Genome-wide association and functional studies identify a role for matrix-Gla protein in osteoarthritis of the hand

Wouter den Hollander[1¶], Cindy G. Boer[2¶], Debbie Hart[3], Michelle S. Yau[4,5], Yolande F.M. Ramos[1], Sarah Metrustry[3], Linda Broer[2], Joris Deelen[1,6], L. Adrienne Cupples[7], Fernando Rivadeneira[2], Margreet Kloppenburg[8], Marjolein Peters[2], Tim D. Spector[3], Albert Hofman[9,10], P. Eline Slagboom[1], Rob G.H.H. Nelissen[11], André G. Uitterlinden[2,9], David T. Felson[5,12] , Anna M. Valdes[13], Ingrid Meulenbelt[1&*], Joyce J.B. van Meurs[2&*]

[1]Department of Medical Statistics and Bioinformatics, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[2]Department of Internal Medicine, Genetic Laboratory, Erasmus Medical Center, Rotterdam, the Netherlands

[3]Department of Twin Research and Genetic Epidemiology, King's College London, London, , UK

[4]Institute for Aging Research, Hebrew SeniorLife, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

 [5]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, MA, USA

 [6]Max Planck Institute for Biology of Ageing, Cologne, Germany

[7]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; The Framingham Heart Study, Framingham, MA, USA

[8]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

[9]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

[10]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[11]Department of Orthopedics, Leiden University Medical Center, Leiden, The Netherlands

[12]Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester, UK

[13] School of Medicine, University of Nottingham, Nottingham, UK

[¶]These authors contributed equally to this work

[&]These authors jointly supervised this work

[*]Corresponding authors

j.vanmeurs@erasmusmc.nl (JBJM)

i.meulenbelt@lumc.nl (IM)

Abstract

Objective Osteoarthritis (OA) is the most common form of arthritis and the leading cause of disability in the elderly. Of all the joints, genetic predisposition is strongest for OA of the hand, however only few genetic risk loci for hand OA have been identified. Our aim was to identify novel genes associated with hand OA and examine the underlying mechanism.

Methods We performed a genome-wide association study of a quantitative measure of hand OA in 12,784 individuals (discovery:8,743, replication:4,011). Genome-wide significant signals were followed up by analysing gene and allele specific expression in a RNA-sequencing dataset (n=96) of human articular cartilage

Results We found two significantly associated loci in the discovery set: at chr12 (P=3.5×10^{-10}) near the MGP gene and at chr12 (P=6.1×10^{-9})near the CCDC91 gene. The DNA variant near the MGP-gene was validated in three additional studies, which resulted in a highly significant association between the MGP-variant and hand OA (rs4764133, Beta$_{meta}$=0.83, P$_{meta}$=$1.8*10^{-15}$). This variant is high linkage disequilibrium with a coding variant in *Matrix Gla-Protein (MGP),* a vitamin K-dependent inhibitor of cartilage calcification. Using RNA-sequencing data from human primary cartilage tissue (n=96), we observed that the hand OA MGP-risk allele was significantly lower expressed compared to the reference allele (40.7%, P<$5*10^{-16}$).

Conclusions Our results indicate that the association between the MGP-variant and increased risk for hand OA is caused by a lower expression of *MGP*, which may increase the burden of hand OA by decreased inhibition of cartilage calcification.

Keywords:

Genetics, Hand Osteoarthritis, genome-wide association study, functional study, Matrix-Gla protein

Introduction

Osteoarthritis (OA) is the most frequent joint disorder worldwide. An estimated 22% of the adult population has a joint affected by OA and this incidence increases to 49% in individuals over 65 years of age[1]. All synovial joints can be affected by OA, with hand OA as one of the most common forms of OA. Hand OA is characterized by osteophyte formation, bony enlargements of finger joints and cartilage degradation in the joints. One of the factors contributing to cartilage degradation is the increase of calcified cartilage in the joint[2,3]. In addition, hand OA is related to the occurrence of OA at other sites, most notably with knee OA[4,5]. Patients affected by hand OA suffer from pain and disability, impacting their quality of life. OA is a leading cause of chronic disability[6], yet currently no effective therapeutic treatments against osteoarthritis are known. It is therefore imperative to dissect the underlying mechanism of disease aetiology as this may enhance effective and targeted drug development.

OA has a strong genetic component. Depending on the joint affected, the heritability of OA is estimated in the range of 40-60%[7,8], with hand OA having the largest heritability, i.e. ~60%[9],[10]. Therefore, in recent years, several large-scale genetic studies have been performed to identify the underlying genes and pathways leading to OA. Multiple significant associated loci for OA of the hip and knee have been identified through genome-wide association studies (GWAS)[11–18]. However, thus far, only one report has described a robust association with OA of the hand[19]. In this previous report, common variants in the *ALDH1A2* and rare variants in chromosome 1p31 were genome-wide significantly associated with hand OA using a discovery cohort of 837 cases and 77,325 controls.

