed recently a large scale meta-analysis across seven OA scans to empower the examination of the association of this variant with hip OA. In male subjects the C allele of DOT1L rs12982744 was found significantly associated with OA (p=7.8×10^{-9}, OR=1.17, 95% CI 1.11 to 1.23) whereas for both genders combined the p value was 8.1×10^{-8} (table 1).21 Interestingly, as for the GDF5 polymorphism, the same DOT1L variant associated with OA has also been associated with height,87 suggesting a role in skeletal formation.

rs12982744 is in the first intron of DOT1L (DOT1-like, histone H3 methyltransferase) gene which encodes an essential and dedicated enzyme for Wnt target gene activation in the intestine and is required for the expression of genes that require high levels of Wnt signalling in drosophila.88 89 Wnt signalling is critical in the formation of cartilage and bone and in the development of the synovial joint. In vitro functional analyses demonstrated a role for DOT1L in chondrogenesis and the protein was found to interact directly with transcription factor 4—a transcription factor interacting with β-catennin—suggesting a role for this gene in the Wnt/β-catennin signalling cascade in developing chondrocytes.86

THE IMPORTANCE OF PHENOTYPE DEFINITION IN GENETIC STUDIES OF OA

Studies in other musculoskeletal traits have demonstrated the increase in power that can be afforded by investigating quantitative traits closer to the underlying biological phenotype. For example, in the largest GWAS for osteoporosis to date 56 loci associated with bone mineral density at genome-wide significance compared to six loci associated with the hard clinical outcome of fracture.90 In contrast to several diseases that can be considered as the extreme of the distribution of a physiological trait, OA is a highly heterogeneous disease affecting the entire joint and is manifested at different or several joint sites (generalised OA). OA is characterised by variable clinical features with possibly different genetic aetiologies. Currently OA ascertainment is based on either radiographically derived or symptomatic criteria, or a combination of these. Radiographic definition considers only pathophysiological joint features scored from radiographs. Symptomatic definition considers OA cases when both radiographic and pathological symptoms such as pain, stiffness, and loss of function at joints are present. For radiographic OA (ROA) several scoring systems exist, but the most widely used is the Kellgren-Lawrence (KL) grading system with grade 2 (definitive small osteophytes and little/mild joint space narrowing) or over being classified as ROA. However, an investigation by the TreatOA consortium on phenotype standardisation noted that among major cohort studies KL scores are interpreted differently, especially for the knee and hip.91 The great variability of disease definition among different studies presents an extra source of phenotype heterogeneity.

There are several arguments for and against using a more homogeneous phenotype for OA by examining cases ascertained
by TJR only. Pain and disability among subjects undergoing TJR are often poorly correlated with radiographic severity and TJR candidates show considerable heterogeneity in these symptomatic and radiographic features of OA.\textsuperscript{92,93} On the other hand, TJR definition for OA has been proposed for randomised clinical trials (RCTs) as it is the main clinical outcome that is representative of severe symptomatic large joint OA.\textsuperscript{94} The familial concordance for hip and knee OA is greater in surgically defined than in radiographically defined disease.\textsuperscript{95–97} In the arcOGEN study the authors were able to investigate the effect of OA phenotype definition on the strength of association of the eight established signals by comparing the results of meta-analyses employing TJR only cases as opposed to studying all cases (TJR and ROA combined). Four signals (rs69796, rs4836732, rs835487, rs12107036) showed stronger evidence for association in the TJR meta-analysis compared to the analysis of all cases, despite the considerable decrease in sample size and number of studies in the discovery and replication sets. Only one signal (rs9350591) was stronger in the meta-analysis employing all cases and three signals (rs10492367, rs8044769, rs10948172) remained relatively unchanged.\textsuperscript{20}

Pain, the most common and discomforting symptom for OA, is also an important phenotype to study, but the limited studies for OA related pain to date have not been able to robustly detect any underlying genetic variants (reviewed in Van Meurs and Uitterlinden\textsuperscript{98}). As with other traits (eg, hypertension and blood pressure measurements\textsuperscript{99,100}) it is anticipated that the examination of underlying, intermediate traits that together synthesise the phenotype of OA but are closer to the biology of the disease could be very advantageous in such a heterogeneous disorder. An example of this approach is the implication of the DOT1L gene in hip OA pathogenesis. The DOT1L locus was first discovered significantly associated with cartilage thickness, as measured by joint space width on radiographs, in a relatively small number of subjects, but did not reach genome-wide significance in a well sized case–control analysis for hip.\textsuperscript{86} Subsequently and upon additional follow-up, large scale, replication efforts in several hip OA case–control datasets, the association of DOT1L with OA was eventually established with genome-wide significance.\textsuperscript{21}

There is some evidence that genetic factors influence joint morphology, specific anatomic pattern of joint involvement, severity, and bone responses in OA at the hip and knee, and so these could represent promising endophenotypes to be studied. For example, morphological features such as the pistol grip deformity (PGD), femoral neck shaft angle (FNSA), the alpha angle and the lateral centre edge (LCE) angle have been associated with hip OA and may be under genetic control.\textsuperscript{101–103} Bone responses to hip OA may be classified as atrophic, normotrophic or hypertrophic, with atrophic OA being a more progressive form of OA than hypertrophic OA.\textsuperscript{104} In the only genetic study for bone response to OA, the risk for definite hip OA among siblings was twofold higher in siblings of index participants who had an atrophic pattern of disease than in siblings whose index case had any degree of osteophyte.\textsuperscript{105}

### SITE AND SEX SPECIFIC DIFFERENCES AT OA LOCI

There is compelling evidence that there are joint specific genetic factors contributing to OA aetiology,\textsuperscript{106} consistent with the significant differences reported in OA prevalence between different skeletal sites.\textsuperscript{10} It is thus not surprising that most of the GWAS for OA conducted thus far have stratified cases according to OA manifested either at the hip or at the knee joint and have identified site specific associations.

