Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene *PLEC*

A.K. Sorial[†*], I.M.J Hofer[†], M. Tselepi[†], K. Cheung[†], E. Parker[†], D.J. Deehan[‡], S.J. Rice[†], J. Loughlin[†*]

[†]Newcastle University, Biosciences Institute, Newcastle upon Tyne, UK

[‡]Freeman Hospital, Newcastle Upon Tyne, UK

Author email addresses:

Antony Sorial:	tony.sorial@newcastle.ac.uk
Ines Hofer:	ines.hofer@newcastle.ac.uk
Maria Tselepi:	maria.tselepi@newcastle.ac.uk
Kathleen Cheung:	kat.cheung@newcastle.ac.uk
Eleanor Parker:	eleanor.parker@newcastle.ac.uk
David Deehan:	david.deehan@newcastle.ac.uk
Sarah Rice:	sarah.rice@newcastle.ac.uk
John Loughlin:	john.loughlin@newcastle.ac.uk

[*]Address correspondence and reprint requests to:

Mr. Antony K. Sorial or Professor John Loughlin, Newcastle University, Biosciences Institute, International Centre for Life, Newcastle upon Tyne, NE1 3BZ, United Kingdom.

Tel: +44 (0)191 241 8988; Fax: +44 (0)191 241 8666. E-mail: tony.sorial@newcastle.ac.uk (A.K. Sorial), john.loughlin@newcastle.ac.uk (J. Loughlin)

Running headline: Epigenetic analysis of the *PLEC* OA risk locus

1

1   **Objective**

2   Osteoarthritis (OA) associated single nucleotide polymorphism (SNP) rs11780978 correlates

3   with differential expression of *PLEC,* and methylation quantitative trait loci (mQTLs) at *PLEC*

4   CpGs in cartilage. This implies that methylation links chondrocyte genotype and phenotype,

5   thus driving the functional effect. *PLEC* encodes plectin, a cytoskeletal protein that enables

6   tissues to respond to mechanical forces. We sought to assess whether *PLEC* functional effects

7   were cartilage specific.

8

9   **Method**

10   Cartilage, fat pad, synovium and peripheral blood were collected from patients undergoing

11   arthroplasty. *PLEC* CpGs were analysed for mQTLs and allelic expression imbalance (AEI)

12   was performed. We focussed on previously reported mQTL clusters neighbouring cg19405177

13   and cg14598846. Plectin was knocked down in a mesenchymal stem cell (MSC) line using

14   CRISPR/Cas9 and cells phenotyped by RNA-sequencing.

15

16   **Results**

17   Novel mQTLs were discovered in fat pad, synovium and peripheral blood at both clusters. The

18   genotype-methylation effect of rs11780978 was stronger in cg14598846 than in cg19405177

19   and stronger in joint tissues than in peripheral blood. We observed AEI in synovium in the

20   same direction as for cartilage. Knocking-down plectin impacted on pathways reported to have

21   a role in OA, including Wnt signalling, glycosaminoglycan biosynthesis and immune

22   regulation.

23

24   **Conclusions**

25   Synovium is also a target of the rs11780978 OA association functionally operating on *PLEC*.

26   In fat pad, mQTLs were identified but these did not correlate with *PLEC* expression, suggesting

27   the functional effect is not joint-wide. Our study highlights interplay between genetic risk,

28   DNA methylation and gene expression in OA, and reveals clear differences between tissues

29   from the same diseased joint.

30

31   **Keywords:** osteoarthritis, genetic risk, epigenetics, gene expression, *PLEC*, CRISPR/Cas9,

32   cytoskeletal

33

2

Introduction

34

35 Osteoarthritis (OA) is an extremely common musculoskeletal disorder that is characterised by

36 focal loss of articular cartilage leading to chronic pain, disability, co-morbidities including

37 cardiovascular disease, and premature death[1-3]. OA is polygenic and in recent years over 90

38 genome-wide significant loci have been mapped in Europeans using high-density single

39 nucleotide polymorphism (SNP) arrays[4-10]. An overwhelming majority of the associated

40 polymorphisms reside in non-protein-coding regions of the genome and, as such, OA genetic

41 susceptibility is presumed to act principally via changes to gene expression. The availability of

42 excised tissue following arthroplasty of an OA joint offers the opportunity to experimentally

43 test this in disease-relevant cells, including cartilage chondrocytes, the single cell type in this

44 tissue[11]. As such, many OA risk alleles have been shown to correlate with alteration in the

45 expression of nearby genes in chondrocytes[12]. These effects are known as expression

46 quantitative trait loci (eQTLs). DNA methylation changes at CG dinucleotides (CpGs) also

47 regulate gene expression and OA-associated SNPs have been shown to correlate with

48 methylation levels in cartilage DNA in *cis*[13]. These effects are known as methylation QTLs

49 (mQTLs). Furthermore, several OA risk loci correlate with both methylation and expression,

50 with alteration of the former being shown experimentally to modulate the latter[14]. These effects

51 are known as methylation and expression QTLs (meQTLs). Overall, these studies indicate that

52 DNA methylation links chondrocyte genotype and phenotype, with this epigenetic mechanism

53 underpinning and driving a large proportion of the functional impact of OA genetic risk.

54 Furthermore, in most instances, OA mQTLs localize to predicted enhancers[13], placing OA in

55 the "enhanceropathy" category of common diseases[15].

56 We previously reported that an OA association marked by SNP rs11780978 (G>A)

57 correlated with methylation of nine CpGs measured using the Illumina Infinium

58 HumanMethylation450 genome-wide CpG array[16]. The SNP and CpGs are all located within

59 the *PLEC* gene on chromosome 8q24.3. We subsequently focused on two of the most

60 significant CpGs from amongst the nine, cg19405177 and cg14598846, and by using

61 pyrosequencing, we replicated the correlation between rs11780978 genotype and CpG

62 methylation in an independent panel of DNAs. The two pyrosequencing assays used to

63 replicate cg19405177 and cg14598846 captured an additional seven CpGs flanking

64 cg19405177 and an additional three CpGs adjacent to cg14598846. The methylation of these

65 additional CpGs also correlated with rs11780978 genotype. Furthermore, using allelic

66 expression imbalance (AEI) analysis (a complementary approach to standard eQTL analysis[17])

67   we demonstrated that the OA risk-conferring A allele of rs11780978 correlated with reduced

68   *PLEC* expression, and that expression of the gene also correlated with methylation at CpGs in

69   the cg19405177 and cg14598846 clusters[16]. These mQTL, eQTL and meQTL studies were all

70   performed on cartilage DNA and RNA, and highlighted *PLEC* as a target of the OA

71   susceptibility mapping to chromosome 8q24.3.

