Osteoarthritis and Cartilage 27 (2019) 1545–1556

Osteoarthritis and Cartilage

Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals

S.J. Rice † *, K. Cheung † ‡, L.N. Reynard †, J. Loughlin †

† Newcastle University, Institute of Genetic Medicine, Newcastle upon Tyne, UK
‡ Newcastle University, Bioinformatics Support Unit, Newcastle upon Tyne, UK

ARTICLE INFO

Article history:
Received 5 March 2019
Accepted 24 May 2019

Keywords:
DNA methylation
Genetic risk
SNPs
Gene expression

SUMMARY

Objective: Osteoarthritis (OA) is polygenic with over 90 independent genome-wide association loci so far reported. A key next step is the identification of target genes and the molecular mechanisms through which this genetic risk operates. The majority of OA risk-conferring alleles are predicted to act by modulating gene expression. DNA methylation at CpG dinucleotides may be a functional conduit through which this occurs and is detectable by mapping methylation quantitative trait loci, or mQTLs. This approach can therefore provide functional insight into OA risk and will prioritize genes for subsequent investigation. That was our goal, with a focus on the largest set of OA loci yet to be reported.
Method: We investigated DNA methylation, genotype and RNA sequencing data derived from the cartilage of patients who had undergone arthroplasty and combined this with *in silico* analyses of expression quantitative trait loci, epigenomes and chromatin interactions.
Results: We investigated 42 OA risk loci and in ten of these we identified 24 CpGs in which methylation correlated with genotype (false discovery rate (FDR) *P*-values ranging from 0.049 to $1.73x10^{-25}$). *In silico* analyses of these mQTLs prioritised genes and regulatory elements at the majority of the ten loci, with *COLGALT2* (encoding a collagen galactosyltransferase), *COL11A2* (encoding a polypeptide chain of type XI collagen) and *WWP2* (encoding a ubiquitin ligase active during chondrogenesis) emerging as particularly compelling target genes.
Conclusion: We have highlighted the pivotal role of DNA methylation as a link between genetic risk and OA and prioritized genes for further investigation.

Introduction

Genetic susceptibility is a major risk factor for osteoarthritis (OA)[1,2]. Genome-wide association studies (GWAS) have revealed the polygenic nature of OA and to date over 90 independent risk loci have been identified[3–6]. The largest and most recent OA GWAS was carried out using the full UK Biobank cohort, with a focus on hip and knee OA[6]. This study utilised the genotypes of up to 455,000 individuals across 17.5 million single nucleotide polymorphisms (SNPs). Meta-analysis with the arcOGEN dataset[7] resulted in 64 genome-wide significant loci, 52 of which were novel. A series of techniques including statistical FINEMAPPING and expression quantitative trait locus (eQTL) co-localisation were employed to identify putative effector genes at these loci.

DNA methylation at CpG dinucleotides is an epigenetic process that regulates gene expression[8]. Using the Illumina 450 genome-wide methylation array, we have shown previously that changes in DNA methylation levels occur in cartilage chondrocytes and that these changes correlate with SNP genotype at around 25% of OA genetic risk loci[9,10]. These CpGs are known as methylation quantitative trait loci, or mQTLs. At several OA loci, including *GDF5* and *RUNX2*, genotype at the associated SNP correlates not only with DNA methylation but also with gene expression[11,12]. This indicates that chondrocytes use DNA methylation as an intermediary between genotype and phenotype, with this epigenetic mechanism underpinning a large proportion of OA disease risk. Identifying mQTLs therefore offers insight into the functional mechanism of OA

* Address correspondence and reprint requests to: S.J. Rice, Newcastle University, Skeletal Research Group, Institute of Genetic Medicine, International Centre for Life, Newcastle upon Tyne, NE1 3BZ, UK. Tel: 44-1912418626; Fax: 44-1912418666.
E-mail addresses: sarah.rice@newcastle.ac.uk (S.J. Rice), kat.cheung@newcastle.ac.uk (K. Cheung), louise.reynard@newcastle.ac.uk (L.N. Reynard), john.loughlin@newcastle.ac.uk (J. Loughlin).

https://doi.org/10.1016/j.joca.2019.05.017

genetic susceptibility. Furthermore, if mQTL CpGs cluster in or close to a particular gene at a risk locus, their mapping prioritises that gene and functional regulators of the gene for further investigation as plausible effectors of the association signal.

In this current study, we focussed on the OA signals that emerged from the analysis of the full UK Biobank cohort[6] and searched our cartilage genetic and epigenetic datasets to identify novel OA mQTLs. We were able to investigate a total of 42 signals and found evidence of mQTLs at ten of these. We subsequently carried out an *in-silico* analysis on these signals, highlighting regulatory elements and their putative gene targets requiring detailed functional analyses.

Methods

mQTL analysis of OA risk loci

Genotype and DNA methylation data were generated previously using the Human Omni Express array and Infinium HumanMethylation450 array (Illumina)[9]. Both datasets were generated from 87 patients who had undergone joint arthroplasty for knee OA ($n = 57$), hip OA ($n = 14$), or neck-of-femur (NOF) fracture ($n = 16$). We used genotype at the UK Biobank GWAS SNP, or, if not captured, genotype at a proxy variant. Proxies were derived from European population data using the LDproxy tool[13]. The proxy with the highest r^2 relative to the GWAS SNP was chosen. At each locus, we investigated DNA methylation 1 Mb upstream and 1 Mb downstream of the association SNP. Linear regression was used to measure the relationship between methylation β-values (ranging from 0 to 1) and genotype (0, 1 or 2 copies of the minor allele) at the SNP. If the minor allele frequency (MAF) was low (<3 homozygotes in our cohort) we combined minor allele homozygotes with heterozygotes and compared these to major allele homozygotes.

RNA-sequencing (RNA-seq)

Hip cartilage RNA-seq data that we had generated previously was used to assess differential expression in genes of interest at each locus. Detailed analytical methods can be found within the original reports of this dataset[14–16].

In silico eQTL analysis

For each locus that had an mQTL, we determined which genes at the locus were expressed in cartilage using our RNA-seq data (TPM>1). We then assessed whether any of these cartilage-expressed genes had an eQTL for the relevant association SNP by searching the genotype-tissue expression database GTEx (https://www.gtexportal.org/home/).

In silico epigenome analysis and investigation of chromatin interactions

ROADMAP (http://www.roadmapepigenomics.org/) was searched for epigenomic data, including chromatin state and presence of enhancers, at signals harbouring an OA mQTL. We focussed on five cell types of relevance to the articulating joint: E006, embryonic stem cell derived mesenchymal stem cells (MSCs); E049, bone marrow derived cultured chondrocytes; E023, MSC derived adipocytes; E025, adipose derived MSCs; E026, bone marrow derived MSCs. We searched the WashU epigenome browser[17] to identify long range chromatin interactions extending from annotated regulatory elements containing CpGs of interest. All publicly available HiC and long-range chromatin interaction datasets were loaded for all cell types with available data and the region was searched visually to identify interactions stemming from the regulatory elements. The human breast cancer cell line MCF7 and chronic myeloid leukaemia cell line K562 interaction schema represent protein factor mediated chromatin interaction data measured by RNA polymerase-II Chromatin Interaction Analysis with Paired-End Tag (PolII ChIA-PET) data[18]. These data were produced as part of the ENCODE project. Data from the human lymphoblastoid cell line GM12878 were produced by RNA CTCF ChIA-PET[19].

Statistical analyses

All mQTL calculations were performed using Matrix eQTL[20]. In this analysis we investigated DNA methylation at 19,761 CpGs in correlation with genotype at 42 SNPs and the *P*-values were adjusted to account for the 19,761 tests performed using a false discovery rate (FDR) estimation based on Benjamini–Hochberg correction[21]. Differential expression analysis between OA and NOF RNA-seq data was carried out with the Bioconductor package DESeq2[22]. Hypothesis testing was performed using the DESeq2 implementation of the Wald test and *P*-values were adjusted accordingly.

