

The genetics and functional analysis of primary osteoarthritis susceptibility

Louise N. Reynard and John Loughlin*

Recent genome-wide association scans (GWASs) along with several adequately powered candidate gene studies have yielded a number of risk alleles for osteoarthritis (OA). This number is now sufficiently large to allow conclusions to be drawn regarding the nature of genetic susceptibility, including the fact that the risk alleles have variable effects depending on sex, ethnicity and on the skeletal site of the disease. Several of the alleles that have emerged from the GWASs are within or close to highly plausible candidate genes, including *RUNX2* and *CHST11*. However, the majority of risk alleles do not map to genes previously reported to play a role in musculoskeletal biology, indicating that the GWAS datasets are telling us something new about the OA disease process. Functional studies have so far revealed that effects on gene expression are likely to be one of the main mechanisms through which OA susceptibility is acting. Epigenetic mechanisms such as DNA methylation also influence OA risk, and integration of genetic, transcriptomic and epigenetic data will allow us to use the genetic discoveries for informed development of new OA biological treatments.

Osteoarthritis (OA) is a late-onset musculoskeletal disease characterised by gradual thinning and loss of articular cartilage of the synovial joints with a concurrent alteration in the physiology of several other joint tissues, including the subchondral bone and the synovium (Ref. 1). The disease is therefore one of the whole joint (Ref. 2). No OA disease modifying drugs are available, with the current treatment regimes principally involving analgesia, physiotherapy and, in severe forms of the disease, joint replacement (Ref. 3); a search of the National Joint Registry reveals that over 120 000 hip and knee procedures are performed each year in the UK, the majority for OA (<http://www.njrcentre.org.uk>).

In most ethnic groups OA is extremely common, with its prevalence and incidence varying depending on the diagnostic criteria used and on the joint examined, with disease of the hands and of the knees being particularly prevalent (Ref. 4). OA has a clear and detrimental impact on wellbeing, with up to one-fifth of affected individuals giving up work or retiring early because of the disease, and this increased morbidity contributes indirectly to an increased mortality (Ref. 5).

Before any molecular genetic investigations were performed, a number of epidemiological studies had provided compelling evidence for a genetic component to OA, placing it into the

Newcastle University, Institute of Cellular Medicine, Newcastle upon Tyne, UK

*Corresponding author: John Loughlin, Newcastle University, Institute of Cellular Medicine, 4th Floor Catherine Cookson Building, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. E-mail: john.loughlin@ncl.ac.uk

category of a polygenic and multifactorial disease (reviewed in Ref. 6). Subsequent genetic linkage studies of small- to medium-sized pedigrees failed to detect any genes harbouring OA risk alleles, although broad regions of the genome were implicated. Successes soon followed, however, when genetic linkage was superseded by association analysis, particularly when the association studies were conducted on large case-control cohorts and with dense maps of polymorphic markers. These studies have been complemented by genome-wide gene expression studies of joint tissues such that we are now in the position whereby molecular genetics is making a contribution towards our understanding of the pathophysiology of this common disease.

In this review we shall focus on some of the most recent and compelling OA genetic results and on the steps taken so far to understand what functional effect OA risk alleles have on gene and protein function.

Genetic association analysis and candidate genes

In the last few years a number of candidate genes have been reported as harbouring risk alleles for OA. These candidates have been chosen principally on the basis that the proteins that they code for are regulators of joint formation and of joint maintenance. Table 1 lists examples of some of the genes studied. What has become apparent from these reports is that OA risk alleles often show ethnic stratification, with association in Asians or in Europeans but not typically in both. This can be attributed to differences between ethnic groups in (1) the frequencies of the risk alleles, (2) the genetic background on which the risk alleles are operating and (3) nongenetic (environmental) factors that modulate the impact of the risk alleles. It is also apparent that the OA associations are often to a particular joint rather than to several different joints. In the eight examples listed in Table 1, five associations are to either the hip or to the knee whereas only three are to both joints. This is an important observation as it suggests that there is only a limited sharing of the effects of particular OA risk alleles between skeletal sites. From a genetics perspective, OA is not therefore a systemic disease of the whole skeleton but is instead a more site-specific disease. Stratification by joint is therefore essential to avoid false negative associations.

Of the genes listed in Table 1 *GDF5* and *SMAD3* deserve further comment. *GDF5* is the most compelling candidate association signal so far reported for OA, with the rs143383 single nucleotide polymorphism (SNP) showing association in both Europeans and Asians and at a significance level of $P < 5.0 \times 10^{-8}$, a threshold that is commonly applied to assess the veracity of genome-wide association signals (Refs 7, 8, 9). The gene codes for a growth factor that is part of the transforming growth factor-beta (TGF- β) superfamily and the associated polymorphism, SNP rs143383, have also been reported to be associated with a variety of other musculoskeletal phenotypes (reviewed in Ref. 19). Polymorphism in this gene appears therefore to have a broad impact on skeletal health. The *SMAD3* association is with the common SNP rs12901499 located in intron 1. *SMAD3* codes for an intracellular signal transducer that also operates in the TGF- β pathway and since the association of rs12901499 to OA has so far been reported in only one publication (Ref. 10), further studies are needed to support the role of this SNP in OA susceptibility. Nevertheless, it is intriguing that rare and penetrant haploinsufficient mutations within *SMAD3* give rise to the aneurysms-OA syndrome, characterised by early-onset OA and aortic lesions (Ref. 20). It therefore appears that SNPs such as rs12901499, which like most common polymorphisms are expected to have relatively moderate effects on gene function, and rare but highly detrimental mutations are acting on *SMAD3* to give rise to the same broad OA phenotype. However, these arise at widely different ages and with differing degrees of severity. This is so far the only clear example of this phenomenon in OA. Since individuals with the *SMAD3* haploinsufficient mutations have aortic abnormalities there is the possibility that carriers of the OA risk allele of rs12901499 may be at increased risk of vascular deformities. Research investigating the cardiovascular system in these carriers may therefore be insightful.