72   *PLEC* encodes plectin, a large cytoskeletal protein that regulates signalling from the

73   extracellular environment to the cell nucleus[18-19]. Plectin enables cells to respond to external

74   mechanical stimuli and forces, such as those experienced by the chondrocyte during joint

75   movement[20]. Although cartilage loss is central to the OA disease process, there are pathological

76   changes to other tissues of the articulating joint that occur concurrent to and following the loss

77   of cartilage[21]. *PLEC* is expressed widely and in this report, we set out to test whether the

78   mQTL, eQTL and meQTL effects that we had discovered were cartilage specific or whether

79   they were active in other joint or non-joint tissues. We also modelled the transcriptional effect

80   that the OA risk-conferring allele of rs11780978 has on *PLEC* by knocking-down plectin in a

81   mesenchymal stem cell line using CRISPR/Cas9 and assessing the impact on global gene

82   expression.

83

Materials and Methods

OA Patients

86 Tissue samples were collected from patients undergoing knee or hip arthroplasty for primary

87 OA at the Newcastle upon Tyne Foundation Trust hospitals. Ethical approval was granted by

88 the Health Research Authority of the National Health Service (research ethics committee

89 reference 14/NE/1212) and each donor provided verbal and written informed consent. We

90 collected macroscopically normal cartilage (36 samples, distal from the lesion and avoiding

91 areas of fibrillated tissue), infrapatellar fat pad (68 samples) and synovium (81 samples). For

92 some patients, more than one joint tissue type was available for analysis. We also collected

93 whole peripheral blood samples from 55 of the patients just prior to their surgery, using EDTA

94 vacutainers for DNA extraction and Tempus™ tubes for RNA extraction (ThermoFisher

95 Scientific). In total, we analysed tissue samples from 202 OA patients. Further patient details

96 can be found in Supplementary Table 1.

97

98

99

Nucleic acid extraction from tissue samples

100 **Nucleic acid extraction from tissue samples**

101 Joint tissue samples were stored frozen and ground to a powder using a mixermill (Retsch

102 Limited) under liquid nitrogen. For cartilage, RNA was extracted using TRIzol (Life

103 Technologies) and DNA was extracted using the E.Z.N.A. Tissue DNA isolation kit (Omega

104 Biotek, VWR). For fat pad and synovium, DNA and RNA were extracted using the E.Z.N.A.

105 DNA/RNA isolation kit (Omega Biotek, VWR). For blood, DNA was extracted using the

106 QIAamp DNA blood mini kit (Qiagen) and RNA was extracted using the Tempus™ Spin RNA

107 isolation kit (ThermoFisher Scientific).

108

109 **Quantitative gene expression**

110 cDNA synthesis and quantitative PCR (qPCR) were performed as described previously[22].

111 Predesigned TaqMan assays (Integrated DNA Technologies) were used to quantify expression

112 of the housekeeping genes *HPRT1*, *18S* and *GAPDH*. For *PLEC,* expression primers were

113 designed using the Roche probe library system and are located in exons 19 and 20, which are

114 present in all isoforms of the gene (primers 5'-CTGCGTAGGAAATACAGTT-3' and 5'-

115 CAGCTGTTCCTTCTCGTCCT-3'). The relative expression of *PLEC* was calculated by the $2^{-\Delta Ct}$

116 method, where ΔCt is the mean Ct value of the three housekeeping genes subtracted from

117 the Ct value of *PLEC*.

118

119 **mQTL and eQTL analysis**

120 Genotyping of association SNP rs11780978 and *PLEC* transcript SNP rs11783799 (A>G),

121 DNA methylation analysis at the cg19405177 and cg14598846 CpG clusters, and *PLEC* AEI

122 analysis using rs11783799 were all performed as described previously using the same protocols

123 and assays[16]. In the methylation analysis, PCRs were performed in duplicate on each sample

124 and the mean calculated, with samples being excluded from the analysis if the methylation

125 between the PCR replicates differed by >5%. In the AEI analysis, PCRs were performed in

126 triplicate on each tissue cDNA and DNA sample, with samples excluded from the analysis if

127 the values between the PCR replicates differed by >5%.

128

129 **Generation of a mesenchymal stem cell line constitutively expressing Cas9 (MSC-Cas9)**

130 We purchased the immortalised human mesenchymal stem cell (MSC) line SCRC-4000

131 (ATCC) and transduced it with the lentiCRISPR v2 plasmid 52961, which contains a Cas9

132 insert and confers puromycin resistance (Addgene). The cells were expanded in complete

133 growth medium consisting of serum free media (PCS-500-030, ATCC) supplemented with

134   Mesenchymal Stem Cell Growth Kit containing L-Alanyl-L-Glutamine (PCS-500-040,

135   ATCC), G418 disulphate salt solution (G8168, Sigma Aldrich) and penicillin-streptomycin

136   (30-2300, ATCC). Selection of Cas9 expressing cells was achieved by addition of 0.35ug/μl

137   of puromycin to the complete growth media. Constitutive expression of Cas9 was confirmed

138   by immunoblot with a Cas9 antibody (7A9-3A3, Cell Signalling Technology).

139

140   **CRISPR/Cas9**

141   CHOPCHOP (https://chopchop.cbu.uib.no/) was used to design guide RNA (gRNA) sequences

142   that would generate a 26bp frame-shift deletion of exon 3 of *PLEC* (primers 5'-

143   CTGTATGAAGACCTCCGCGA-3' and 5'-CCCCGAGAGGACCTCCAGCA-3'). This exon

144   is present in all isoforms of the gene. The gRNAs were then purchased as synthetic Alt-R®

145   CRISPR-Cas9 crRNAs alongside Alt-R® CRISPR-Cas9 tracrRNA (Integrated DNA

146   Technologies). MSC-Cas9 cells were plated at sub-confluence (45,000 cells/well) in 6-well

147   plates (VWR) and allowed to adhere for 24 hours. For deletions (4 replicates), paired crRNAs

148   were annealed to tracrRNA in a 1:1:2 ratio at 95°C for 5 minutes, whilst for controls (4

149   replicates), tracrRNA alone was used. Deletion and control mixes were separately combined

150   with Dharmafect1 (Dharmacon) in serum free media (PCS-500-030, ATCC) to produce a

151   transfection mix. 200ul of media/well was aspirated and replaced with transfection mix for 24

152   hours, after which the media was changed to complete growth medium and the cells expanded

153   in T182.5 flasks (VWR) in monolayer and until confluent (between 10-14 days), with two

154   changes of media per week. Cells were detached by trypsinization, washed in phosphate-

155   buffered saline (PBS), spun and the cell pellets frozen. DNA, RNA and protein were

156   simultaneously extracted from pellets using the NucleoSpin TriPrep Kit (Macherey-Nagel,

157   supplied by ThermoFisher). Deletion of the target region was confirmed by end-point PCR of

158   DNA using primers flanking the guides (primers 5'-GAGGGGTTGAGCTGGATTCC-3' and

159   5'-CCAGCACCCACTCTGTAGAT-3') and by Sanger sequencing.