Results

OA loci investigated

The UK Biobank GWAS identified 52 novel OA association signals[9]. We excluded two of these as a SNP identification number was not reported. We excluded a further five as the SNP MAF was 5% or less, which would preclude us from having an adequate number of minor-allele carriers to test for mQTLs in our 87 patients. Of the remaining 45 loci, eight were not directly genotyped on our array and also lacked a proxy variant in high LD ($r^2 > 0.7$); these were also excluded. This left 37 of the 52 novel loci that could be analysed by us. In addition to novel loci, the authors of the UK Biobank GWAS also reported on previously identified OA signals. For four of these, the association SNP in their GWAS was in modest LD ($r^2 < 0.7$) with the SNP that had originally been reported as associated with OA, which may indicate a separate and novel signal. We therefore included these in our analysis. These are the SNPs rs10974438 (*GLIS3*), rs2396502 (*RUNX2*), rs2820443 (*TGFB2*) and rs4775006 (*ALDH1A2*). The authors also reported an association to rs12901372, a *SMAD3* SNP that had previously shown suggestive evidence of association to OA. We also included this SNP. In total therefore, we analysed 42 loci; 37 novel and five previously reported (Supl. Table 2).

Identification of OA mQTLs

At each locus, we searched for correlations between genotype and methylation at all CpGs within a 2 Mb region flanking the association SNP. In total, we analysed methylation at 19,761 CpGs across the 42 loci (Supl. Table 3). This identified ten loci at which there was a significant correlation (FDR *P*-value <0.05) between genotype and methylation at one or more CpGs (Supl. Table 2). There were 24 such CpGs in total, with the largest number (nine) occurring at locus 10. Stratification by joint site, sex, and status (OA or NOF) did not identify any additional mQTLs. Below we discuss the results of each locus including the *in-silico* analyses.

Locus 1

rs11583641 (C>T, OA risk allele = C, MAF = 0.17) is located at chromosome 1q25.3. The SNP is not on our genotyping array and we therefore used the proxy rs10911472 ($r^2 = 0.99$) and investigated 291 CpGs at this locus. Genotype at rs10911472 correlated with methylation at one CpG, located 6 kb from the association

SNP: cg18131582 (FDR $P = 0.0027$; Table I). Both the SNP and CpG are located within the gene body of *COLGALT2* [Fig. 1(A)]. The region encompassing cg18131582 is marked as an enhancer in multiple relevant cell types in ROADMAP [Fig. 1(B)]. The region containing the CpG was also marked as an area of DNaseI hypersensitivity and contained binding sites for multiple transcription factors [Fig. 1(B)]. The OA risk C allele corresponded with decreased levels of methylation [Fig. 1(C)].

GTEx revealed one eQTL corresponding to genotype at the association signal; a *COLGALT2* eQTL in adrenal gland tissue, with the C allele correlating with a relative increase in gene expression. Our cartilage RNA-seq data showed a significant increase in *COLGALT2* expression in OA compared to NOF controls ($P = 8.43\text{x}10^{-8}$, Fig. 1(D)). Long range chromatin interactions were identified between the enhancer region containing cg18131582 and sequences towards the 3′ untranslated region (3′UTR) of *COLGALT2* in K562 cells [Supl. Fig. 1(A)].

Locus 2

rs62182810 (G>A, OA risk allele = A, MAF = 0.39) is located at chromosome 2q33.2. The SNP is not on our genotyping array and we therefore used the proxy rs2305417 ($r^2 = 0.95$) and investigated 147 CpGs. Genotype at rs2305417 correlated with methylation at one CpG, cg10114877 (FDR $P = 3.09\text{x}10^{-6}$; Table I), which is located 41.7 kb from the association SNP [Supl. Fig. 2(A)]. The OA risk allele correlated with an increase in methylation of the CpG [Supl. Fig. 2(B)]. GTEx eQTLs were identified for three cartilage-expressed genes from locus 2: *ICAL1*, *AL2CR8* and *NBEAL1*. All are located upstream of *RAPH1*. There was a significant increase in *ALS2CR8* expression in OA ($P = 0.02$; Supl. Fig. 2(C)).

Locus 3

rs11732213 (T>C, OA risk allele = T, MAF = 0.15) is located within an intron of *SLBP* on chromosome 4p16.3. The SNP is not on our array and we therefore used the proxy rs798756 ($r^2 = 0.99$) and investigated 1618 CpGs. Genotype at rs798756 correlated with methylation at two intergenic CpGs, cg20987369 and cg25007799 (FDR *P* of 0.0027 and 0.019, respectively; Table I), both located 125 kb upstream of the association SNP and falling only 85bp apart [Fig. 2(A)]. The OA risk allele correlated with a decrease in methylation for both CpGs [Fig. 2(B)]. GTEx identified eQTLs across multiple tissue types for several cartilage-expressed genes at this locus. There was a significant increase in expression in OA of three of these: *TMEM129* ($P = 3.05\text{x}10^{-7}$), *TACC3* ($P = 1.81\text{x}10^{-13}$) and *FGFR3* ($P = 6.99\text{x}10^{-19}$) [Fig. 2(C)].

Locus 4

rs9277552 (C>T, OA risk allele = C, MAF = 0.22) is located on chromosome 6p21.32. The SNP is not on our array and we therefore used the proxy rs9277557 ($r^2 = 0.81$) and investigated 3,574 CpGs. Genotype at rs9277557 correlated with methylation at four CpGs: cg25491704, cg02197634, cg02375585 and cg13921245 (FDR *P*-values of 0.0021, 0.011, 0.028 and 0.040, respectively; Table I). These span a 42 kb region on chromosome 6 and flank the association SNP [Fig. 3(A)]. This is a gene-rich region coding for HLA and non-HLA genes, including *COL11A2*.

The OA risk allele correlated with an increase in methylation at cg02375585 but a decrease in methylation at cg02197634, cg25491704 and cg13921245 [Fig. 3(C)]. These latter three CpGs fall within the gene body of *HLA-DPB1*, with cg02197634 and cg25491704 being only 4bp apart [Fig. 3(B)]. These two CpGs are located within an intron of *HLA-DPB1* and in the promoter region of the syntenic gene *HLA-DPA1*, with this promoter demonstrating long range chromatin interactions with the 3′UTR of *HLA-DPB1* [Supl. Fig. 1(B)]. This region is also one of DNase I hypersensitivity and transcription factor binding, with cg02197634 and cg25491704 being located within the binding site of POLR2A, the largest subunit of the eukaryotic mRNA synthesising enzyme RNA polymerase II.

cg02375585 is located within an intron of *HLA-DPB2*, which is a part of a repressed heterochromatic sequence that physically interacts with the promoter of *COL11A2* [Fig. 3(A) and Supl. Fig. 1(C)]. GTEx identified eQTLs across several tissue types for four cartilage-