Genetic association analysis and genome-wide association scans

Candidate gene studies have the advantage of targeting the known, combined with the attraction that the investigating team has only to genotype a relatively small number of polymorphisms to cover the gene of interest. These studies, however, have the clear disadvantage that what is known is

Table 1. Examples of genes reported as associated with osteoarthritis (OA) through candidate gene studies

Gene	Chromosome	Associated polymorphism	Protein	Protein function	Joint	Ethnic group	Refs
<i>GDF5</i>	20q11.22	rs143383	Growth differentiation factor 5	Regulator of cell growth and differentiation	Knee and hip	Asians and Europeans	7–9
<i>SMAD3</i>	15q22.33	rs12901499	Mothers against decapentaplegic homologue 3	Signal transducer and transcriptional modulator	Knee and hip	Europeans No Asian reports yet	10
<i>ASPN</i>	9q22.31	Triplet repeat of the codon for aspartic acid	Asporin	Cartilage extracellular protein, member of the small leucine-rich proteoglycan family	Knee	Asians Not Europeans	11
<i>EDG2/LPAR1</i>	9q31.3	rs10980705	Lysophosphatidic acid receptor 1	Lysophosphatidic acid receptor	Knee	Asians Not Europeans	12,13
<i>CALM1</i>	14q24-q31	rs12885713	Calmodulin	Subunit of phosphorylase kinase	Hip	Asians Not Europeans	14,15
<i>DIO2</i>	14q31.1	rs225014 and rs12885300	Type II iodothyronine deiodinase	Catalyses the activation of thyroid hormone	Hip	Europeans and Asians	16
<i>DIO3</i>	14q32.31	rs945006	Type III iodothyronine deiodinase	Catalyses the inactivation of thyroid hormone	Knee and hip	Europeans No Asian reports yet	17
<i>SREBF2</i>	22q13.2	rs2228314	Sterol regulatory element-binding protein 2	Transcription factor controlling cholesterol homeostasis	Knee	Europeans No Asian reports yet	18

The genetics and functional analysis of primary osteoarthritis susceptibility

Table 2. Genome-wide association scans performed utilising high genome coverage and large cohorts

Study	Ethnicity of discovery cohort and cohort size	Ethnicity of replication cohort and overall cohort size for discovery and replication combined	Outcome	Refs
Rotterdam	European – Netherlands 1341 cases and 3496 controls	European and North American of European descent 14 938 cases and 39 000 controls	One signal, at chromosome 7q22	21
Tokyo	Asian – Japanese 899 cases and 3396 controls	Asian and European 1879 cases and 3396 controls	One signal, at the HLA locus	22
arcOGEN	European – UK 3177 cases and 4894 controls	European and North American of European descent 13 768 cases and 53 286 controls	Three signals	23
	European – UK 7410 cases and 11 009 controls	European 14 883 cases and 53 947 controls	Eight signals	24

rarely comprehensive and that novel signals will not therefore be uncovered. The genome-wide association scan (GWAS) circumvents these deficiencies and, when performed on large cohorts and with replication, has the power to overcome the issue of multiple testing. In OA there have so far been three reports of GWAS studies combining the key requirements of extensive coverage, large cohorts and replication in additional cohorts (Table 2).

The Rotterdam study reported a single association signal for knee OA with a P -value of 8×10^{-8} to a region of chromosome 7q22 encompassing six known protein-coding genes (Ref. 21, Table 3), none of which had previously been implicated in OA. A subsequent analysis with additional cohorts marginally increased the significance of the association (Ref. 25). The Tokyo study reported an association signal to two SNPs located within a 340 kb region of the HLA locus on chromosome 6p, with P -values $< 7 \times 10^{-8}$ (Ref. 22). The association is also with knee OA. The arcOGEN study was performed in two stages. Initially, a GWAS was performed on 3177 cases and three association signals were reported, but none surpassed the genome-wide significance threshold of $P \leq 5.0 \times 10^{-8}$ (Ref. 23). The signals were to *MICAL3* in knee and/or hip OA with a P -value of 2.3×10^{-5} , to *C6orf130* in

knee OA with a P -value of 2.7×10^{-5} and to *COL11A1* in hip OA with a P -value of 1.2×10^{-5} . The number of cases genotyped was then increased to 7410 and the GWAS was re-run. This enhanced power identified eight novel signals, five of which were genome-wide significant (Ref. 24, Table 3). Five of the arcOGEN signals encompass two or more protein-coding genes with the remaining three signals containing single genes. Several of the genes are plausible functional candidates, including *RUNX2*, which codes for a transcription factor active in joint tissues, and *CHST11*, which codes for an enzyme that sulfates cartilage proteoglycan. However, the majority of the genes have not previously been suggested to have a role in OA meaning that the arcOGEN study has provided very novel insights into OA aetiology. Body mass index (BMI) is a known risk factor for OA but only one of the arcOGEN signals, to *FTO*, was attenuated after BMI adjustment; polymorphism in *FTO* is one of the strongest associations so far reported for BMI and obesity risk.