In this study, we aimed to identify novel genes and pathways involved in the aetiology of OA of the hand by performing a large-scale GWAS. We used a semi- quantitative measure for OA of the hand in order to increase statistical power. We gathered a large sample size of 12,754 individuals for analysis., by combining data from three studies in the discovery phase and an additional three cohorts for

80   replication. Next, we conducted functional follow-up of our top finding to investigate the underlying

81   mechanism.

82

83   **METHODS**

84   **Discovery GWAS, replication and meta-analysis**

85   For a detailed description on the GWAS methods, participating studies, quality control procedures for

86   genotyping and imputation, see supplementary text S1 and Table S1.

87   **Detailed phenotype description of KL sum-score**

88   We have used a semi-quantitative bilateral measure of osteoarthritis of the hand based on the radiographic

89   Kellgren and Lawrence score (KL-score)[20]. Using radiographs of both hands, the KL-score was determined for

90   each joint in the hand. Using these KL-scores we defined the KLsum-score: the total KL-score, the sum, of the

91   following hand joints for both hands (left and right): all Distal Interphalangeal (DIP) joints, all Proximial

92   Interphalangeal (PIP) joints, all Metacarpophalangeal (MCP) joints , the Interphalangeal (IP) joint and the first

93   Carpometacarpal (CMC1) joint. Which gives the sum of 15 joints on each hand, and in total 30 joints for both hands

94   together, resulting in a minimum score of 0 and a maximum score of 120. The Leiden Studies cohort no Kellgren-

95   Lawrence scoring was done of the MCP joints, resulting in a KLsum-score of maximum 88. Individuals lacking KL-

96   grading for both or one hands and individuals with missing age or gender information were excluded from all

97   analyses in all cohorts. As the KL-sum score has a skewed distribution the top finding of the meta-analysis was

98   repeated in the discovery cohorts using a Poisson regression.

99

100   **Visualization of the associated loci and the regulatory landscape**

101   For the top GWAS associated SNP, the LD region (r2>0.8) was determined using the 1000G Phase-1

102   population using the HaploReg V3 tool[21]. Using the ROADMAP generated reference epigenomes we

103   determined if any of the variants in high LD were located in potential gene regulatory regions in primary

104   osteoblasts (generated by ENCODE) and bone marrow derived chondrocytes (ROADMAP). [22,23].. The

105   18-state chromatin reference epigenomes were downloaded from the ROADMAP epigenomes data

106   portal[23]. SNPs and regulatory annotations were visualized busing the UCSC genome-browser on

107   GRCh37/hg19[24]. For each variant it was also determined if the alternative allele would disrupt a

108   protein binding motif, this was done using the HaploReg V3 tool[21].

109

110   **RNA-sequencing data**

111   Post RNA isolation (Qiagen RNeasy Mini Kit, RIN>7) of 40 knee (15 paired preserved (P) and OA lesioned

112   (OAL), 7 P only, 3 OAL only) and 28 hip (6 paired P and OAL, 14 P only, 2 OAL only) cartilage samples

113   (Supplementary Table S2), paired-end 2x100bp RNA library sequencing (Illumina TruSeq RNA-Library

114   Prep Kit, Illumina HiSeq2000) resulted in an average of 10 million fragments per sample. Reads were

115   aligned using GSNAP against GRCh37/hg19, in which SNPs from the Genome of the Netherlands

116   consortium withMAF>1% were masked to prevent alignment bias. Number of fragments per gene were

117   used to assess quantile-adjusted conditional maximum likelihood (qCML) (edgeR, R-

118   package).Subsequently, differential gene expression analysis was performed pairwise between P and

119   OAL samples for which we had RNA of both (n=21). ASE was assessed using SNVMix2[25] with default

120   settings(min coverage=25, 10 reads per allele). The extent of allele specific expression (ASE) was defined

121   as the fraction of risk allele among all counts at the respective location. Meta-analysis was done only

122   across P samples or OAL when no P counterpart sample was present. P-values were calculated using

123   canonical binominal test (metagen R-package).