Table I
The significant genotype–methylation associations identified

Locus	Chr.	GWAS SNP	SNP position (hg19)	Proxy SNP	r^2	CpG	CpG position (hg19)	Uncorrected *P*-value	FDR-corrected *P*-value	Slope (95% CI)
1	1	rs11583641	183906245	rs10911472	0.99	cg18131582	183912305	$1.92\text{x}10^{-6}$	0.0027	0.52 (0.32–0.72)
2	2	rs62182810	204387482	rs2305417	0.95	cg10114877	204427199	$9.37\text{x}10^{-10}$	$3.09\text{x}10^{-6}$	0.91 (0.65–1.17)
3	4	rs11732213	1704244	rs798756	0.99	cg20987369	1579572	$1.55\text{x}10^{-6}$	0.0027	0.33 (0.21–0.46)
						cg25007799	1579657	$1.96\text{x}10^{-5}$	0.019	0.71 (0.40–1.02)
4	6	rs9277552	33055501	rs9277557	0.81	cg25491704	33048879	$9.43\text{x}10^{-7}$	0.0021	0.93 (0.58–1.28)
						cg02197634	33048875	$1.05\text{x}10^{-5}$	0.011	1.02 (0.59–1.46)
						cg02375585	33091111	$3.11\text{x}10^{-5}$	0.028	−0.68 (−0.98–−0.37)
						cg13921245	33053791	$4.86\text{x}10^{-5}$	0.04	0.37 (0.20–0.54)
5	8	rs60890741	130768504	rs12542856	0.97	cg18170545	131080137	$6.20\text{x}10^{-5}$	0.049	−0.34 (−0.49–−0.18)
6	12	rs317630	69637847	rs317646	0.99	cg22375663	69725435	$8.45\text{x}10^{-15}$	$4.18\text{x}10^{-11}$	1.06 (0.84–1.28)
7	15	rs35206230	75097780	rs1378940	0.96	cg20040747	74715105	$1.83\text{x}10^{-6}$	0.0027	−0.27 (−0.37–−0.17)
						cg10253484	75165896	$3.12\text{x}10^{-5}$	0.028	0.36 (0.20–0.52)
8	16	rs6499244	69735271	rs7359336	0.91	cg26661922	69951706	$8.27\text{x}10^{-6}$	0.0096	−0.46 (−0.66–−0.27)
						cg26736200	69951820	$9.98\text{x}10^{-6}$	0.011	−0.51 (−0.72–−0.29)
9	17	rs2953013	29496343	Not required	n/a	cg16779580	29740988	$4.86\text{x}10^{-6}$	0.006	−0.51 (−0.72–−0.30)
10	17	rs62063281	44038785	rs17650901	0.98	cg17117718	43663208	$8.78\text{x}10^{-30}$	$1.73\text{x}10^{-25}$	1.63 (1.45–1.82)
						cg15633388	44266530	$1.93\text{x}10^{-21}$	$1.91\text{x}10^{-17}$	−1.52 (−1.75–−1.28)
						cg18228076	43983362	$9.35\text{x}10^{-14}$	$3.69\text{x}10^{-10}$	−1.09 (−1.33–−0.85)
						cg15295732	43942128	$1.10\text{x}10^{-9}$	$3.09\text{x}10^{-6}$	−0.41 (−0.52–−0.29)
						cg23616531	44269258	$2.07\text{x}10^{-7}$	0.00051	−0.39 (−0.52–−0.25)
						cg10826688	43714992	$1.48\text{x}10^{-6}$	0.0027	−0.31 (−0.43–−0.19)
						cg11117266	43971461	$1.75\text{x}10^{-6}$	0.0027	−0.49 (−0.68–−0.30)
						cg16520312	43971471	$2.76\text{x}10^{-6}$	0.0036	−0.32 (−0.44–−0.19)
						cg01934064	44064242	$4.33\text{x}10^{-5}$	0.037	0.31 (0.17–0.45)

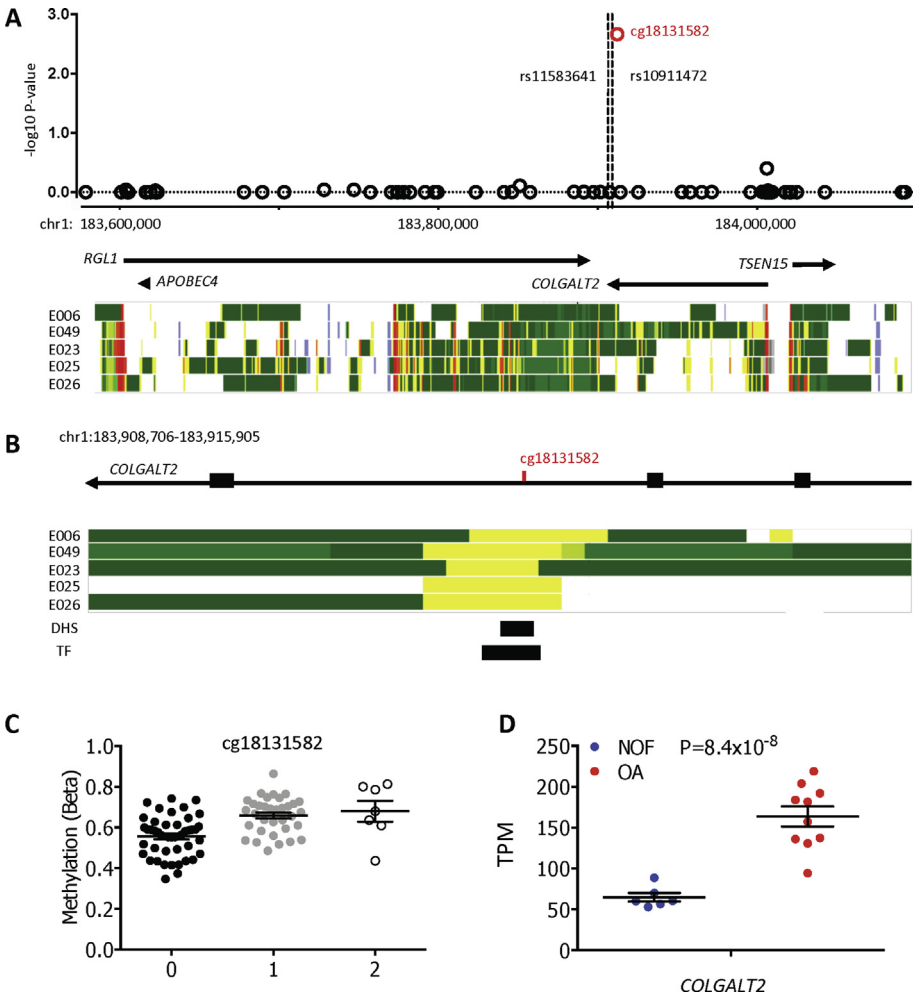

Fig. 1. Genotype at rs11583641/rs10911472 correlates with the methylation of CpG cg18131582, located in the gene body of *COLGALT2*. (**A**) The plot shows the association between genotype and methylation levels of CpGs probes within the locus. The *x*-axis represents the genomic position of the CpGs and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG, with cg18131582 highlighted in red. The locations of rs11583641 and rs10911472 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in five relevant cell types: red, active transcription start site (TSS); green, strong transcription; orange, active enhancer. (**B**) Zoomed in image showing the location of cg18131582 in an enhancer. DHS, DNaseI hypersensitivity; TF, transcription factor binding sites. (**C**) The association between genotype at rs10911472 and cg18131582 methylation for all 87 samples (FDR *P*-value = 0.0027). The level of methylation is shown as the β-value. Horizontal lines represent the mean and standard error of the mean (SEM). 0, major allele homozygote; 1, heterozygote; 2, minor allele homozygote. (**D**) Expression of *COLGALT2* in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

expressed genes at this locus: *HLA-DPA1, HLA-DPB1, COL11A2* and *HSD17B8*. There was a significant increase in expression of *COL11A2* in OA cartilage ($P = 7.7\text{x}10^{-5}$; Fig. 3(D)).

Locus 5

rs60890741 (C>CA, OA risk allele = C, MAF = 0.18) is located on chromosome 8q24.2 and marks a region of LD spanning 24.5 kb rs60890741 is not on our array and we therefore used the proxy rs12542856 ($r^2 = 0.97$) and analysed 166 CpGs. We identified a single CpG, cg18170545, at which methylation significantly correlated with genotype at rs12542856 (FDR $P = 0.049$; Table I). This CpG falls within an intron of *ASAP1* [Supl. Fig. 3(A)] in a predicted enhancer region containing multiple transcription factor binding sites [Supl. Fig. 3(B)]. The CpG is hypermethylated in cartilage, with the OA risk allele correlating with increased methylation [Supl. Fig. 3(C)]. There were no cartilage-expressed genes at this locus that had GTEx eQTLs.

Locus 6

rs317630 (C>T, OA risk allele = T, MAF = 0.23) is located on chromosome 12q15 and falls within an intron of *CPSF6* [Supl. Fig. 4(A)]. rs317630 is not on our array and we therefore used the proxy rs317646 ($r^2 = 0.99$) and analysed 239 CpGs. We identified a single CpG, cg22375663, at which methylation significantly correlated with genotype at rs317646 (FDR P=4.18x10^{-11}; Table I). This CpG is intergenic, located 88 kb upstream of the association SNP and falls within a region marked as a bivalent enhancer and promoter [Supl. Fig. 4(B)]. cg22375663 falls within the binding site for POLR2A. The OA risk allele correlates with an increase in methylation of the CpG [Supl. Fig. 4(C)]. GTEx identified eQTLs across multiple tissue types for three cartilage-expressed genes at this locus: *CPSF6*, *LYZ* and *YEATS4*. There was no differential expression for any of these between OA and NOF cartilage [Supl. Fig. 4(D)].