One key observation from the arcOGEN study was that stratification by sex, by joint and by the OA ascertainment criteria used was critical in the discovery of the association signals. Regarding sex, the chromosome 6p21.1 signal was only relevant to male disease whereas the 3q28, 9q33.1

Table 3. The signals from the Rotterdam genome-wide association scan (GWAS) and from the full arcOGEN GWAS

Study	SNP	Chromosome	P-value	Stratum	Gene	Protein name	Summary of protein function
Rotterdam	rs3815148	7q22	8.0×10^{-8}	Knee	<i>PRKAR2B</i>	protein kinase-cAMP-dependent-regulatory-type II- β	Regulator of cAMP signalling
					<i>HPB1</i>	HMG-box transcription factor 1	Transcription factor
					<i>COG5</i>	component of oligomeric golgi complex 5	Golgi morphology and function
					<i>GPR22</i>	G protein-coupled receptor 22	Receptor and membrane protein
					<i>DUS4L</i>	dihydrouridine synthase 4-like	Unknown
					<i>BCAP29</i>	B-cell receptor-associated protein 29	Unknown
arcOGEN	rs6976	3p21.1	7.24×10^{-11}	TJR	<i>STAB1</i>	Stabilin-1	Transmembrane receptor
					<i>NT5DC2</i>	5'-nucleotidase domain-containing protein 2	Unknown
					<i>PBRM1</i>	Polybromo-1	Chromatin remodeling and transcriptional activation
					<i>GNL3</i>	Guanine nucleotide-binding protein-like 3	Nuclear protein important for stem cell proliferation
					<i>GLT8D1</i>	Glycosyltransferase 8 domain-containing protein 1	Glycosyltransferase
					<i>SPCS1</i>	Signal peptidase complex subunit 1	Unknown

(continued on next page)

The genetics and functional analysis of primary osteoarthritis susceptibility

Table 3. The signals from the Rotterdam genome-wide association scan (GWAS) and from the full arcOGEN GWAS (continued)

Study	SNP	Chromosome	P-value	Stratum	Gene	Protein name	Summary of protein function
					NEK4	Serine/threonine-protein kinase Nek4	Kinase
					ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	Serine protease inhibitor – plasma
					ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	Serine protease inhibitor – plasma
					ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	Serine protease inhibitor – plasma
					MUSTN1	Musculoskeletal embryonic nuclear protein 1	Nuclear protein – expressed during skeletogenesis
					TMEM110	Transmembrane protein 110	Unknown
arcOGEN	rs12107036	3q28	6.71×10^{-8}	TKR-F	TP63	Tumor protein 63	Transcription factor
arcOGEN	rs10948172	6p21.1	7.92×10^{-8}	Male	SUPT3H	Transcription initiation protein SPT3 homologue	Unknown
					RUNX2	Runt-related transcription factor 2	Transcription factor
arcOGEN	rs9350591	6q13-q14.1	2.42×10^{-9}	Hip	COL12A1	Collagen α -1(XII) chain	FACIT collagen, interacts with type I collagen
					COX7A2	Cytochrome c oxidase subunit 7A2, mitochondrial	One of the polypeptides of cytochrome C oxidase
					TMEM30A	Cell cycle control protein 50A	Transmembrane protein, transporter

(continued on next page)

The genetics and functional analysis of primary osteoarthritis susceptibility

Table 3. The signals from the Rotterdam genome-wide association scan (GWAS) and from the full arcOGEN GWAS (continued)

Study	SNP	Chromosome	P-value	Stratum	Gene	Protein name	Summary of protein function
arcOGEN	rs4836732	9q33.1	6.11×10^{-10}	THR-F	<i>PAPPA</i>	Pappalysin-1	Secreted metalloproteinase
arcOGEN	rs10492367	12p11.22	1.48×10^{-8}	Hip	<i>KLHDC5</i>	Kelch domain-containing protein 5	Protein ubiquitination in innate immunity
arcOGEN	rs835487	12q23.3	1.64×10^{-8}	THR	<i>CHST11</i>	Carbohydrate sulfotransferase 11	Protein ubiquitination
arcOGEN	rs8044769	16q12.2	6.85×10^{-8}	Female	<i>FTO</i>	α -ketoglutarate-dependent dioxygenase FTO	Hormone
					<i>FILIP1</i>	Filamin-A-interacting protein 1	Interacts with filamin – filamins bind to actin
					<i>SENP6</i>	Sentrin-specific protease 6	Ubiquitin-like molecule ligated to target proteins
					<i>MYO6</i>	Myosin-VI	Intracellular vesicle and organelle transport
					<i>IMPG1</i>	Interphotoreceptor matrix proteoglycan 1	Chondroitin sulfate proteoglycan expressed in the retina
					<i>ASTN2</i>	Astrotactin-2	Neural migration
					<i>TRIM32</i>	E3 ubiquitin-protein ligase TRIM32	Protein ubiquitination in innate immunity
					<i>PTHLH</i>	Parathyroid hormone-related protein	Hormone

TJR, total joint replacement; TKR-F, total knee replacement in females; THR-F, total hip replacement in females; THR, total hip replacement.

The genetics and functional analysis of primary osteoarthritis susceptibility

and 16q12.2 signals were only relevant to female disease. Regarding joint, the 3q28 signal was only relevant to knee OA whereas the 6q13-q14.1, 9q33.1, 12p11.22 and 12q23.3 signals were only relevant to hip OA. Regarding ascertainment criteria, all the arcOGEN cases had radiographic evidence of OA but in over 80% of the cases the disease was so severe that the individuals had also undergone hip or knee joint replacement surgery. A focus on these more severe cases aided the identification of several of the signals.

The minor allele frequencies (MAFs) of the SNPs that marked the arcOGEN association signals were quite common, all >10%. An analysis of the initial arcOGEN GWAS for less common alleles identified another association, with SNP rs11842874 (Ref. 26). This SNP has a MAF of 7% and resides within an intron of *MCF2L*, which codes for the protein guanine nucleotide exchange factor DBS. The association surpasses the genome-wide significance threshold, with a *P*-value of 2.1×10^{-8} .