160

161   **Immunoblotting**

162   Up to 10μg of total protein was resolved on NuPAGE™ 3-8% Tris-Acetate Protein Gels

163   (Invitrogen) in 1x NuPAGE™ Tris-Acetate SDS Running Buffer (Invitrogen). Protein was

164   transferred through a semi-dry transfer at 200mA/gel for 3 hours. Blots were probed with anti-

165   plectin guinea pig polyclonal serum (GP21, Progen)/goat anti-guinea pig (AB6908, Abcam)

166 and anti-GAPDH (MAB374, Sigma Aldrich)/goat anti-mouse (P0447, Dako) as a loading

167 control.

168

RNA-sequencing

170 For each sample, the TruSeq Stranded mRNA library prep kit (Illumina) was used to prepare a

171 sequencing library from 1.5µg of total RNA. All samples were then sequenced to a depth of

172 ~52-71 million paired end reads on an Illumina NextSeq platform. Quality control of raw

173 sequencing reads was performed using FastQC (v0.11.7). All samples passed QC. Reads were

174 mapped to the hg38 (ENSEMBL version 91) reference genome using Salmon (v0.14.1) in

175 quasi-mapping mode on default settings[23]. All samples except one achieved alignment rates of

176 ~88%. Counts were summarised to gene level and imported into R using tximport[24]. Principle

177 component analysis (PCA) revealed the sample with the low alignment rate to be an outlier

178 and this was removed from further analysis, leaving four *PLEC* deletion and three controls.

179 Unwanted variation in the dataset was estimated using surrogate variable analysis (svaseq[25]).

180 The number of surrogate variables was estimated at two and these were included into the model

181 for differential expression analysis. Differential expression analysis between deletion and

182 control samples was performed using the Wald test implemented in DESeq2[26]. The false

183 discovery rate (FDR) was controlled using the Benjamini-Hochberg procedure. Genes with an

184 FDR $P<0.05$ were considered significant. Gene set enrichment analysis (GSEA[27]) was used to

185 discover overrepresented pathways in the set of up- and down-regulated genes. The full gene

186 list was ranked by signed $\log10(P$ value); positive sign if up-regulated and negative if down-

187 regulated. This pre-ranked list was used as input into the fgsea package in Bioconductor to

188 query enrichment of Reactome pathways[28]. Pathways with an FDR $P<0.25$ were considered

189 significant, as recommended by the developers of GSEA[27]. To visualise splice junctions, reads

190 were aligned to hg38 reference genome using STAR (v2.4.2a) short read aligner using default

191 settings[29] and viewed in Integrative Genomics Viewer (IGV[30]).

192

Statistical analyses

194 When correlating association between rs11780978 genotype and methylation of CpGs, *P*

195 values were calculated using the Kruskal-Wallis test. For comparison of global methylation

196 between tissues, Kruskal-Wallis with Dunn's multiple comparisons was used. For quantitative

197 *PLEC* expression analyses, *P* values were calculated using a Mann-Whitney 2-tailed exact test.

198 For *PLEC* AEI, *P* values were calculated using Wilcoxon's matched pairs signed rank test.

Results

PLEC cartilage mQTLs are present in other tissues

We previously discovered *PLEC* mQTLs in cartilage DNA[16]. For cg19405177 and its seven flanking CpGs (located in a 112bp interval), the OA risk-conferring A allele of rs11780978 correlated with increased methylation, whereas for cg14598846 and its three adjacent CpGs (located in a 23bp interval), the A allele correlated with decreased methylation. In this report, we first assessed if these mQTLs were also active in DNA from fat pad, synovium and blood. We also included a new cohort of cartilage DNAs. Figure 1 provides the data for cg19405177 and cg14598846, whilst the data for all 12 CpGs across the four tissues can be found in Supplementary Figures 1 (cartilage), 2 (fat pad), 3 (synovium) and 4 (blood). The mQTLs were replicated in the new batch of cartilage DNAs and in the same directions as before. The mQTLs were also present in the other tissues and also in the same direction.

When we plotted the methylation data by tissue type but independent of genotype, there were significant differences in methylation levels (Supplementary Figure 5). For example, cg19405177 demonstrated higher methylation values in fat pad and synovium compared to cartilage and blood, whilst cg14598846 demonstrated higher methylation values in blood compared to joint tissues. To emphasize this variability in CpG methylation in tissues from the same patient, we plotted the cg19405177 and cg14598846 data for those patients in whom we had studied two or more tissues. There were 18 such patients with data generated at cg19405177 and 20 patients with data generated at cg14598846 (Figure 2). The variability in methylation between the tissues of an individual is particularly clear for patient 6550 at cg19405177, and patient 6599 at cg14598846, with a methylation range of >30% across the four tissues tested for each CpG.

We next constructed heat maps of the percentage effect of rs11780978 genotype on methylation at each CpG in each tissue. For cartilage, we included data from the replication study of our previous report[16] to maximise power. For the cg19405177 CpGs, the average genotypic effect for all eight CpGs combined was 31.3% in cartilage, 17.4% in fat pad, 21.2 % in synovium and 9.3% in blood (Figure 3A). For the cg14598846 CpGs, the average genotypic effect for all four CpGs combined were comparable across the joint tissues, with values of 85.0% in cartilage, 81.4% in fat pad and 87.5% in synovium (Figure 3B). Although the genotypic effect in blood was lower, it was still high at 43.5%. At the individual CpG level, there were some striking differences between tissues. For example, cg19405177 had a

231   genotypic effect of >30% in cartilage, synovium and blood but only 3.9% in fat pad. These

232   large differences imply intra-tissue differences in the nature of the mQTLs.

233   Overall, this analysis highlighted that the previously reported *PLEC* mQTLs are active

234   in non-cartilaginous tissues and that the effects are stronger in joint tissues than blood.

235   Although the genotypic effect of rs11780978 is greater in cartilage for the cg19405177 CpGs,

236   this is not the case for the cg14598846 CpGs.

237

238   **The *PLEC* rs11780978 eQTL is active in synovium**

239   Before testing for the *PLEC* rs11780978 eQTL by AEI analysis, we measured the relative

240   expression of the gene in fat pad, synovium and blood along with cartilage as a comparator

241   (Supplementary Figure 6). The qPCR data revealed that *PLEC* expression was equivalent

242   between cartilage and synovium, was higher in fat pad and lower in blood.