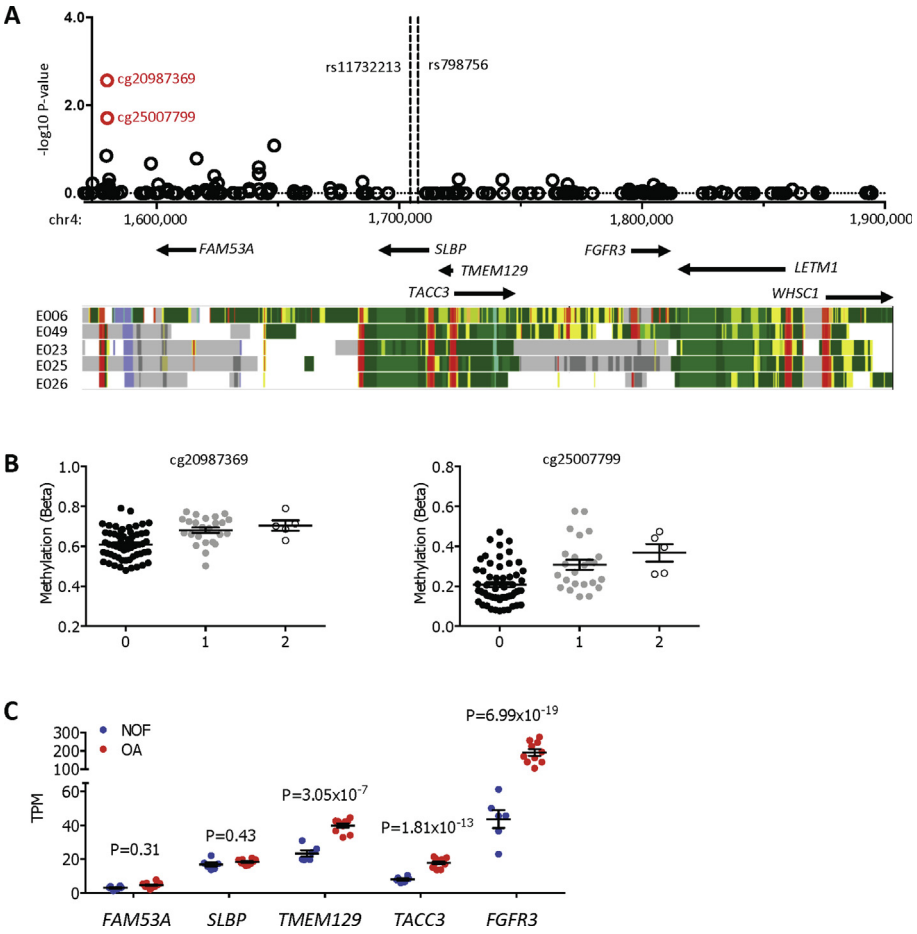

Fig. 2. Genotype at rs11732213/rs798756 correlates with the methylation of CpGs cg20987369 and cg25007799. (**A**) The plot shows the association between genotype and methylation levels of CpGs within the locus. The *x*-axis represents the genomic position of the CpGs and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG, with cg20987369 and cg25007799 highlighted in red. The locations of rs11732213 and rs798756 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in five relevant cell types: red, active TSS; green, strong transcription; orange, active enhancer. (**B**) The association between genotype at rs798756 and methylation at cg20987369 and cg25007799 for all 87 samples (FDR *P*-values = 0.0027 and 0.019, respectively). The level of methylation is shown as the β-value. Horizontal lines represent the mean and SEM. 0, major allele homozygote; 1, heterozygote; 2, minor allele homozygote. (**C**) Expression of genes at the locus in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

Locus 7

rs35206230 (C>T, OA risk allele = T, MAF = 0.24) is located on chromosome 15q24.1. rs35206230 is not on our array and we therefore used the proxy rs1378940 (r^2 = 0.96) and analysed 761 CpGs. Genotype at rs1378940 correlated with methylation at two CpGs, cg20040747 and cg10253484 (FDR *P*-values of 0.0027 and 0.028, respectively; Table I).

cg20040747 is located within an intron of *SEMA7A* [Fig. 4(A)] within a DNaseI hypersensitive enhancer region containing transcription factor binding sites [Fig. 4(B)]. cg10253484 is located 451 kb downstream of cg20040747, within the promoter region of *SCAMP2* [Fig. 4(B)]. Chromatin interactions were identified between this promoter and that of the neighbouring gene *MPI* [Supl. Fig. 1(D)]. The OA risk is associated with a decreased methylation of cg20040747 but an increased methylation of cg10253484 [Fig. 4(C)].

GTEx identified eQTLs at nine cartilage-expressed genes from within this locus: *CSK*, *LMAN1L*, *ULK3*, *SCAMP2*, *MPI*, *FAM219B*, *RPP25*, *SCAMP5* and *PPCDC*. There was no differential expression for any of these between OA and NOF cartilage [Fig. 4(D)].

Locus 8

rs6499244 (A>T, OA risk allele = A, MAF = 0.29) is located on chromosome 16q22.1. rs6499244 is not on our array and we therefore used the proxy rs7359336 (r^2 = 0.91) and analysed 453 CpGs. Genotype at rs7359336 correlated with methylation at two CpGs, cg26661922 and cg26736200 (FDR *P*-values of 0.0096 and 0.011, respectively; Table I). Both CpGs fall within the body of *WWP2* [Fig. 5(A)] in a region marked as having enhancer activity [Fig. 5(B)], with the OA risk allele correlating with increased methylation at each CpG [Fig. 5(C)]. GTEx identified eQTLs at five cartilage-expressed genes at this locus: *CLEC18A*, *IL34*, *NFAT5*, *NOB1* and *PDXDC2P*. There was a significant increase in expression of *CLEC18A* in OA cartilage (*P* = 0.02; Fig. 5(D)).

Locus 9

rs2953013 (T>G, OA risk allele = G, MAF = 0.32) is located on chromosome 17q11.2. This SNP was directly genotyped on the array and we analysed 313 CpGs. Genotype at rs2953013 correlated with methylation at one CpG, cg16779580 (FDR