The large arcOGEN dataset has so far been used in two related studies, the first investigating an association between DNA variants of the mitochondrial genome and OA and the second investigating an overlap between alleles for height, BMI and OA. For many years dysfunction of mitochondria has been implicated in OA aetiology and there is some evidence that polymorphism of mitochondrial DNA may mediate this, although the genetic studies so far performed have been on relatively small cohorts (reviewed in Ref. 27). The array used by arcOGEN provided robust genotype data for 48 common mitochondrial variants, which provided good coverage of the mitochondrial genome. However, no association with OA was observed in the comparison of 7393 arcOGEN cases and 5122 controls (Ref. 28). It appears unlikely therefore, that common mitochondrial DNA variants are major contributors to OA susceptibility, at least not in the north European population that is represented by the arcOGEN study. It has been postulated that there may be overlap between the development of OA and natural variation in skeletogenesis, based on the assumption that efficient developmental patterning of the skeleton is essential for creation and maintenance of healthy joints. The discovery that *GDF5* SNP rs143383 is one of the many polymorphisms associated with natural variation in height

(Ref. 29) clearly links skeletogenesis and OA. Furthermore, the subsequent discovery that due to its correlation with height, the OA-associated T-allele of this SNP is subjected to positive selection implies that OA genetic aetiology may partly be hitchhiking onto the evolution events of skeletogenesis (Ref. 30). However, although there was some initial suggestion from the arcOGEN study of an excess overlap of association signals between OA and height and between OA and BMI, only the BMI signal at *FTO* replicated (Ref. 31). This suggests that, apart from a few clear examples, there is not yet any compelling evidence of a major correlation between the alleles for height and OA or between the alleles for BMI and OA.

Functional studies on *GDF5* SNP rs143383

Having identified compelling genetic signals, the next step is the functional analyses of these to assess how the associated alleles modulate gene or protein function. In this regard, studies on OA-associated SNP rs143383 act as an exemplar.

rs143383 is a C to T transition located within the 5' untranslated region (5'UTR) of *GDF5*, which codes for the extracellular signalling molecule growth differentiation factor 5. In the paper that first reported the association between rs143383 and OA, the investigators demonstrated that the OA-risk T-allele of the SNP-mediated reduced mRNA expression relative to the C-allele in an *in vitro* luciferase assay conducted in a chondrocyte cell line; chondrocytes are the only cell type present in cartilage (Ref. 32). This experiment therefore highlighted rs143383 as the actual functional SNP responsible for the association signal. *GDF5* protein is an anabolic extracellular signalling molecule required for joint formation and maintenance and this functional study implied that OA susceptibility at rs143383 acts by reducing the levels of *GDF5* protein. Soon after this *in vitro* study, it was demonstrated that the T-allele correlated with reduced expression of *GDF5* in cartilage (Ref. 33). This result was obtained by an analysis of the expression of the gene using RNA directly extracted from the cartilage of OA patients, who had undergone elective joint replacement surgery. The patients studied were heterozygous for rs143383, allowing a direct comparison of the expression of the T-allele versus the C-allele. The allelic expression imbalance (AEI) between the C- and T-alleles of rs143383 also occurs in

Table 4. Functional studies performed on *GDF5* SNP rs143383

Functional study	Outcome	Refs
Luciferase assay in chondrocytic cell line HCS-2/8	T-allele shows reduced expression relative to the C-allele	32
<i>GDF5</i> allelic expression analysis (AEI) on cartilage RNA from OA patients	T-allele shows reduced expression relative to the C-allele	33
<i>GDF5</i> AEI on RNA extracted from other joint tissues	T-allele shows reduced expression relative to the C-allele	34
Investigation of additional <i>GDF5</i> 5'UTR SNPs	rs143383's effect on gene expression is modulated by rs143384	34
Investigation of <i>GDF5</i> 3'UTR SNPs	rs56366915 independently influences gene expression	34
EMSA, ChIP and RNA knockdown in liposarcoma cell line SW872	Sp1, Sp3, P15 and DEAF-1 bind differentially to the rs143383 alleles	35
CpG DNA methylation analysis	Demethylation exacerbates the AEI mediated by rs143383	40

other joint tissues (Ref. 34), emphasising that OA is a disease of the whole joint. A second *GDF5* 5'UTR SNP, rs143384, was discovered that modulates the effect that rs143383 has on gene expression. This result revealed that the activity of an OA susceptibility allele could be context-specific. A search of the *GDF5* 3'UTR then identified rs56366915 as a SNP that also causes AEI, but independently of rs143383 (Ref. 34). The subsequent use of electrophoretic mobility shift assays (EMSAs), EMSA-supershifts, chromatin immunoprecipitation (ChIP) and RNA knockdown identified the trans-acting factors Sp1, Sp3, P15 and DEAF-1 as binding differentially to the T- and C-alleles of rs143383 and mediating the AEI (Ref. 35). A summary of these functional studies is listed in Table 4.

rs143383, rs143384 and rs56366915 are common SNPs. A sequence analysis of *GDF5* in a cohort of 992 OA cases and 944 controls did not detect any additional common variants, implying that there are no other *GDF5* polymorphisms that can, at the population level, influence OA susceptibility (Ref. 36). This sequence analysis did however detect six unique variants, one of which is located in the promoter of *GDF5*. A subsequent functional study demonstrated that this promoter variant is a binding site for the trans-acting factor YY1 and that the unique A-allele of the variant can neutralise the reduced expression

that is mediated by the OA-associated T-allele of rs143383 (Ref. 37). This result again demonstrates that the activity of a susceptibility allele is context-specific whereas YY1, such as Sp1, Sp3, P15 and DEAF-1, is a trans-acting factor that can modulate *GDF5* expression and which could therefore be exploited to potentially alleviate the OA risk coded for by this gene.

In their C-allele forms rs143383 and rs143384 create CpG sites. Such sites are potentially amenable to epigenetic regulation by DNA methylation. Since epigenetic mechanisms are known to play a role in chondrogenesis and OA pathogenesis (Refs 38, 39) it was hypothesised that rs143383 and rs143384 may be targets for epigenetic control and that was indeed shown to be the case (Ref. 40). Methylation of the *GDF5* 5'UTR was demonstrated to occur in cell lines and in joint tissues, with demethylation leading to increased *GDF5* expression. In a cell line heterozygous for rs143383, this demethylation was found to exacerbate the AEI mediated by rs143383. This result clearly demonstrated that the genetic effect that rs143383 has on *GDF5* expression is modifiable by nongenetic factors.