243   We performed AEI as described previously[16]. In summary, we identified rs11780978

244   heterozygotes who were also heterozygous for the *PLEC* transcript SNP rs11783799 (r^2=0.93;

245   the OA risk conferring A allele of rs11780978 correlates with the G allele of rs11783799). We

246   then used rs11783799 to quantify the relative allelic expression of *PLEC* in the compound

247   heterozygotes. The blood cDNA samples consistently failed quality control, with the PCR

248   replicates differing by >5% in all samples. This is likely a reflection of the lower expression of

249   *PLEC* in this tissue and we could not therefore generate blood AEI data. We were able to

250   generate AEI data for 19 fat pad and 20 synovium compound heterozygotes (Figure 4). AEI

251   was detectable in both tissues but only in synovium did the direction of AEI significantly

252   correlate with an allele of rs11783799, with reduced expression of *PLEC* correlating with the

253   G allele of the SNP in the combined analysis (*P*=0.0054; Figure 4B). In fat pad, AEI was

254   observed in a majority of the patients, but this did not correlate with a particular allele (*P*=0.31;

255   Figure 4A).

256   To emphasize the variability of AEI in tissues from the same patient, we next plotted

257   the data for those patients in whom we had studied fat pad and synovium (Supplementary

258   Figure 7). There were six such patients with clear variability in the G/A ratio between the two

259   tissues tested for most of these, with patients 5680 and 5725 demonstrating particularly clear

260   inter-tissue variability in the ratio; in both, the G allele is more highly expressed in fat pad but

261   more lowly expressed in synovium relative to the A allele.

262   The significant AEI in synovium was in the same direction that we had observed in

263   cartilage[16], with the G allele of rs11783799, which marks the OA risk conferring A allele of

264   rs11780978, demonstrating reduced *PLEC* expression. In that previous report, we had studied

9

265   19 cartilage compound heterozygotes and the mean G/A ratio was 0.89. The mean synovium

266   G/A ratio we detected here was 0.92.

267   We next plotted the *PLEC* AEI data for the fat pad and synovium compound

268   heterozygotes against their individual methylation data. In fat pad, there was no evidence of a

269   significant correlation between methylation and AEI at cg19405177, cg14598846 or the

270   additional 10 CpGs measured (all *P* values >0.05; Supplementary Figure 8). However, in

271   synovium, three of the CpGs correlated with AEI; CpG3 from the cg19405177 cluster

272   (*P*=0.049), and cg14598846 (*P*=0.039) and CpG2 (*P*=0.047) from the cg14598846 cluster

273   (Supplementary Figure 9). These three CpGs had also shown meQTL activity in our previous

274   cartilage study[16]. For CpG3 from the cg19405177 cluster, the same direction of effect is seen

275   in cartilage and synovium, with increased methylation correlating with decreased AEI (the $\log_2$

276   AEI ratio moves towards zero with increasing methylation). However, for cg14598846 and

277   CpG2 from the cg14598846 cluster, the effects are in opposite directions between cartilage and

278   synovium; in cartilage, increased methylation correlated with decreased AEI whereas in

279   synovium, increased methylation correlated with increased AEI (the $\log_2$ AEI ratio moves

280   away from zero with increasing methylation).

281   In summary, the *PLEC* rs11780978 eQTL previously observed in cartilage also

282   operates in synovium. Furthermore, we observed correlations between *PLEC* expression and

283   CpG methylation in synovium. These meQTLs are shared with cartilage but are not necessarily

284   operating in the same manner.

285

Knock-down of *PLEC* and plectin elicits a response

287   Our previous data and the results generated here highlight that the OA risk conferring A allele

288   of rs11780978 correlates with decreased expression of *PLEC* in cartilage and synovium. We

289   modelled this effect by knocking-down plectin. We used an immortalised human MSC cell line

290   that is amenable to CRISPR-based editing[31] and transduced it to constitutively express Cas9.

291   We then generated a 26bp frame-shift deletion of *PLEC* exon 3, with four replicates for deletion

292   and four control replicates. The cells were cultured in monolayer for up to 14 days and the

293   deletion was confirmed by PCR and Sanger sequencing, whilst a near-total plectin knock-out

294   was observed by immunoblotting using an antibody that binds to the C-terminus of the protein

295   (Figure 5A-C). The isolated RNA was then subjected to next-generation mRNA sequencing.

296   This revealed alternate splicing occurring in *PLEC* in the deletion samples, with exon 3 being

297   completely skipped in a subset of transcripts (Figure 5D). No exon skipping was seen in the

298   control samples. The level of *PLEC* mRNA was 26% lower in the deletion samples compared

299   to the controls (log$_2$ fold change of -0.44; FDR P=0.001; Supplementary Table 2). In summary,

300   this data implies that the Cas9-mediated deletion impacted on the transcription of *PLEC* and

301   on the translation of *PLEC* mRNA into plectin.

302   We next assessed the effect of this plectin knock-down on the transcriptome. In addition

303   to *PLEC*, the deletion had a significant effect (FDR P<0.05) on the expression of 269 genes

304   (Figure 6A and B, Supplementary Table 2). Pathway analysis revealed upregulation of

305   processes involving immune regulation and protein post-translation modification (Figure 6C),

306   and downregulation of Wnt signalling, metabolic pathways and glycosaminoglycan

307   biosynthesis, amongst others (Figure 6D). In conclusion, this model system highlighted that

308   depletion of *PLEC* and plectin impacted upon a range of pathways reported to play a critical

309   role in cartilage biology and OA.

310

Discussion

312   The mapping of risk loci for polygenic traits is now a relatively straightforward procedure,

313   with the next major step in complex trait analysis being the transition from association signal

314   to functional characterization[32]. For OA, considerable strides have been made in this regard in

315   recent years[33-35]. Our interest is in the epigenetic dimension of OA genetic risk and especially

316   the role of DNA methylation. This is the best characterized and most stable epigenetic

317   modification and modulates gene expression by disrupting the binding of transcription factors

318   and by facilitating the binding of methyl-binding proteins, which initiate compaction of

319   chromatin and gene silencing[36]. Last year, we demonstrated that the OA association signal

320   marked by SNP rs11780978 correlated with methylation of CpGs within *PLEC* and with

321   variable expression of the gene[16]. These analyses were chondro-centric and here, we set out to

322   determine whether these functional effects were more widespread. If so, an exclusive focus on

323   cartilage as a target tissue of the rs11780978 association signal would not be merited.

324   We discovered that the rs11780978-*PLEC* eQTLs, mQTLs and meQTLs that are active

325   in cartilage are also active in synovium, but with some differences; the genotypic effect is

326   substantially greater in cartilage for the cg19405177 cluster of CpGs, whilst the cg14598846

327   meQTLs act in opposite directions between the two tissues. In fat pad, we observed the mQTLs

328   but *PLEC* AEI did not correlate with rs11780978 genotype. Instead, the OA risk-conferring A

329   allele of rs11780978 demonstrated increased and decreased expression relative to the non-risk

330   allele across the 19 fat pad samples studied, with no overall significant difference when the

331   data was combined. There are therefore eQTLs operating on *PLEC* in fat pad, but these are

332   unrelated to rs11780978. These distinctions point toward regulatory variability in gene

333   expression mediated by the same association between joint tissues. In blood, the mQTLs were

334   detectable but their genotypic effect was much weaker in comparison to the three joint tissues.

335   Such variable strengths of mQTLs effects between cells from different lineages are expected[37].

336   We were unable to test blood for *PLEC* eQTLs due to the low expression level of the gene. In

337   summary, we observed the *PLEC* mQTLs across all four tissues tested with the joint tissues

338   more similar to each other than to blood. This was exemplified by the mQTL genotypic effects

339   being much stronger in joint tissues. Since synovium was subject to eQTLs and meQTLs, it

340   too should be investigated as a target of the rs11780978 association signal.