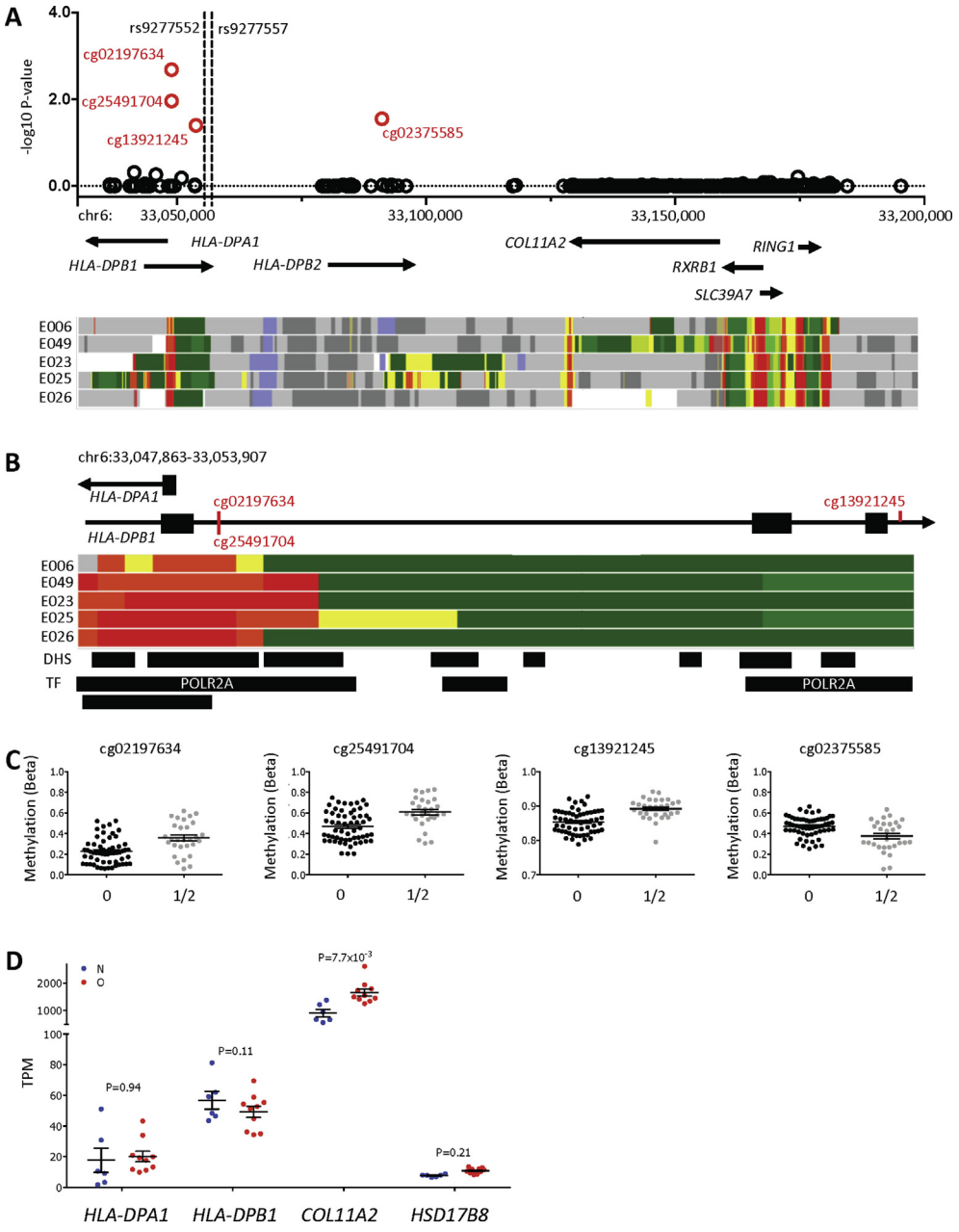

Fig. 3. Genotype at rs9277552/rs9277557 correlates with the methylation of four CpGs. (**A**) The plot shows the association between genotype and methylation levels of CpGs within the locus. The *x*-axis represents the genomic position of the CpGs and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG probe. The four significant CpGs are highlighted in red. The locations of rs9277552 and rs9277557 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in five relevant cell types: red, active TSS; green, strong transcription; orange, active enhancer. (**B**) Zoomed in image of *HLA-DPB1* showing the location of three of the CpGs relative to regulatory elements. DHS, DNaseI hypersensitivity; TF, transcription factor binding sites. (**C**) The association between genotype at rs9277557 and methylation at the four CpGs for all 87 samples (FDR *P*-values range from 0.0021 to 0.04). The level of methylation is shown as the β-value. Horizontal lines represent the mean and SEM. Due to the relatively modest minor allele frequency (MAF) for rs9277557, minor allele homozygotes were combined with heterozygotes (labelled 1/2) and compared to major allele homozygotes (labelled 0). (**D**) Expression of genes at the locus in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

$P = 0.006$; Table I). cg16779580 resides within *RAB11FIP4* [Supl. Fig. 5(A)] and specifically within an intronic enhancer region of the gene [Supl. Fig. 5(B)], with the OA risk allele correlating with decreased methylation of the CpG [Supl. Fig. 5(C)]. GTEx identified eQTLs at three cartilage-expressed genes at this locus: *NF1*, *EV12A* and *RAB11FIP4*. A significantly increased expression of *RAB11FIP4* was observed in OA cartilage ($P = 0.05$; Supl. Fig. 5(D)).

Locus 10

rs62063281 (T>C, OA risk allele = C, MAF = 0.22) is located on chromosome 17q21.31. This SNP was not directly genotyped on the array and we therefore used the proxy rs17650901 ($r^2 = 0.98$) and analysed 501 CpGs. Genotype at rs17650901 correlated with methylation at nine CpGs, with FDR *P*-values ranging from 0.037 to $1.73\text{x}10^{-25}$ (Table I).

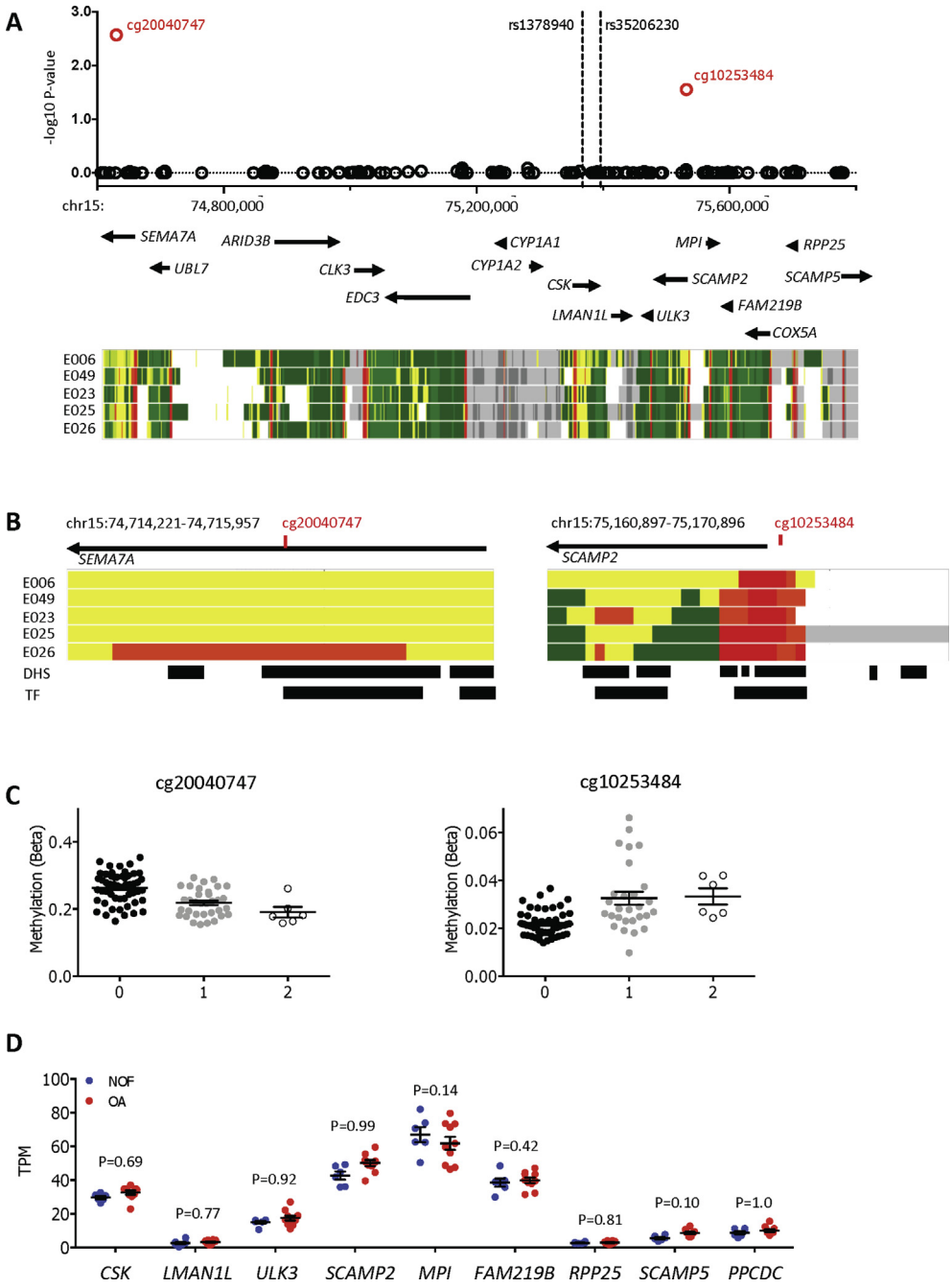

Fig. 4. Genotype at rs35206230/rs1378940 correlates with the methylation of CpGs cg20040747 and cg10253484. (**A**) The plot shows the association between genotype and methylation levels of CpGs within the locus. The *x*-axis represents the genomic position of the CpG probes, and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG probe, with cg20040747 and cg10253484 highlighted in red. The locations of rs35206230 and rs1378940 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in relevant cell types: red, active TSS; green, strong transcription; orange, active enhancer. (**B**) Zoomed in image of *SEMA7A*, showing the location of cg20040747 within an enhancer, and of *SCAMP2*, showing the location of cg10253484 within a promoter. DHS, DNaseI hypersensitivity; TF, transcription factor binding sites. (**C**) The association between rs1378940 genotype and methylation at cg20040747 and cg10253484 for all 87 samples (FDR *P*-values = 0.0027 and 0.028, respectively). The level of methylation is shown as the β-value. Horizontal lines represent the mean and SEM. 0, major allele homozygote; 1, heterozygote; 2, minor allele homozygote. (**D**) Expression of genes at the locus in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

The nine positive CpGs demonstrate hypo- and hyper-methylation, with the OA risk allele correlating with increased and decreased levels of methylation [Fig. 6(C)]. Several of the CpGs reside within genes and within potential functional domains [Fig. 6(A)] with four meriting particular comment. cg11117266 and cg16520312, which are located just 10bp apart, fall within the promoter region of *MAPT* [Fig. 6(B)], whilst cg15633388 and cg23616531 are located within the promoter region of *KANSL1* and an adjacent enhancer element [Fig. 6(C)]. These regulatory elements of *MAPT* and *KANSL1* physically interact with each other [Supl. Fig. 1(E)], implying coordinated functionality of DNA sequences harbouring the four CpGs.