124

125   **TaqMan assay**

126   **Conventional TaqMan genotyping was performed on both genomic-DNA (gDNA), articular cartilage**

127   **and Subchondral bone cDNA. An allele-specific custom TaqMan assay for rs1800801 (Thermo Fisher**

128   **Scientific) was used to quantify the allele ratio in cDNA samples and were normalized against the**

129 **gDNA ratio, which was used as an 1:1 allele ratio reference. Each sample has been measured in four**

130 **(cartilage) or eight (subchondral bone) times, while calculations and statistics were performed as**

131 **described previously**[19,26]**. Cartilage samples which yielded fewer than four measurements (N=2)**

132 **were discarded prior to further analyses. All subchondral bone samples were assessed by eight**

133 **technical replicates.**

134

135

136 **RESULTS**

137 *GWAS of KLsum-score*

138 We conducted a genome-wide association study (GWAS) of a semi-quantitative measure of hand OA, a

139 bilateral summed score of Kellgren- and Lawrence scores[20], that grades radiographic OA severity,

140 across all hand joints (KLsum-score, range of 0 to 120). The discovery set consisted of three Rotterdam

141 Study cohorts (RSI, RSII and, RSIII) and included 8,743 participants with KLsum-scores. Replication was

142 done in another 4,011 individuals from three different cohorts; Leiden studies(LS), Framingham heart

143 study(FHS), and Twins-UK (TUK). General characteristics of the discovery cohorts and replication cohorts

144 can be found in Supplementary Table S3 and in Supplementary Text S1.

145 The discovery analysis yielded two novel independent genome-wide significant loci (P≤5*10^{-8})

146 on chromosome 12, an intergenic region between *MGP* and *ERP27* and an intronic region in *CCDC91*. We

147 also identified seven other novel loci with suggestive significance (P<5*10^{-6}) (Figure 1). In total, nine loci

148 were selected for replication in 4,011 individuals from three different cohorts (LS, FHS, and TUK). Using a

149 Bonferroni corrected P-value<5.56*10^{-3}, we significantly replicated one of nine loci, rs4764133

150 (Beta$_{meta}$=0.83, SE$_{meta}$=0.10, P-value$_{replication}$=3.4*10^{-07}, P-value$_{meta}$=1.8*10^{-15}) with the same direction of

151 effect as identified in the discovery analysis (Table 1, and Supplementary Figure S1 Forrest plot). This

152 locus maintained genome-wide significance and another locus near *ENPP3* reached near genome-wide

153 significance (chr6:132063842:D, $Beta_{meta}$=0.58, SE_{meta}=0.11, P-value$_{meta}$=3.8*10^{-7}) in the combined

154 discovery and replication joint meta-analysis (Table 1). Since the KLsum-score has a skewed distribution

155 the top hit was also re-analysed in the discovery set using a Poisson regression (rs4764133,

156 $Beta_{poisson}$=0.12, $SE_{poisson}$=0.02, P-value$_{poisson}$=1.98*10^{-11}).

157   Our top replicated and genome-wide significant finding, rs4764133 [T] (P_{meta}=1.80*10^{-15},

158 Beta=0.83, MAF=0.39) is located in a non-coding intergenic region between *MGP* (Matrix Gla-protein)

159 and *ERP27* (Endoplasmic Reticulum Protein 27). However, variants in high linkage disequilibrium (LD)

160 with rs4764133 (r^2≥0.8) span a ~80Kb region encompassing multiple genes, including *MGP* and an open-

161 reading frame *C12orf60* (Figure 2A). Moreover, several of these variants are located in an mRNA

162 transcript, including a nonsynonymous variant in *MGP*, and variants in 3' and 5'UTR of *MGP* and

163 *C12orf60* (Table 2, Figure 2B). The nonsynonymous variant in *MGP,* rs4236, is predicted to be non-

164 damaging (STIFT=1, tolerated; PolyPhen=0, benign) causing a threonine to alanine amino acid

165 substitution. Two variants are located in predicted active promoter region of *MGP (*rs1800801) and

166 *C12orf60 (*rs9668569) in chondrogenic cells and primary osteoblasts (Table 2).

167   Next, we investigated the association of rs4764133 with bilateral severe hand OA and bilateral

168 finger OA using the discovery set (RS-I, RS-II and RS-III). We found a strong association with finger OA (P-

169 value=3.09*10-08, OR=1.25) and nominal significant association with severe hand OA (p-value=2.80*10-

170 2, OR=1.36), which has a low frequency in the population (Table S4), we also see a nominal significant

171 association with cartilage thickness in the hip joint (minimal joint space width. mJSW). To see if

172 rs4764133 also confers risk for other forms of osteoarthritis, i.e. osteoarthritis of the hip and knee, we

173 used the GWAS summary data of the treat the treat OA consortium[27] and the recently published

174 mJSW meta-analysis[18]. No association was found between rs4764133 and hip or knee OA (Table S4).

175   However, we did find a nominal significant association between rs rs1049897 (r2= 0.98 with

176   rs4764133)(P-value=1.28*10-2, Beta=-0.398).

177

178   *Gene expression analyses*

179   In order to identify potential causal genes located in the LD block surrounding rs4764133, we assessed

180   gene expression of *MGP, ERP27, ART4, SMOC3 (C12orf69) and C12orf60* in articular cartilage, the

181   primary OA affected tissue. RNA sequencing was obtained on articular cartilage from primary OA

182   patients who had total joint replacement surgeries of either the knee (n=25) or hip (n=22) joint.