There are as yet no animal models directly targeting the *GDF5* SNPs discussed above. However, the brachypodism mouse has offered an opportunity to assess what effect a reduction in *GDF5* has on joint physiology. The brachypodism

mutation is a frame shift of *Gdf5* that causes a premature termination codon, resulting in a null allele. Homozygotes display a number of musculoskeletal defects whereas heterozygotes show no such developmental abnormalities but are at risk of developing an OA phenotype when challenged (Ref. 41). By being subjected to a reduction rather than a total loss of *GDF5* the brachypodism heterozygote mimics the effect of rs143383 in human OA and emphasises the need to maintain appropriate levels of *GDF5* to ensure healthy joint function.

Functional studies on other OA susceptibility loci

There have been published reports on the functional analysis of other OA loci, with particularly insightful data generated recently for the 7q22 signal from the Rotterdam study and for *DIO2*.

As mentioned earlier, the Rotterdam study reported an association signal to a region of chromosome 7q22 encompassing six known protein-coding genes (Table 3). The expression of these genes was assessed using cells extracted and cultured from the joint tissues of OA patients. Expression was also investigated in mouse joint tissues and in zebrafish embryos. The results revealed that all the genes showed a near universal expression pattern, apart from *GPR22* (Refs 21, 25). A subsequent study that examined the expression of the genes using RNA directly extracted from patient joint tissues discovered that genotype at the 7q22 association signal correlated with expression of *HBP1* in cartilage (Ref. 42). It would appear likely therefore that part, or all, of the 7q22 association signal is accounted for by a regulatory polymorphism that modulates *HBP1* expression.

DIO2 codes for iodothyronine-deiodinase enzyme type 2 (D2), a selenoprotein that converts intracellular inactive thyroid hormone to its active form. A common *DIO2* haplotype composed of the C-allele of SNP rs225014 and the C-allele of SNP rs12885300 is associated with OA in Europeans and Asians (Ref. 16, Table 1). An analysis of *DIO2* expression revealed that the gene was subject to AEI with the C-allele of rs225014 correlating with increased expression in cartilage (Ref. 43). Furthermore, immunohistochemistry revealed an increased amount of D2 positive cells in OA versus healthy cartilage. These results suggest that modulating the level of active thyroid

hormone in joint tissues, via the increased expression of *DIO2* and the subsequent availability of D2, is a contributing factor in OA aetiology. Methylation analysis of CpG sites within and close to *DIO2* in both OA diseased cartilage and intact cartilage have revealed that, as for *GDF5*, the gene is also subjected to epigenetic regulation related to disease status (Ref. 44).

GDF5, *HBP1* and *DIO2* are all examples of where the OA risk allele mediates its effect by altering gene expression. As more susceptibility loci are identified for common human diseases it is becoming apparent that the majority of associated alleles contribute to disease risk by influencing gene expression, typically by modulating rates of transcription or of transcript stability (Ref. 45). OA genetic susceptibility is following this trend and as such the initial functional test on all new OA signals should be an assessment of the effect that the risk allele has on gene expression.

Clinical implications and applications

One of the early translational aims of the genetic investigation of common complex diseases was that it would offer a predictive capacity to clinicians in their attempts to identify individuals at risk of disease initiation and progression. However, it has become apparent that the vast majority of disease risk alleles contribute individually only very modestly to heritability and as such there is limited current scope for their use as predictive tools (Ref. 46). Clearly, this scenario is likely to change as more risk alleles are identified (see below) but for the vast majority of diseases this is realistically a medium term goal at best. Alternatively, by identifying the principal pathways that are the recipients of genetic susceptibility then there is scope for exploiting these insights for new treatment development. As yet the number of OA risk loci is quite small but it does appear from the list of genes in Tables 1 and 3 that regulation of cell differentiation could be one of the key pathways for future exploitation.

Research in progress and outstanding research questions

One of the clear messages to have emerged from the GWASs performed on common diseases is that the case-control sample sizes used have to be large in order to detect what are small individual contributions from the risk alleles. Of the three GWASs listed in Table 2 only arcOGEN had adequate sample size to identify multiple

genome-wide significant signals. If additional novel OA signals are to be discovered, it is critical that genotyping at the genome-wide level continues on all available samples and with the concurrent use of meta-analyses. The number of OA cases will also need to be expanded to provide the power to account for susceptibility differences between ethnic groups, between the two sexes and between different joints; without case sample sizes in the tens of thousands the vast majority of OA susceptibility alleles will go undetected. It has also become evident that rare variants can also contribute to common disease risk (Refs 47, 48) and that these are often not captured by the current GWAS arrays. As such imputation, re-sequencing and exome and whole-genome analyses (Ref. 49) are likely to have a role in the discovery of rare risk alleles for OA.

OA genetic studies have quite sensibly focused on clinical forms of the disease but it is becoming apparent that breakthroughs can also be achieved by studying alternative but related phenotypes and a very clear example of this has recently been published. The Rotterdam group who discovered the 7q22 locus subsequently examined joint-space width (JSW) at the hip in their population cohorts (Ref. 50). JSW is a proxy for cartilage thickness, with a narrowing indicating thinner cartilage. A GWAS was performed and an association was detected with SNP rs12982744, with a P -value of 1.1×10^{-11} . The G-allele of the SNP was associated with an increase in JSW and a secondary analysis revealed that this allele also correlated with a reduced risk of hip OA. The P -value for the OA association was more modest than that for JSW, at 1.0×10^{-4} , but the principle was clearly established that the use of a proxy phenotype can further our understanding of OA. rs12982744 resides within *DOT1L*, which codes for the enzyme histone-lysine *N*-methyltransferase, H3 lysine-79 specific. As the name implies, this enzyme is a histone methyltransferase that methylates lysine-79 of histone H3. Using the mouse ATDC5 chondrogenesis model and micromass cultures the investigators demonstrated that knockdown of *Dot1l* leads to reduced proteoglycan and collagen content and that this may be mediated by modulation of the wnt signalling pathway. The investigators also demonstrated expression of the gene during mouse limb development.