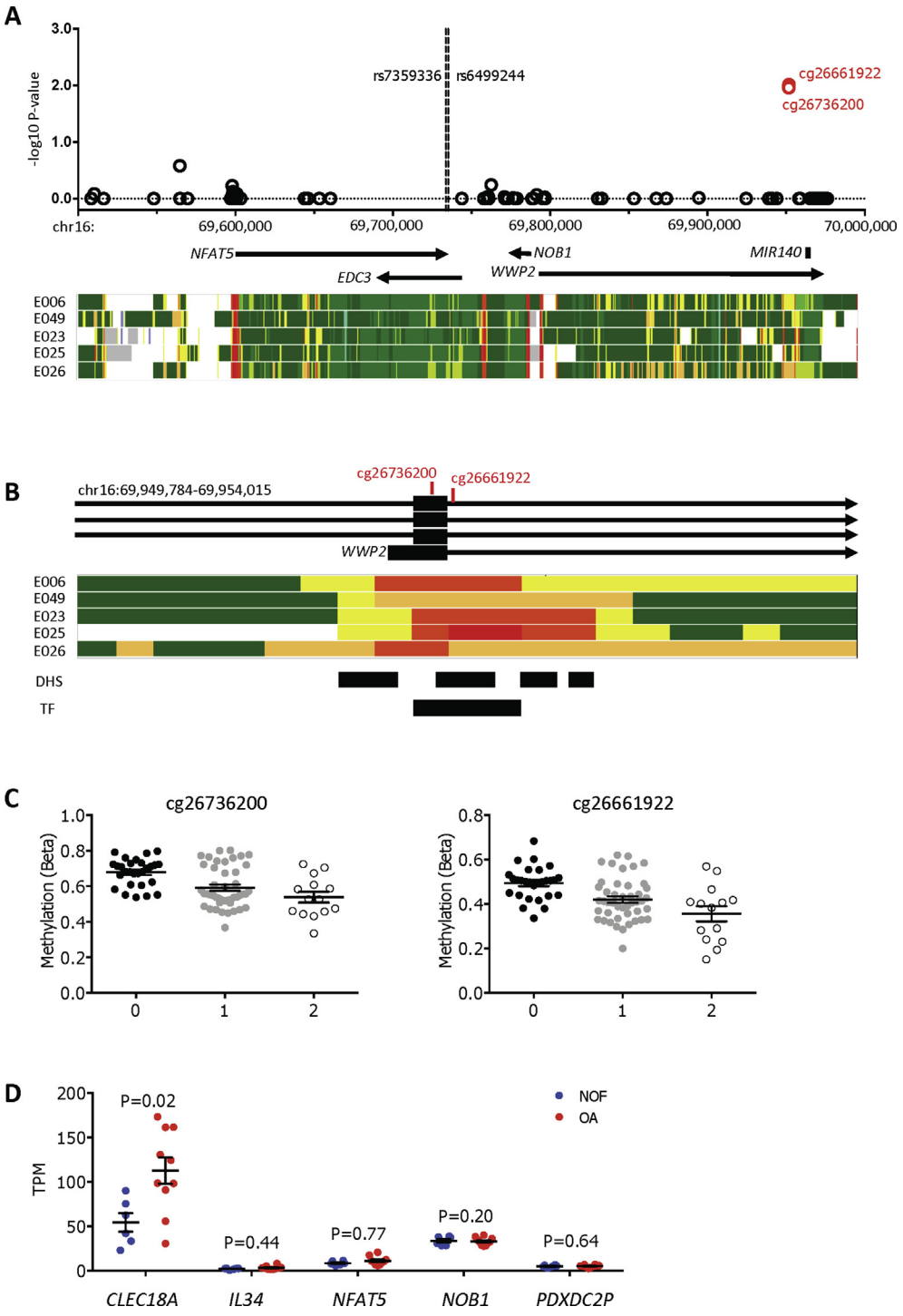

Fig. 5. Genotype at rs6499244/rs7359336 correlates with the methylation of CpGs cg26736200 and cg26661922. (**A**) The plot shows the association between genotype and methylation levels of CpGs within the locus. The *x*-axis represents the genomic position of the CpGs and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG probe, with cg26736200 and cg26661922 highlighted in red. The locations of rs6499244 and rs7359336 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in relevant cell types: red, active TSS; green, strong transcription; orange, active enhancer. (**B**) Zoomed in image of *WWP2* showing the location of cg26736200 and cg26661922 within promoter and enhancer regions. DHS, DNaseI hypersensitivity; TF, transcription factor binding sites. (**C**) The association between rs7359336 genotype and methylation at cg26736200 and cg26661922 for all 87 samples (FDR *P*-values = 0.0096 and 0.011, respectively). The level of methylation is shown as the β-value. Horizontal lines represent the mean and SEM. 0, major allele homozygote; 1, heterozygote; 2, minor allele homozygote. (**D**) Expression of genes at the locus in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