341   Finally, we modelled the effect of the rs11780978 risk-conferring allele by knocking

342   down plectin in a cell line. *PLEC* mRNA was detectable in the RNA-sequencing data but at a

343   significantly lower level compared to control. The knock-down was very efficient at the protein

344   level, with a near total absence of plectin observed. The knock-down impacted on a range of

345   cellular pathways. Of particular note, both the innate and the acquired immune response

346   showed an upregulation, an effect that has been reported on previously in OA for both cartilage

347   and synovium[38-39]. Downregulation of Wnt signalling was also of interest, as this pathway is

348   critical to cartilage homeostasis[33,40]. Furthermore, there is growing evidence of an interplay

349   between Wnt signalling and inflammation in joint tissues[41]. Our model system did not include

350   mechanical load, a force which plectin allows cells to respond to[18-20]. Aberrant mechanical

351   load, including only moderate change, can elicit an inflammatory response that has been termed

352   "mechanoflammation"[42]. Our data suggests that even in the absence of load, the loss of plectin

353   is eliciting this response. A future model incorporating load may therefore reveal yet more

354   insight into a potential key role of plectin as a regulator of inflammation in response to

355   mechanical load and joint tissue injury in OA.

356   In conclusion, our investigations have highlighted the interplay between OA genetic

357   risk, DNA methylation and gene expression, and reveal clear differences in these effects for

358   tissues from the same diseased joint. Furthermore, by recapitulating the effect of the risk locus

359   at *PLEC*, we have linked OA genetic susceptibility with a number of cellular pathways, thus

360   providing mechanistic insight into OA aetiology.

361

362

363

364

365

References

366

367   1. Glyn-Jones S, Palmer AJ, Agricola R, Price AJ, Vincent TL, Weinans H, et al.
368   Osteoarthritis. Lancet 2015; 386: 376-387.

369   2. Kendzerska T, Juni P, King LK, Croxford R, Stanaitis I, Hawker GA. The longitudinal
370   relationship between hand, hip and knee osteoarthritis and cardiovascular events: a
371   population-based cohort study. Osteoarthritis Cartilage 2017; 25: 1771-1780.

372   3. Hawker GA, Croxford R, Bierman AS, Harvey PJ, Ravi B, Stanaitis I, et al. All-cause
373   mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a
374   population based cohort study. PLoS One 2014; 9: e91286.

375   4. arcOGEN Consortium, arcOGEN Collaborators, Zeggini E, Panoutsopoulou K, Southam
376   L, Rayner NW, et al. Identification of new susceptibility loci for osteoarthritis (arcOGEN):
377   a genome-wide association study. Lancet 2012; 380: 815-823.

378   5. Reynard LN. Analysis of genetics and DNA methylation in osteoarthritis: What have we
379   learnt about the disease? Semin Cell Dev Biol 2017; 62: 57-66.

380   6. Styrkarsdottir U, Helgason H, Sigurdsson A, Norddahl GL, Agustsdottir AB, Reynard LN,
381   et al. Whole-genome sequencing identifies rare genotypes in COMP and CHADL
382   associated with high risk of hip osteoarthritis. Nat Genet 2017; 49: 801-805.

383   7. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, et al.
384   Genome-wide analyses using UK Biobank data provide insights into the genetic
385   architecture of osteoarthritis. Nat Genet 2018; 50: 549-558.

386   8. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stefansson OA, Sigurdsson JK, et al.
387   Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11,
388   COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet 2018; 50: 1681-
389   1687.

390   9. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J,

391   et al. Identification of new therapeutic targets for osteoarthritis through genome-wide

392   analyses of UK Biobank data. Nat Genet 2019; 51: 230-236.

393   10. Styrkarsdottir U, Stefansson OA, Gunnarsdottir K, Thorleifsson G, Lund SH, Stefansdottir

394   L, et al. GWAS of bone size yields twelve loci that also affect height, BMD, osteoarthritis

395   or fractures. Nat Commun 2019; 10: 2054.

396   11. den Hollander W, Pulyakhina I, Boer C, Bomer N, van der Breggen R, Arindrarto W, et al.

397   Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue:

398   Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. Arthritis Rheumatol

399   2019; 71: 561-570.

400   12. Shepherd C, Reese AE, Reynard LN, Loughlin J. Expression analysis of the osteoarthritis

401   genetic susceptibility mapping to the matrix Gla protein gene MGP. Arthritis Res Ther

402   2019; 21: 149.

403   13. Rice SJ, Cheung K, Reynard LN, Loughlin J. Discovery and analysis of methylation

404   quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals.

405   Osteoarthritis Cartilage 2019; 27: 1545-1556.

406   14. Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression

407   of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis

408   susceptibility SNP rs143383. Hum Genet 2014; 133: 1059-1073.

409   15. Smith E, Shilatifard A. Enhancer biology and enhanceropathies. Nat Struct Mol Biol 2014;

410   21: 210-219.

411   16. Rice SJ, Tselepi M, Sorial AK, Aubourg G, Shepherd C, Almarza D, et al. Prioritization of

412   PLEC and GRINA as Osteoarthritis Risk Genes Through the Identification and

413   Characterization of Novel Methylation Quantitative Trait Loci. Arthritis Rheumatol 2019;

414   71: 1285-1296.

415   17. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. Nat

416   Rev Genet 2010; 11: 533-538.

417   18. Castanon MJ, Walko G, Winter L, Wiche G. Plectin-intermediate filament partnership in

418   skin, skeletal muscle, and peripheral nerve. Histochem Cell Biol 2013; 140: 33-53.

419   19. Rezniczek GA, Winter L, Walko G, Wiche G. Functional and Genetic Analysis of Plectin

420   in Skin and Muscle. Methods Enzymol 2016; 569: 235-259.

421   20. Goldmann WH. Intermediate filaments and cellular mechanics. Cell Biol Int 2018; 42: 132-

422   138.

423   21. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint

424   as an organ. Arthritis Rheum 2012; 64: 1697-1707.

425   22. Gee F, Rushton MD, Loughlin J, Reynard LN. Correlation of the osteoarthritis

426   susceptibility variants that map to chromosome 20q13 with an expression quantitative trait

427   locus operating on NCOA3 and with functional variation at the polymorphism

428   rs116855380. Arthritis Rheumatol 2015; 67: 2923-2932.

429   23. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-

430   aware quantification of transcript expression. Nat Methods 2017; 14: 417-419.

431   24. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level

432   estimates improve gene-level inferences. F1000Res 2015; 4: 1521.

433   25. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data.

434   Nucleic Acids Res 2014; 42.

435   26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for

436   RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.

437   27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene

438   set enrichment analysis: a knowledge-based approach for interpreting genome-wide

439   expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.