Pain is a symptom of clinical OA and has been considered a logical therapeutic target for OA

treatment. However, the current clinical trial data on the use of targeted pain biologics have revealed that adverse events can outweigh any benefits of managing OA pain (Ref. 51). Nevertheless, there have been several reports testing for genetic association between known pain loci and OA pain, including to the genes *P2RX7*, *PACE4* and *TRPV1* (Refs 52, 53, 54). In comparison with the OA GWAS studies discussed above, the significant P -values obtained were relatively modest. Despite this, these studies do support a more comprehensive and powered analysis of the genetics of OA pain.

Gene expression microarray analyses provide an opportunity to assess which particular pathways are implicated in disease initiation and progression, or in the establishment of a particular cell type. As such they have enhanced our understanding of OA pathogenesis and Table 5 provides details of six such recent studies including a summary of their key findings (Refs 55, 56, 57, 58, 59, 60). One clear conclusion is that gene expression varies qualitatively and quantitatively between cartilage collected from different skeletal sites, such as the knees and the hips (Refs 57, 58). This is reminiscent of the genetic discovery discussed earlier of OA risk alleles rarely being systemic in their effects and often showing joint-specific effects. The microarrays have also highlighted genes that show significant up- or down-regulation in OA and several of the genes from within the arcOGEN signals are among these, including *COL12A1*, *MYO6*, *CHST11* and *PAPPA*. In the future, RNA sequencing (RNAseq) will supersede microarrays, as it offers a number of advantages including the ability to discover novel transcripts and alternative splice forms (Refs 61, 62). A natural progression from gene expression is proteomic analysis but this is only now receiving the detailed and comprehensive attention that it deserves. Investigators have, for example, characterised the protein profiles in the joint tissues and fluids of OA patients and then compared these with the profiles from healthy groups and from patients with a different joint disease, such as rheumatoid arthritis (Refs 63, 64). Cartilages from different anatomical sites have also been compared (Ref. 65). The aim has been to enlighten our understanding of the pathophysiology of the disease and to assist in the identity of biomarkers that may be used to

Table 5. Recent global gene-expression microarray studies relevant to human osteoarthritis (OA)

Joint or cell type	Key findings	Refs
Knee cartilage	Diseased and nondiseased knee cartilage from the same donors demonstrates significant differences in the expression of a range of genes, including <i>WISP1</i> and <i>IGFBP3</i> . Whether this reflects a causal role in OA or is instead a consequence of the disease was not determined	55
Trabecular bone from the iliac crest	The gene profile of iliac crest from OA patients is significantly different from the gene profile of iliac crest from younger non-OA patients and from older non-OA patients, particularly for genes coding for proteins regulating bone, implying a systemic bone phenotype related to OA	56
Knee cartilage	The gene profile of knee cartilage from patients with knee OA is significantly different from that of knee cartilage from patients without knee OA. Genes that showed differential expression included those coding for proteins involved in bone formation, in the synthesis of collagens, cytokines and growth factors	57
Hip cartilage	The gene profile of hip cartilage from patients with hip OA is significantly different to that of hip cartilage from patients without hip OA. Gene expression in OA hip cartilage demonstrates a number of pathway overlaps with that of OA knee cartilage but also a striking number of differences, implying major aetiological variability between OA at the two sites	58
Cultured fibroblasts from knee synovial tissue	Compared with fibroblasts from the synovia of healthy knees, gene expression in OA fibroblasts clustered tightly and has an expression profile significantly distinct from healthy and from rheumatoid arthritis fibroblasts	59
Cartilage and growth plate	A comparison of gene expression between the growth-plate and articular cartilage of healthy adolescent donors identified the wnt pathway and several downstream target genes as brakes on chondrocyte hypertrophy, which is a known key step in OA progression	60

assess disease status and disease progression (Ref. 66). It is clear from these studies that there are significant differences between the OA and the non-OA proteome profiles and that some of the protein differences are detectable in accessible fluids such as urine, which makes their use in future clinical trials of OA therapeutics a realistic possibility.

As noted earlier, epigenetics contributes to OA pathogenesis as a regulator of gene expression irrespective of genotype and as a modulator of genetic risk (Refs 38, 39, 40). Much more work, however, needs to be performed before we get even a vague understanding of the actual impact of epigenetics on the disease. Careful

decisions will have to be made between which tissues to analyse and although cartilage is an obvious priority tissue, it is clear that all joint tissues have to be considered. The developmental time points studied will also be critical since epigenetic events are likely to be temporally regulated. The type of epigenetic mark is also an issue, with DNA CpG methylation and microRNAs amenable to relatively high-throughput analyses whereas histone modifications currently are not.

Ultimately the goal of all such genetic, transcriptomic and epigenetic datasets will be their integration to identify the regulatory pathways and networks that are definitive of

and causal to the disease (Ref. 67). At that point we may be in a position to use our genetic insights for informed development of new biological treatments for OA.

Acknowledgements and funding

We are grateful to Arthritis Research UK, the National Institute for Health Research, the JGW Patterson Foundation, the Nuffield Foundation and the Dr William Harker Foundation for funding our group's research.