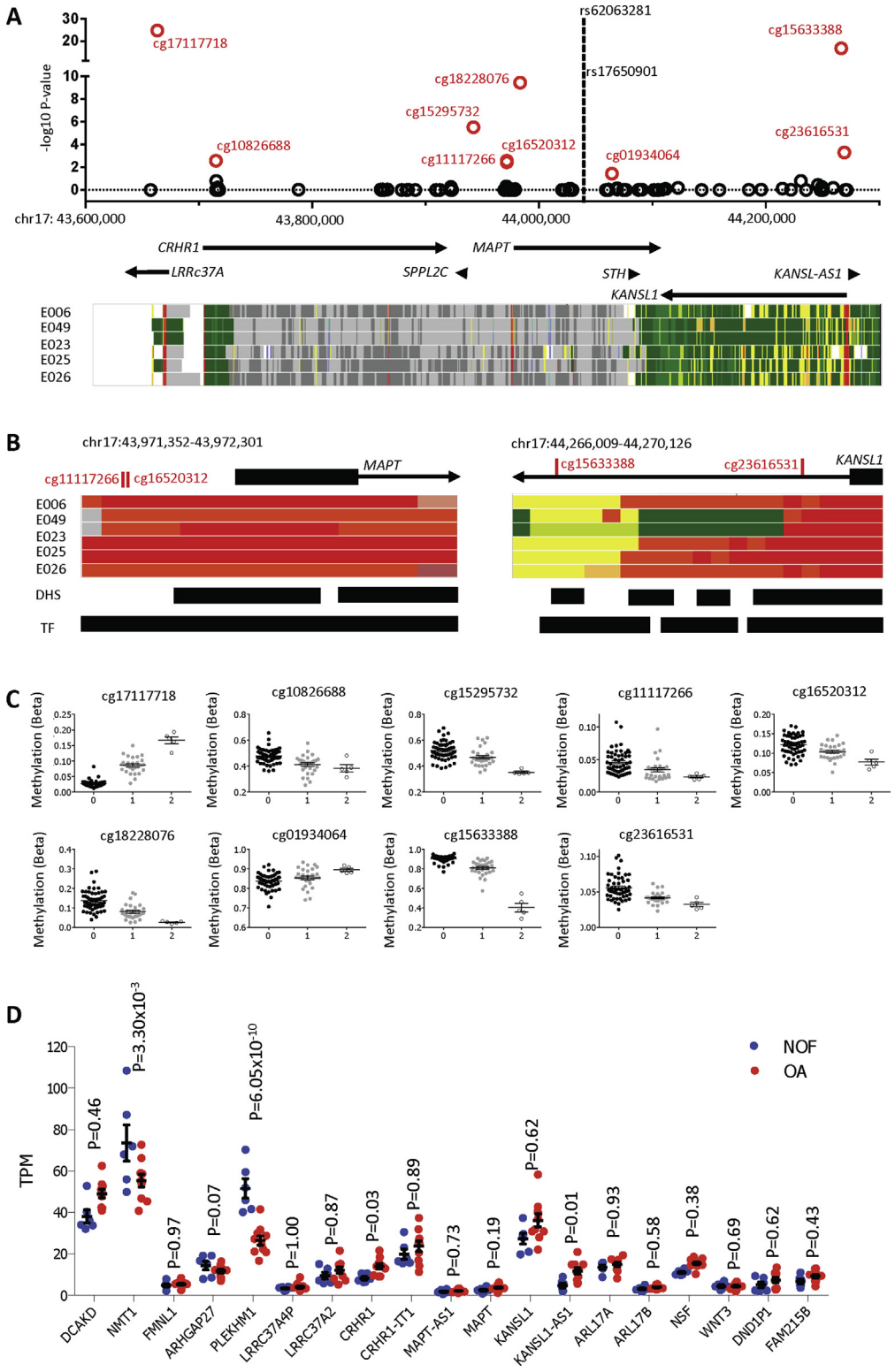

Fig. 6. Genotype at rs62063281/rs17650901 correlates with the methylation status of nine CpGs. (**A**) The plot shows the association between genotype and methylation levels of CpGs within the region. The *x*-axis represents the genomic position of the CpG probes and the *y*-axis the Benjamini-Hochberg corrected -$\log_{10}$ *P*-value of the correlation between genotype and M-value at each CpG. Each open circle represents a single CpG probe, with the nine CpGs highlighted in red. The locations of rs62063281and rs17650901 are indicated. The genes within the region analysed are shown below the association plot, with the gene direction indicated by arrows. The lower panel shows data from the ROADMAP epigenome database, marking chromatin states in relevant cell types: red, active TSS; green, strong transcription; orange, active enhancer. (**B**) Zoomed in images of *MAPT*, showing the location of cg11117266 and cg16520312 within the promoter of the gene, and of *KANSL1*, showing the location of cg15633388 and cg23616531 within the promoter of the gene and within an adjacent enhancer. DHS, DNaseI hypersensitivity; TF, transcription factor binding sites. (**C**) The association between rs17650901 genotype and methylation at the nine CpGs for all 87 samples (FDR *P*-values range from $1.73\text{x}10^{-25}$ to 0.037). The level of methylation is shown as the β-value. Horizontal lines represent the mean and SEM. 0, major allele homozygote; 1, heterozygote; 2, minor allele homozygote. (**D**) Expression of genes at the locus in cartilage from OA and NOF patients. TPM, transcripts per million. Bars represent the mean and the SEM. *P*-values were calculated using a Wald test within the DESeq2 package.

GTEx identified eQTLs for 19 cartilage-expressed genes at this locus, including *MAPT* and *KANSL1* [Fig. 6(D)]. Four showed a significant differential expression between OA and NOF cartilage, with *NMT1* ($P = 3.30\text{x}10^{-3}$) and *PLEKHM1* ($P = 6.05\text{x}10^{-10}$) demonstrating reduced expression in OA, and *CRHR1* (P=0.03) and *KANSL1-AS1* ($P = 0.01$) demonstrating increased expression.

Discussion

GWAS are widely used in the identification of risk-conferring variants in a range of complex traits. Despite this, there remains an enormous disparity between the discovery of GWAS signals and subsequent functional analyses, which are essential in their interpretation and the identification of intervenable targets[23]. In this report, we utilised genome wide genetic and epigenetic datasets and discovered *cis* mQTLs operating at novel OA risk loci. This highlights the potential mechanistic importance of DNA methylation in mediating gene expression. This study aims to bridge the gap between GWAS and the execution of successful functional analyses, by prioritising disease-associated regulatory elements and genes. There have now been 18 OA mQTLs identified to date. This is likely a conservative number considering the low genomic coverage of the Illumina methylation array, which captures less than 1% of CpGs. Our results emphasise the key function of DNA methylation in mediating a large proportion of OA susceptibility. Due to the joint specificity of both the cartilage methylome and OA GWA signals, we found it somewhat surprising that we did not identify any hip- or knee-specific mQTLs. However, this is consistent with our previous analyses[9,10]. We hypothesise that, whilst OA mQTLs are active at both joint sites, they might not necessarily exert a functional effect at all sites due, perhaps, to variability in chromatin state, or distinct compensatory mechanisms within a joint. Additionally, it is feasible that the enhancers marked by mQTLs are functionally active during different stages of the lifespan and could be integral for correct cartilage development and maintenance within the joints at distinct time points. At each novel mQTL we analysed cartilage RNA-seq data and online datasets, including eQTL and long-range chromatin interaction data, to highlight functionally relevant genes and regulatory elements that could be the targets of OA genetic risk. In this regard, loci 1, 4 and 8 were particularly compelling.

At locus 1, (rs11583641, C>T), cg18131582 falls within a predicted *COLGALT2* enhancer. There is a physical interaction between this enhancer and the body of the gene, which implies that the enhancer could regulate *COLGALT2* expression. The gene encodes procollagen galactosyltransferase 2, an enzyme that post-translationally glycosylates collagen[24]. Collagens are core structural and functional components of the cartilage extracellular matrix (ECM) with their post-translational modification being key to their activity[25]. Polymorphism in *COLGALT2* has previously been associated with skeletal development[26] whilst in our RNA-seq data, there was a highly significant increased expression of the gene in OA hip cartilage. This increase in *COLGALT2* expression has also been observed in knee cartilage[27]. In a recent study of the OA cartilage transcriptome performed on a Netherlands cohort of 47 patient samples[28], evidence of allelic expression imbalance at *COLGALT2* was reported, implying that the gene is subject to cartilage eQTLs. We hypothesise that in OA there is an attempted reparative response of the cartilage involving new collagen synthesis with a concurrent upregulation of *COLGALT2*. This increased expression requires methylation of the enhancer falling within the gene body of *COLGALT2*. However, the OA risk allele attenuates this response, manifested as lower enhancer methylation and an inadequate increase in *COLGALT2* expression. This ultimately has a negative impact on the post-translational modification of newly synthesised collagens, hindering repair.

At locus 4, (rs9277552, C>T), there were four mQTL CpGs, three within *HLA-DPB1* and one within *HLA-DPB2*. The *HLA-DPB2* CpG, cg02375585, marks a heterochromatic region that interacts with the promoter of *COL11A2*. This gene encodes the α2 polypeptide chain of type XI collagen, a critical structural component of articular cartilage. Type XI collagen is essential for the stability of type II collagen fibrils, which comprise 90% of the collagen network of the cartilage ECM[29]. Mutations in *COL11A2* cause Stickler syndrome, a skeletal abnormality that often presents with early-onset OA[30]. In our RNA-seq data, *COL11A2* is abundantly expressed with a highly significant increased expression in OA vs non-OA cartilage. Similarly, a significant increase in *COLL11A2* expression was detected in OA knee cartilage[27]. This may reflect a cartilage reparative response. Unfortunately, the HLA locus was excluded from analysis in the Netherlands cartilage transcriptome study[28]. On GTEx, the OA risk-conferring C allele of rs9277552 correlates with both increased and decreased expression of the gene in non-cartilaginous tissues. We speculate that in cartilage, this risk conferring allele will also correlate with differential expression of the gene and that akin to *COLGALT2,* this expression change is mediated by differential enhancer methylation.