Sex differences have also been reported in the prevalence of OA,\textsuperscript{107} with female sex being an important risk factor for OA. Epidemiological studies have suggested that oestrogen loss may be accompanied by an increase in the prevalence and incidence of knee and hip OA in females,\textsuperscript{108} which may partly explain the sex differences in the prevalence of OA. Genetic studies that have stratified for sex have identified some clear differences.

The most compelling example is the significantly different (p=0.003) effect size estimate between both sexes at the DOT1L polymorphism (OR=1.17, 95% CI 1.11 to 1.23, p=7.8×10⁻⁷ in males vs OR=1.05, 95% CI 1.00 to 1.10, p=0.04 in females), and some of the loci identified by the arcOGEN study (ASTN2, TP63, and FTO significantly associated with OA in females and SUPT3H, CDC5L locus in males).\textsuperscript{20,21}

### TRANSLATIONAL POTENTIAL OF CURRENT FINDINGS

It is universally accepted that in characterising the genetic aetiology of common multifactorial diseases that can be ascribed to common variation, the GWAS approach has been very fruitful. However, because of the modest and small effect sizes exerted by the majority of common variants the translational potential of GWAS findings has been extensively criticised. Small effect sizes, however, should not undermine the biological importance of the genetically implicated genes.

An excellent example is the recent implication of CHST11 in OA. CHST11 codes for an enzyme responsible for the formation of CS, an important cartilage proteoglycan, with proteoglycan modulation being a currently active area of OA therapeutic development. CS is used as a symptomatic, slow acting drug for OA recommended by the latest OA Research Society International (OARSI) treatment guidelines but, despite extensive trials, evidence for its effectiveness remains controversial.\textsuperscript{109} The implication of CHST11 as a risk locus for OA suggests that alternate therapeutic approaches targeting the same pathway may be clinically beneficial.

The implication of the PITHLH locus in risk of OA may pave the way for exploring recently developed novel anabolic treatments for osteoporosis (peptide fragments based on parathyroid hormone) in the management of OA.

The genetic association with FTO confirms existing epidemiological evidence of the interplay between obesity and OA and highlights existing clinical recommendations that weight loss regimens may offer symptom relief and avoidance.

Furthermore, the biological insights afforded by the novel robust associations represent the largest, though indirect, translational contribution of these GWAS findings to OA.

### FUTURE STUDIES IN OA

In line with other common complex diseases, all OA associated variants thus far collectively explain only a small fraction (<10%) of the genetic component. There are possibly several more common variants to be discovered for OA through larger scale meta-analytical efforts,\textsuperscript{23} but also low frequency and rare variants, structural variants, gene–environment interactions, and epigenetic changes are likely to contribute substantially towards this missing heritability.\textsuperscript{110}

As the new era of next generation sequencing (NGS) association studies is emerging, the field of complex disease genetics is now focusing on the contribution of low minor allele frequency (MAF 1–5%) and rare variants (MAF <1%). Such variants may have larger effect sizes, higher penetrance, and point to causal
genes or functional units (eg, regulatory regions) more readily. Studies of rare variation in OA are currently underway.

The study of less heterogeneous, narrower OA endophenotypes closer to the biology of the disease is likely to lead to many more common and low frequency/rare variants underpinning specific and clinically relevant processes of disease development and progression. In addition, large scale studies investigating interactions between genetic and environmental risk factors can conceivably help shape approaches of disease management. Ultimately coupling all these genetic variants to function through functional studies and by integration with data generated from transcriptomics—the study of gene expression—and epigenomics—the study of epigenetic modifications such as DNA methylation, histone modifications, etc, in control of gene expression—will shape future genomics research in OA.

**SUMMARY**

Over the past few years GWAS in individuals of European and Asian ethnicity have collectively robustly identified 15 OA associated variants with genome-wide significance. All of the variants that have been detected thus far are common in frequency—which is by definition what GWAS are designed for—and appear to confer small to modest effect sizes. Fine mapping is required to identify which are the causal variants at the established loci. In addition, functional work is required to establish the causal genes, particularly for the loci that encompass many genes in regions of extended linkage disequilibrium. Despite these limitations the research that has been carried out thus far has provided insights into the biological processes that underlie OA susceptibility and has revealed some candidates with translational potential. The future outlook for OA genetics appears likely to be shaped by larger meta-analytical efforts to identify additional susceptibility loci, NGS approaches that can interrogate low frequency and rare variation, expanded and tighter OA phenotype definitions, and the integration of genetic variation studies with epigenetics and transcriptomics.

**Contributors**

KP and EZ contributed equally to this manuscript. **Competing interests**

EZ and KP are funded by the Wellcome Trust (098051). KP is funded by Arthritis Research UK (19542).

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Open Access**

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

---

**REFERENCES**


polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. Ann Rheum Dis [Published Online First 16 Mar 2013].