References

- 1 Brandt, K.D., Dieppe, P. and Radin, E.L. (2008) Etiopathogenesis of osteoarthritis. *Rheumatic Disease Clinics of North America* 34, 531-559
- 2 Loeser, R.F. et al. (2012) Osteoarthritis: a disease of the joint as an organ. *Arthritis and Rheumatism* 64, 1697-1707
- 3 Bennell, K.L., Hunter, D.J. and Hinman, R.S. (2012) Management of osteoarthritis of the knee. *British Medical Journal* 345, e4934
- 4 Pereira, D. et al. (2011) The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthritis and Cartilage* 19, 1270-1285
- 5 Nüesch, E. et al. (2011) All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. *British Medical Journal* 342, d1165
- 6 Loughlin, J. (2002) Genome studies and linkage in primary osteoarthritis. *Rheumatic Disease Clinics of North America* 28, 95-109
- 7 Chapman, K. et al. (2008) A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5'UTR of GDF5 with osteoarthritis susceptibility. *Human Molecular Genetics* 17, 1497-1504
- 8 Evangelou, E. et al. (2009) Large-scale analysis of association between *GDF5* and *FRZB* variants and osteoarthritis of the hip, knee and hand. *Arthritis and Rheumatism* 60, 1710-1721
- 9 Valdes, A.M. et al. (2011) The *GDF5* rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Annals of the Rheumatic Diseases* 70, 873-875
- 10 Valdes, A.M. et al. (2010) Genetic variation in the *SMAD3* gene is associated with hip and knee osteoarthritis. *Arthritis and Rheumatism* 62, 2347-2352
- 11 Nakamura, T. et al. (2007) Meta-analysis of association between the *ASPN* D-repeat and osteoarthritis. *Human Molecular Genetics* 16, 1676-1681
- 12 Mototani, H. et al. (2008) A functional SNP in *EDG2* increases susceptibility to knee osteoarthritis in Japanese. *Human Molecular Genetics* 17, 1790-1797
- 13 Dieguez-Gonzalez, R. et al. (2009) Testing the druggable endothelial differentiation gene 2 knee osteoarthritis genetic factor for replication in a wide range of sample collections. *Annals of the Rheumatic Diseases* 68, 1017-1021
- 14 Mototani, H. et al. (2005) A functional single nucleotide polymorphism in the core promoter region of *CALM1* is associated with hip osteoarthritis in Japanese. *Human Molecular Genetics* 14, 1009-1017
- 15 Loughlin, J. et al. (2006) The *CALM1* core promoter polymorphism is not associated with hip osteoarthritis in a United Kingdom Caucasian population. *Osteoarthritis and Cartilage* 14, 295-298
- 16 Meulenbelt, I. et al. (2008) Identification of *DIO2* as a new susceptibility locus for symptomatic osteoarthritis. *Human Molecular Genetics* 17, 1867-1875
- 17 Meulenbelt, I. et al. (2011) Meta-analysis of genes modulating intracellular T3 bio-availability reveal a possible role for the *DIO3* gene in osteoarthritis susceptibility. *Annals of the Rheumatic Diseases* 70, 164-167
- 18 Kostopoulou, F. et al. (2012) Central role of *SREBP-2* in the pathogenesis of osteoarthritis. *PLoS ONE* 7, e35753
- 19 Loughlin, J. (2011) Genetics of osteoarthritis. *Current Opinion in Rheumatology* 23, 479-483
- 20 van de Laar, I.M. et al. (2011) Mutations in *SMAD3* cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nature Genetics* 43, 121-126
- 21 Kerkhof, H.J. et al. (2010) A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis and Rheumatism* 62, 499-510
- 22 Nakajima, M. et al. (2010) New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS ONE* 5, e9723
- 23 Panoutsopoulou, K. et al. (2011) Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Annals of the Rheumatic Diseases* 70, 864-867
- 24 arcOGEN Consortium and arcOGEN collaborators. (2012) Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 380, 815-823

- 25 Evangelou, E. et al. (2011) Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Annals of the Rheumatic Diseases* 70, 349-355
- 26 Day-Williams, A.G. et al. (2011) A variant in MCF2L is associated with osteoarthritis. *American Journal of Human Genetics* 89, 446-450
- 27 Blanco, F.J., Rego, I., Ruiz-Romero, C. (2011) The role of mitochondria in osteoarthritis. *Nature Reviews. Rheumatology* 7, 161-169
- 28 Hudson, G. et al. (2013) No evidence of an association between mitochondrial DNA variants and osteoarthritis in 7393 cases and 5122 controls. *Annals of the Rheumatic Diseases* 72, 136-139
- 29 Sanna, S. et al. (2008) Common variants in the GDF5-UGCC region are associated with variation in human height. *Nature Genetics* 40, 198-203
- 30 Wu, D.D. et al. (2012) Positive selection on the osteoarthritis-risk and decreased-height associated variants at the GDF5 gene in east Asians. *PLoS ONE* 7, e42553
- 31 Elliott, K.S. et al. Evaluation of the genetic overlap between osteoarthritis with body mass index and height using genome-wide association scan data. *Annals of the Rheumatic Diseases* (in press) <http://www.ncbi.nlm.nih.gov/pubmed/22956599>
- 32 Miyamoto, Y. et al. (2007) A functional polymorphism in the 5'UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nature Genetics* 39, 529-533
- 33 Southam, L. et al. (2007) An SNP in the 5'UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Human Molecular Genetics* 16, 2226-2232
- 34 Egly, R. et al. (2009) Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis and Rheumatism* 60, 2055-2064
- 35 Syddall, C.M. et al. (2012) The identification of transacting factors that differentially regulate the expression of GDF5 via the osteoarthritis associated SNP rs143383. *Osteoarthritis and Cartilage* 20 (Suppl 1), S46
- 36 Dodd, A.W. et al. (2011) Deep sequencing of GDF5 reveals the absence of rare variants at this important osteoarthritis susceptibility locus. *Osteoarthritis and Cartilage* 19, 430-434
- 37 Dodd, A.W., Syddall, C.M. and Loughlin, J. A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *European Journal of Human Genetics* (in press) <http://www.ncbi.nlm.nih.gov/pubmed/22929025>
- 38 Barter, M.J., Bui, C. and Young, D.A. (2012) Epigenetic mechanisms in cartilage and osteoarthritis: DNA methylation, histone modifications and microRNAs. *Osteoarthritis and Cartilage* 20, 339-349
- 39 Swingler, T.E. et al. (2012) The expression and function of microRNAs in chondrogenesis and osteoarthritis. *Arthritis and Rheumatism* 64, 1909-1919
- 40 Reynard, L.N. et al. (2011) Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. *Human Molecular Genetics* 20, 3450-3460
- 41 Daans, M., Luyten, F.P. and Lories, R.J. (2011) GDF5 deficiency in mice is associated with instability-driven joint damage, gait and subchondral bone changes. *Annals of the Rheumatic Diseases* 70, 208-213
- 42 Raine, E.V. et al. (2012) Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Annals of the Rheumatic Diseases* 71, 2020-2027
- 43 Bos, S.D. et al. (2012) Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Annals of the Rheumatic Diseases* 71, 1254-1258
- 44 den Hollander, W. et al. (2012) CpG sites of osteoarthritis susceptibility gene DIO2 are differentially methylated in arthritic compared to preserved cartilage. *Osteoarthritis and Cartilage* 20 (Suppl 1), S196
- 45 Montgomery, S.B. and Dermitzakis, E.T. (2011) From expression QTLs to personalized transcriptomics. *Nature Reviews. Genetics* 12, 277-282
- 46 Rodriguez-Fontenla, C. et al. (2012) Genetic risk load and age at symptom onset of knee osteoarthritis. *Journal of Orthopaedic Research* 30, 905-909
- 47 Nelson, M.R. et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337, 100-104
- 48 Keinan, A. and Clark, A.G. (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336, 740-743
- 49 Tennessen, J.A. et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64-69
- 50 Castaño Betancourt, M.C. et al. (2012) Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proceedings of the National*