At locus 8, (rs6499244, A>T), there were two mQTL CpGs, both located in an enhancer within *WWP2*. This gene encodes NEDD4-like E3 ubiquitin-protein ligase, a protein that is expressed at high abundance in cartilage and which plays an essential role in chondrogenesis via TGFβ/Smad signaling[31]. Knockout of the protein results in altered craniofacial patterning and palatogenesis[32,33]. In our search of GTEx, we did not find any *WWP2* eQTLs correlating with the OA association signal and there was no differential expression of the gene between our OA and non-OA cartilage (data not shown). However, the Netherlands cartilage transcriptome study reported allelic expression at the gene[27]. The transcript SNPs used are in modest LD with the association SNP, with the most significant *WWP2* transcript SNP, rs1052429 (A/G), having an r^2 of only 0.22 with rs6499244. Despite this, the OA risk conferring A allele of rs6499244 occurs almost exclusively on a haplotype containing the A allele of rs1052429, which correlates with increased *WWP2* expression. We therefore hypothesise that the risk allele at this locus leads to increased methylation of the enhancer and subsequent increased expression of *WWP2*, which is detrimental to cartilage health. *WWP2* encompasses microRNA 140, which also regulates chondrocyte activity[34] and could also be a target of this mQTL.

The locus with the largest number of significant CpGs in our study was locus 10, with nine including three with FDR *P*-values $<1.0\text{x}10^{-15}$. This is an extremely complex locus of very high LD containing a large number of polymorphisms. These span an interval of 945 kb and contain at least nine recently evolved common haplotypes, including inversions and repeats[35]. The large number of highly significant CpGs at this locus, which are spread across a number of genes, many of which are expressed in cartilage, make this a highly intriguing signal. None of the 19 cartilage-expressed genes at the locus that have eQTLs in GTEx were reported as demonstrating allelic expression in the Netherlands cartilage transcriptome study. However, this could simply reflect the relatively modest sample size used in that transcriptome report. Future functional studies of this complex locus may require a gene by gene approach.

In summary, we have identified novel mQTLs correlating to genotype at OA risk loci. Through this analysis, we have more than doubled the number of such mQTLs and in several instances, we have highlighted putative effector genes through the co-localisation of CpGs to regulatory elements; *COLGALT2*, *COL11A2*

and *WWP2* represent clear examples. Our results reveal potential mechanisms by which the OA-encoded risk at several of the loci mediate their effect and as such, the significant CpGs and their regulatory elements, and putative target genes now warrant a detailed functional investigation in the context of OA molecular aetiology.

Author contributions

SJR, JL, and LNR designed the study. KC carried out the regression analysis of methylation and genotype data. SJR interpreted the analysis results, collated data from online databases, and prepared the manuscript figures. SJR and JL prepared the main manuscript text. All authors contributed to writing or reviewing the manuscript and final approval.

Competing interest statement

There are no conflicts of interest.

Role of the funding source

This work was supported by Versus Arthritis (grant 20771), by the Medical Research Council and Arthritis Research UK as part of the MRC-Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing (CIMA) UK (grant references JXR 10641, MR/P020941/1 and MR/R502182/1), and by the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement number No. 305815 (D-BOARD). The funders had no role in the design, data collection, analysis and interpretation of data, in the writing of the manuscript or in the decision to submit the manuscript for publication.

Data availability

All data are available from the corresponding author upon request. RNA-seq data are available from GEO (GSE111358).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2019.05.017.

References

1. Loughlin J. Genetic indicators and susceptibility to osteoarthritis. Br J Sports Med 2011;45:278–82.
2. Reynard LN. Analysis of genetics and DNA methylation in osteoarthritis: what have we learnt about the disease? Semin Cell Dev Biol 2017;62:57–66.
3. Styrkarsdottir U, Helgason H, Sigurdsson A, Norddahl GL, Agustsdottir AB, Reynard LN, *et al.* Whole-genome sequencing identifies rare genotypes in COMP and CHADL associated with high risk of hip osteoarthritis. Nat Genet 2017;49:801–5.
4. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, *et al.* Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet 2018;50:549–58.
5. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stefansson OA, Sigurdsson JK, *et al.* Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet 2018;50:1681–7.
6. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, *et al.* Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank. Nat Genet 2019;51:230–6.
7. arcOGEN Consortium, arcOGEN Collaborators, Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet 2012;380:815–23.
8. Young DA, Bui C, Barter MJ. Understanding CpG methylation in the context of osteoarthritis. Epigenomics 2012;4:593–5.
9. Rushton MD, Reynard LN, Young DA, Shepherd C, Aubourg G, Gee F, *et al.* Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. Hum Mol Genet 2016;24:7432–44.
10. Rice SJ, Tselepi M, Sorial AK, Aubourg G, Shepherd C, Almarza D, et al. Prioritization of PLEC and GRINA as osteoarthritis risk genes through the identification and characterization of novel methylation quantitative trait loci. Arthritis Rheum (in press). doi: 10.1002/art.40849. [Epub ahead of print].
11. Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis SNP rs143383. Hum Genet 2014;133:1059–73.
12. Rice SJ, Aubourg G, Sorial AK, Almarza D, Tselepi M, Deehan DJ, *et al.* Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. Hum Mol Genet 2018;27:3464–74.
13. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31:3555–7.
14. Bui C, Barter MJ, Scott JL, Xu Y, Galler M, Reynard LN, *et al.* cAMP response element-binding (BREB) recruitment following a specific CpG demethylation leads to the elevated expression of the matrix metalloproteinase 13 in human articular chondrocytes and osteoarthritis. FASEB J 2012;26:3000–11.
15. Shepherd C, Skelton AJ, Rushton MD, Reynard LN, Loughlin J. Expression analysis of the osteoarthritis genetic susceptibility locus mapping to an intron of the MCF2L gene and marked by the polymorphism rs11842874. BMC Med Genet 2015;16:108.
16. Shepherd C, Zhu D, Skelton AJ, Combe J, Threadgold H, Zhu L, *et al.* Functional characterisation of the osteoarthritis genetic risk residing at ALDH1A2 identifies rs12915901 as a key target variant. Arthritis Rheum 2018;70:1577–87.
17. Zhou X, Lowdon RF, Li D, Lawson HA, Madden PA, Costello JF, *et al.* Exploring long-range genome interactions using the WashU Epigenome Browser. Nat Methods 2013;10:375–6.
18. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012;148:84–98.
19. Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, *et al.* CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell 2015;163:1611–27.
20. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 2012;28:1353–8.
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 1995;57:289–300.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
23. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. Am J Hum Genet 2018;102:717.
24. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Mol Cell Biol 2009;4:943–52.

25. Hardin JA, Cobelli N, Santambrogio L. Consequences of metabolic and oxidative modifications of cartilage tissue. Nat Rev Rheumatol 2015;11:521–9.
26. Zhao J, Li M, Bradfield JP, Zhang H, Mentch FD, Wang K, *et al*. The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. BMC Med Genet 2010;11:96.
27. Soul J, Dunn SL, Anand S, Serracino-Inglott F, Schwartz J, Boot-Handford RP, *et al*. Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage. Ann Rheum Dis 2018;77:423.
28. den Hollander W, Pulyakhina I, Boer C, Bomer N, van der Braggen R, Arindrarto W, et al. Annotating transcriptional effects of genetic variants in disease-relevant tissue: transcriptome-wide allelic imbalance in osteoarthritic cartilage. Arthritis Rheum (in press). doi: 10.1002/art.40748. [Epub ahead of print].
29. Blaschke UK, Eikenberry EF, Hulmes DJ, Galla HJ, Bruckner P. Collagen XI nucleates self-assembly and limits lateral growth of cartilage fibrils. J Biol Chem 2000;275:370–8.
30. Acke FR, Malfait F, Vanakker OM, Steyaert W, De Leeneer K, Mortier G, *et al*. Novel pathogenic COL11A1/COL11A2 variants in Stickler syndrome detected by targeted NGS and exome sequencing. Mol Genet Metab 2014;113:230–5.
31. Chantry A. WWP2 ubiquitin ligase and its isoforms: new biological insight and promising disease targets. Cell Cycle 2011;15:2437–9.
32. Zou W, Chen X, Shim JH, Huang Z, Brady N, Hu D, *et al*. The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. Nat Cell Biol 2011;13:59–65.
33. Nakamura Y, Yamamoto K, He X, Otsuki B, Kim Y, Murao H, *et al*. Wwp2 is essential for palatogenesis mediated by the interaction between Sox9 and mediator subunit. Nat Commun 2011;2:251.
34. Proctor CJ, Smith GR. Computer simulation models as a tool to investigate the role of microRNAs in osteoarthritis. PLoS One 2017;12. e0187568.
35. Boettger LM, Handsaker RE, Zody MC, McCarroll SA. Structural haplotypes and recent evolution of the human 17q21.31 region. Nat Genet 2014;44:881–5.