440   28. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database

441   of reactions, pathways and biological processes. Nucleic Acids Res 2011; 39: D691-697.

442   29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast

443   universal RNA-seq aligner. Bioinformatics 2013; 29: 15-21.

444   30. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-

445   performance genomics data visualization and exploration. Brief Bioinform 2013; 14: 178-

446   192.

447   31. Farhang N, Brunger JM, Stover JD, Thakore PI, Lawrence B, Guilak F, et al. (*) CRISPR-

448   Based Epigenome Editing of Cytokine Receptors for the Promotion of Cell Survival and

449   Tissue Deposition in Inflammatory Environments. Tissue Eng Part A 2017; 23: 738-749.

450   32. Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: From Association to Function. Am

451   J Hum Genet 2018; 102: 717-730.

452   33. Monteagudo S, Cornelis FMF, Aznar-Lopez C, Yibmantasiri P, Guns LA, Carmeliet P, et

453   al. DOT1L safeguards cartilage homeostasis and protects against osteoarthritis. Nat

454   Commun 2017; 8: 15889.

455   34. Shepherd C, Zhu D, Skelton AJ, Combe J, Threadgold H, Zhu L, et al. Functional

456   Characterization of the Osteoarthritis Genetic Risk Residing at ALDH1A2 Identifies

457   rs12915901 as a Key Target Variant. Arthritis Rheumatol 2018; 70: 1577-1587.

458   35. Klein JC, Keith A, Rice SJ, Shepherd C, Agarwal V, Loughlin J, et al. Functional testing

459   of thousands of osteoarthritis-associated variants for regulatory activity. Nat Commun

460   2019; 10: 2434.

461   36. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation.

462   Nat Rev Genet 2016; 17: 551-565.

463   37. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, et al. Methylation

464   quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental

465   stage, and tissue type. BMC Genomics 2014; 15: 145.

466   38. Berenbaum F, van den Berg WB. Inflammation in osteoarthritis: changing views.

467   Osteoarthritis Cartilage 2015; 23: 1823-1824.

468   39. Miller RJ, Malfait AM, Miller RE. The innate immune response as a mediator of

469   osteoarthritis pain. Osteoarthritis Cartilage 2019.

470   40. Monteagudo S, Lories RJ. Cushioning the cartilage: a canonical Wnt restricting matter. Nat

471   Rev Rheumatol 2017; 13: 670-681.

472   41. Claudel M, Jouzeau JY, Cailotto F. Secreted Frizzled-related proteins (sFRPs) in osteo-

473   articular diseases: much more than simple antagonists of Wnt signaling? FEBS J 2019;

474   286: 4832-4851.

475   42. Vincent TL. Mechanoflammation in osteoarthritis pathogenesis. Semin Arthritis Rheum

476   2019; 49: S36-S38.

477

478   **Author Contributions**
479   AKS and JL designed the study. AKS, IMJH, MT and SJR generated the experimental data.
480   AKS, IMJH, KC and EP analysed the data and prepared the manuscript figures. AKS, IMH,
481   and JL prepared the main manuscript text. DJD provided patient samples for analyses. All
482   authors were involved in drafting the article or revising it critically for intellectual content, and
483   all authors approved the final version to be published.

484

496

Competing interest

498   The authors declare that they have no competing interests.

499

Data Availability

501   All data are available from the corresponding authors upon request. Raw RNA-seq data has

502   been deposited into the NCBI GEO datasets repository with accession number GSE143725.

503

Legends to figures

505   **Figure 1.** Association between rs11780978 genotype and methylation at CpGs cg19405177

506   and cg14598846 in cartilage, fat pad, synovium and blood DNA. *P* values were calculated

507   using the Kruskal-Wallis test. Horizontal lines and error bars show the mean ± SEM.

508   n = the number of patients providing data per CpG site for each tissue.

509

510   **Figure 2.** Differences in methylation within tissues of the same patient. Methylation at CpGs

511   cg19405177 (**A**) and cg14598846 (**B**) were plotted for those individuals in whom methylation

512   data was generated for two or more tissues. Numbers on the x-axis refer to the anonymised

513   identification number assigned to patients at recruitment. Data is grouped by genotype at

514   rs11780978.

515

516   **Figure 3.** Heat maps displaying the percentage effect of genotype at rs11780978 upon

517   methylation at CpG clusters cg19405177 (**A**) and cg14598846 (**B**) for the four patient tissues.

518   The correlation between genotype and methylation was determined using adjusted r^2 values

519   calculated using a standard least squares linear regression model and expressed as a percentage.

520   Data analysis was performed using SAS JMP Statistical Data Visualization Software and heat

521   maps produced in PRISM. Patients contributing data to the cg19405177 cluster 1: cartilage,

522   n=81; fat pad, n=39; synovium, n=64; blood, n=43. Patients contributing data to the

523   cg14598846 cluster: cartilage, n=103; fat pad, n=68; synovium, n=79; blood, n=54.

524

525 **Figure 4.** *PLEC* allelic expression imbalance (AEI) in fat-pad (**A**) and synovium (**B**). AEI
526 analysis was conducted using transcript SNP rs11783799. The risk/non-risk allelic ratio is
527 plotted, with a ratio <1 indicating decreased *PLEC* expression from the risk allele. For each
528 patient, the mean of the DNA ratio (black, n=3 technical repeats) and the mean of the cDNA
529 ratio (grey, n=3 technical repeats) is plotted. Numbers on the x-axis refer to the anonymised
530 identification number assigned to patients at recruitment. The values for DNA and cDNA for
531 all patients combined are represented as box plots to the right, in which the lines within the box
532 represent the median, the box represents the 25th to 75th percentile, and the whiskers represent
533 the maximum and minimum values. *P* values were calculated using Wilcoxon's matched pairs
534 signed rank test.

535

536 **Figure 5.** Knock-down of plectin using CRISPR/Cas9. (**A**) The 26bp sequence of *PLEC* exon
537 3 deleted. (**B**) Representative agarose gel image showing the PCR product for control (non-
538 deleted) and deletion. (**C**) Representative immunoblot comparing plectin levels in control
539 versus deletion, with GAPDH as a loading control. (**D**) Sashimi plot showing splicing of
540 merged *PLEC* deletion (red) and control (blue) samples. Data were viewed on IGV genome
541 browser. The multiple isoforms of *PLEC* are shown on the bottom of the panel. Read coverages
542 show the presence of reads within exons. The lines illustrate the splicing between exons. The
543 values on the lines show the number of junction spanning reads. The merged deletion track
544 shows the absence of the 26bp region deleted within *PLEC* exon 3 (indicated by the box).