- Academy of Sciences of the United States of America 109, 8218-8223
- 51 Seidel, M.F. and Lane, N.E. (2012) Control of arthritis pain with anti-nerve-growth factor: risk and benefit. *Current Rheumatology Reports* 14, 583-588
- 52 Sorge, R.E. et al. (2012) Genetically determined P2X7 receptor pore formation regulates variability in chronic pain sensitivity. *Nature Medicine* 18, 595-599
- 53 Malfait, A.M. et al. (2012) A role for PACE4 in osteoarthritis pain: evidence from human genetic association and null mutant phenotype. *Annals of the Rheumatic Diseases* 71, 1042-1048
- 54 Valdes, A.M. et al. (2011) The Ile585Val TRPV1 variant is involved in risk of painful knee osteoarthritis. *Annals of the Rheumatic Diseases* 70, 1556-1561
- 55 Geyer, M. et al. (2009) Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes on osteoarthritis pathophysiology. *Osteoarthritis and Cartilage* 17, 328-335
- 56 Sánchez-Sabaté, E. et al. (2009) Identification of differentially expressed genes in trabecular bone from the iliac crest of osteoarthritic patients. *Osteoarthritis and Cartilage* 17, 1106-1114
- 57 Karlsson, C. et al. (2010) Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis and Cartilage* 18, 581-592
- 58 Xu, Y. et al. (2012) Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis and Cartilage* 20, 1029-1038
- 59 Del Rey, M.J. et al. (2012) Transcriptome analysis reveals specific changes in osteoarthritis synovial fibroblasts. *Annals of the Rheumatic Diseases* 71, 275-280
- 60 Leijten, J.C. et al. (2012) Gremlin 1, frizzled-related protein, and Dkk-1 are key regulators of human articular cartilage homeostasis. *Arthritis and Rheumatism* 64, 3302-3312
- 61 Ozsolak, F. and Milos, P.M. (2012) RNA sequencing: advances, challenges and opportunities. *Nature Review. Genetics* 12, 87-98
- 62 Xu, Y. et al. (2012) Comparison of osteoarthritis and normal hip cartilage transcriptomes using RNA-seq reveals new candidate gene targets and associated pathways. *Osteoarthritis and Cartilage* 20 (Suppl 1), S68
- 63 Henrotin, Y. et al. (2012) Fibulin 3 peptides Fib3-1 and Fib3-2 are potential biomarkers of osteoarthritis. *Arthritis and Rheumatism* 64, 2260-2267
- 64 Mateos, J. et al. (2012) Differential protein profiling of synovial fluid from rheumatoid arthritis and osteoarthritis patients using LC-MALDI TOF/TOF. *Journal of Proteomics* 75, 2869-2878
- 65 Önnérfford, P. et al. (2012) Quantitative proteomic analysis of eight cartilaginous tissues reveals characteristic differences as well as similarities between subgroups. *Journal of Biological Chemistry* 287, 18913-18924
- 66 Mobasheri, A. (2011) Applications of proteomics to osteoarthritis, a musculoskeletal disease characterized by aging. *Frontiers in Physiology* 2, 108
- 67 Califano, A. et al. (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature Genetics* 44, 841-847

Further reading, resources and contacts

There are several organizations promoting research into OA and encouraging public understanding of the disease. The principal international organisation is Osteoarthritis Research Society International (OARSI); others include the European League against Rheumatism (EULAR) and the American College of Rheumatology (ACR):

<http://www.oarsi.org/>

<http://www.eular.org/>

<http://www.rheumatology.org/>

Funding bodies (both governmental and charitable) that substantially support OA research include the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the Arthritis Foundation and the Canadian Arthritis Network in North America, and Arthritis Research UK and the Dutch Arthritis Foundation in Europe:

<http://www.niams.nih.gov/>

(continued on next page)

Further reading, resources and contacts (*continued*)

<http://www.arthritis.org/>

<http://www.arthritisnetwork.ca/>

<http://www.arthritisresearchuk.org/>

<http://www.reumafonds.nl/>

Features associated with this article

Tables

Table 1. Examples of genes reported as associated with osteoarthritis (OA) through candidate gene studies.

Table 2. Genome-wide association scans performed utilising high genome coverage and large cohorts.

Table 3. The signals from the Rotterdam genome-wide association scan (GWAS) and from the full arcOGEN GWAS.

Table 4. Functional studies performed on *GDF5* SNP rs143383.

Table 5. Recent global gene-expression microarray studies relevant to human osteoarthritis (OA).

Citation details for this article

Louise N. Reynard and John Loughlin (2013) The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev. Mol. Med.* Vol. 15, e2, February 2013, doi:10.1017/erm.2013.4