545

546 **Figure 6.** Differentially expressed genes and pathways. (**A**) Volcano plot illustrating
547 significantly differentially expressed genes between *PLEC* deletion and control samples. The
548 x-axis shows the $\log_2$ fold change between condition and the y-axis shows the $-\log_{10}$(FDR).
549 Genes with an FDR *P*<0.05 are shown in red. (**B**) Heatmap of significantly differentially
550 expressed genes between *PLEC* deletion and control. Normalised counts were used for plotting.
551 Hierarchical clustering was used to cluster samples and genes. The heatmap shows that samples
552 cluster by condition (D0-Del, deletions; D0-Neg, controls). (**C**) Top 25 significantly
553 upregulated Reactome pathways found using GSEA. (**D**) Top 25 significantly downregulated
554 Reactome pathways found using GSEA. All genes were ranked using signed $\log_{10}$(*P*-value)
555 and this pre-ranked list used as input into GSEA to query enrichment of Reactome pathways.

556

557 **Supplementary Figure 1.** Association between rs11780978 genotype and methylation at CpG
558 clusters cg19405177 (**A**) and cg14598846 (**B**) in cartilage DNA. *P* values were calculated using

19

559   the Kruskal-Wallis test. Horizontal lines and error bars show the mean ± SEM. n = the number

560   of patients providing data per CpG site.

561

562   **Supplementary Figure 2.** Association between rs11780978 genotype and methylation at CpG

563   clusters cg19405177 (**A**) and cg14598846 (**B**) in fat pad DNA. *P* values were calculated using

564   the Kruskal-Wallis test. Horizontal lines and error bars show the mean ± SEM. n = the number

565   of patients providing data per CpG site.

566

567   **Supplementary Figure 3.** Association between rs11780978 genotype and methylation at CpG

568   clusters cg19405177 (**A**) and cg14598846 (**B**) in synovium DNA. *P* values were calculated

569   using the Kruskal-Wallis test. Horizontal lines and error bars show the mean ± SEM. n = the

570   number of patients providing data per CpG site.

571

572   **Supplementary Figure 4.** Association between rs11780978 genotype and methylation at CpG

573   clusters cg19405177 (**A**) and cg14598846 (**B**) in blood DNA. *P* values were calculated using

574   the Kruskal-Wallis test. Horizontal lines and error bars show the mean ± SEM. n = the number

575   of patients providing data per CpG site.

576

577   **Supplementary Figure 5.** Methylation at CpG clusters cg19405177 (**A**) and cg14598846 (**B**)

578   for the four patient tissues without stratification by rs11780978 genotype. *P* values were

579   calculated using Kruskal-Wallis with Dunn's multiple comparisons. Horizontal lines and error

580   bars show the mean ± SEM. n = the number of patients providing data per tissue and at each

581   CpG site. $*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$ $****P \leq 0.0001$.

582

583   **Supplementary Figure 6.** Expression of *PLEC* in tissue samples from OA patients. *PLEC*

584   mRNA levels were measured by qPCR in cartilage (n=10), fat pad (n=9), synovium (n=10)

585   and blood (n=10). Horizontal lines and error bars show the mean ± SEM. *P* values were

586   calculated using a Mann-Whitney 2-tailed exact test. $**P \leq 0.01$; $***P \leq 0.001$

587

588   **Supplementary Figure 7.** The six patients with AEI data for fat pad (FP) and synovium

589   (SYN). The risk/non-risk allelic ratio is plotted, with a ratio <1 indicating decreased *PLEC*

590   expression from the risk allele. For each patient, the mean of the DNA ratio (black, n=3

591   technical repeats) and the mean of the cDNA ratio (grey, n=3 technical repeats) is plotted.

592   Numbers on the x-axis refer to the anonymised identification number assigned to patients at

593   recruitment.

594

595   **Supplementary Figure 8.** Methylation expression quantitative trait locus analyses in fat pad.

596   *PLEC* $\log_2$ allelic expression imbalance (AEI) ratios were plotted against methylation at all

597   CpGs in the cg19405177 (**A**) and cg14598846 (**B**) clusters. The square of the correlation

598   coefficient (r^2) and *P* values were calculated by linear regression analysis using a standard least

599   squares model.

600

601   **Supplementary Figure 9.** Methylation expression quantitative trait locus analyses in

602   synovium. *PLEC* $\log_2$ allelic expression imbalance (AEI) ratios were plotted against

603   methylation at all CpGs in the cg19405177 (**A**) and cg14598846 (**B**) clusters. The square of

604   the correlation coefficient (r^2) and *P* values were calculated by linear regression analysis using
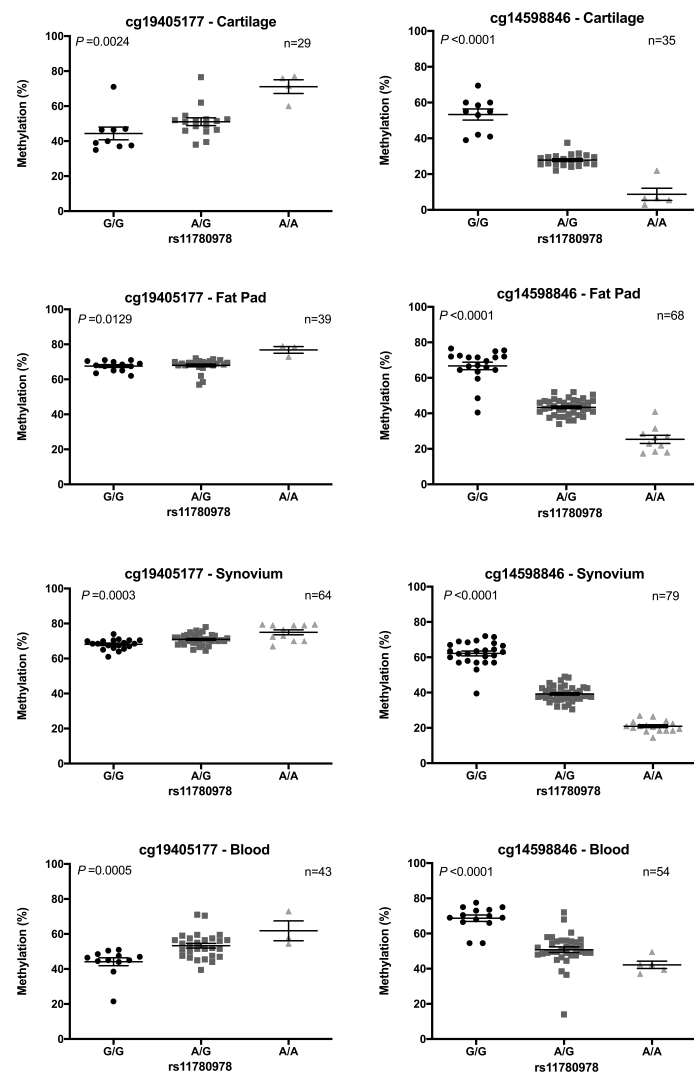
605   a standard least squares model.
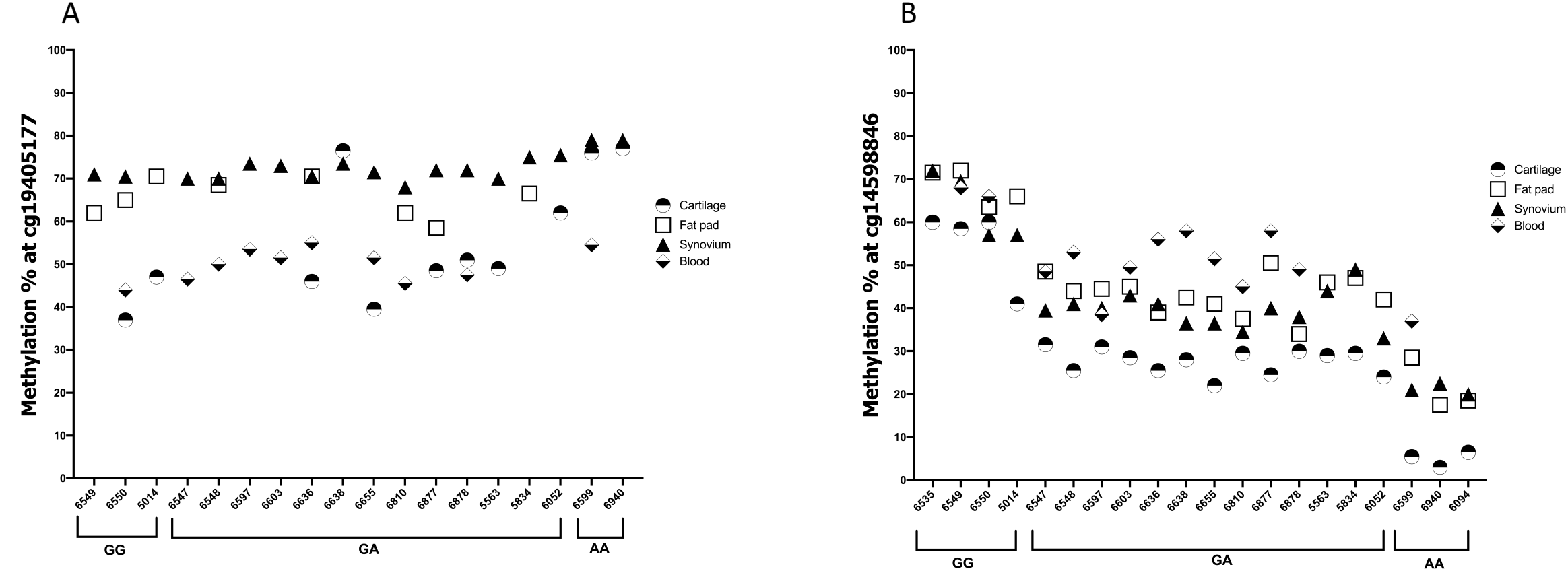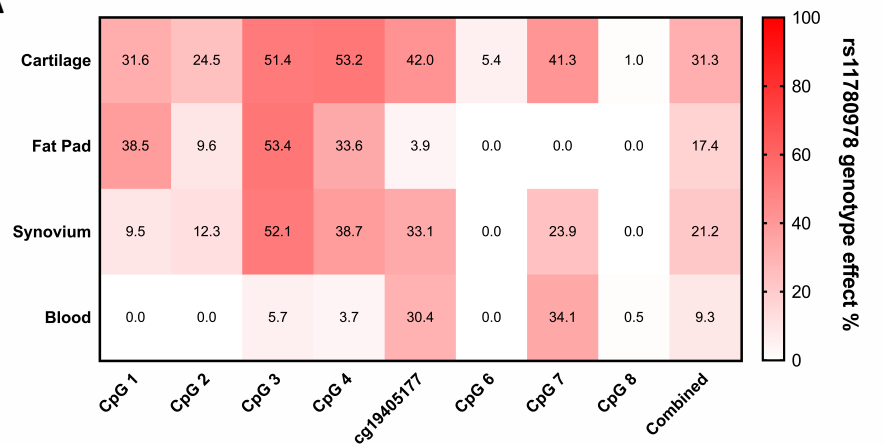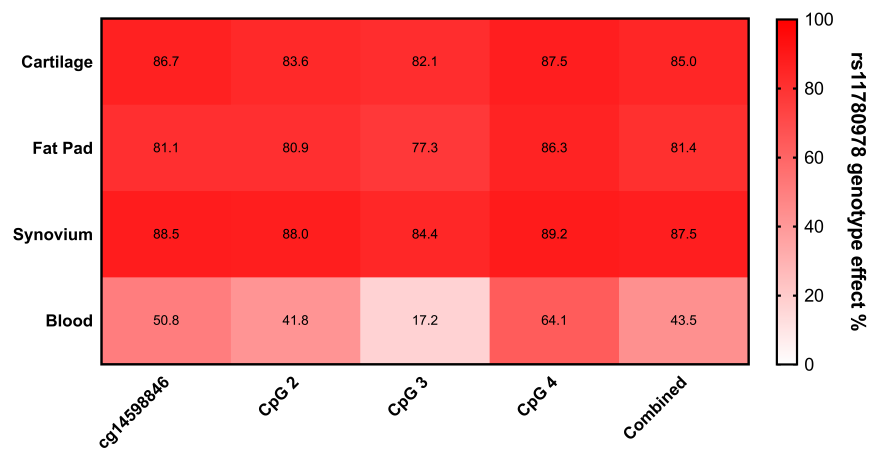
Figure 1

Figure 2

Figure 3

A

	CpG 1	CpG 2	CpG 3	CpG 4	cg19405177	CpG 6	CpG 7	CpG 8	Combined
Cartilage	31.6	24.5	51.4	53.2	42.0	5.4	41.3	1.0	31.3
Fat Pad	38.5	9.6	53.4	33.6	3.9	0.0	0.0	0.0	17.4
Synovium	9.5	12.3	52.1	38.7	33.1	0.0	23.9	0.0	21.2
Blood	0.0	0.0	5.7	3.7	30.4	0.0	34.1	0.5	9.3

rs11780978 genotype effect %

B

	cg14598846	CpG 2	CpG 3	CpG 4	Combined
Cartilage	86.7	83.6	82.1	87.5	85.0
Fat Pad	81.1	80.9	77.3	86.3	81.4
Synovium	88.5	88.0	84.4	89.2	87.5
Blood	50.8	41.8	17.2	64.1	43.5

rs11780978 genotype effect %

Figure 4

Figure 5

A

26bp deletion

5' CTCCAGCAGGGAGATGAGGTTGTGGCCATCGCGGAG 3'   Control
5' CTCCA   CGGAG 3'   Deletion

B

Control   Deletion

DNA

C

Control   Deletion

Plectin

GAPDH

Protein

D

Exon 3

Del_merged_hg38.bam

Deletion

Neg_merged_hg38.bam

Control

143936024   143937210   143938396   143939583

PLEC

Figure 6

A

FDR < 0.05
- No
- Yes

B

C

Upregulated Pathways

PLEC deletion vs control

REACTOME Pathways

D

Downregulated Pathways