183   Expression levels of *ERP27,C12orf60,ART4 and SMOC3* were substantially lower than *MGP* expression

184   levels in articular cartilage (Figure. S2A). Nonetheless, neither *MGP, ERP27, ART4 nor SMOC3* and

185   *C12orf60* showed significant difference in gene expression between paired preserved(P) and OA

186   lesioned(OAL) articular cartilage. However, while these genes are not differentially expressed in OA

187   affected cartilage, it is possible that the identified GWAS SNPs affect gene transcription. When we

188   analysed the relationship between the top SNP and expression analysis in a classical eQTL (expression

189   Quantitative Trait Loci) analysis, we did not to detect significant correlations between rs1049897, rs4236

190   or rs1800801 and absolute *MGP, ERP27, ART4, SMOC3* or *C12orf60* expression levels (Figure S2B).

191   However, we did observe several variants in high LD located in the mRNA transcript of *MGP* and

192   *C12orf60*, allowing us to assess allele specific expression (ASE) for these genes. We were unable to study

193   ASE for *ART4*, *SMOC3* and *ERP27,* since no SNP in high LD with rs4764133 is present in the coding region.

194   In ASE the influence of exonic alleles on gene expression *in-cis* is measured within heterozygote

195   subjects, circumventing strong effects from environmental or trans-acting influences. This property

196   results in ASE analysis to be a more statistically powerful approach, when compared to classical eQTL

197   analysis [28]. Subsequently, we found that the OA risk alleles for three coding variants in high LD with

198   the lead variant, rs4236 (Figure. 3SA, 39.6% C allele, P<5*10-16), rs1049897 (Figure. S3B, 44.4% A allele,

199 P<5*10-10), and rs1800801 (Figure. 3A, 40.7% T allele, P<5*10-16), were significantly correlated with

200 lower expression of *MGP*, marking imbalanced expression among heterozygotes, independent of the

201 disease status of the articular cartilage. No allele specific expression was observed between SNPs

202 rs11276, rs3088189 and rs1861698 (residing in *C12orf60* and in high LD with the lead SNP,r^2>0.8, Table

203 2). Technical and biological replication was performed using a custom allele specific TaqMan assay for

204 rs1800801 in eight additional heterozygous individuals for which we isolated RNA from P cartilage (n=2),

205 OAL (n=2) or both (n=4) from primary knee OA patients and confirmed the observed imbalance in

206 preserved articular cartilage (Figure 3B, relative allelic difference=0.92, $P<1*10^{-6}$), as well as in 8 knee

207 subchondral bone samples (Figure 3C, relative allelic difference=0.78, $P<1*10^{-4}$).

208

209 **Discussion**

210 Here, we show for the first time, that there is a robust genome-wide significant association between

211 rs4764133, located near *MGP,* and hand OA. Furthermore we performed functional validation showing

212 that *MGP* coding variants in LD with rs4764133 are associated with allele-specific expression of *MGP*

213 which may increase risk of hand OA by lowering inhibition of articular cartilage calcification, since MGP

214 is an essential inhibitor of cartilage calcification[29,30]. These findings suggest that MGP could be

215 considered a prioritized drug target for hand OA, since genetically supported drug targets double the

216 success rate of therapeutics in clinical development[31].

217   *MGP* is an essential inhibitor of cartilage calcification, and genetic deficiencies of *MGP* in

218 humans and mice have been linked to abnormal mineralization of soft tissues, including cartilaginous

219 tissue[29,32]. Furthermore *MGP* has been previously implicated in relation to OA. A small candidate

220 study reported marginally significant association between hand OA and genetic variants in *MGP*

221 (rs1800802 and rs4236)[33]. This is consistent with our findings that the minor allele for rs4764133 and

222 related coding variants in high LD (r2>0.8), rs1800802 and rs4236, increase the risk of hand OA and that

9

223 we found high expression of *MGP* in both preserved and OA lesioned articular cartilage. In contrast,

224 another study showed that an MGP protein complex is excreted by healthy articular chondrocytes, but

225 not by OA affected chondrocytes[34], although we only assed *MGP* expression and not MGP protein

226 complex excretion.

227   Although the loci with allele specific expression (ASE) are known to be enriched for eQTLs[35],

228 we were unable to detect an association between the MGP-genotype and *MGP* RNA-expression levels in

229 cartilage. This could have been due to our modest sample size (knee joint, n=25 and or hip joint, n=22) in

230 combination with large heterogeneity of the tissue. Notably, the available cartilage samples originated

231 from different joint sites (knee, hip) and different disease stage (preserved versus affected), and had

232 large age range of the individuals. Also, It is known that ASE is a more powerful technique then classical

233 eQTL analysis to identify functional SNPs influencing expression of genes[28]. While the extent of

234 imbalance could be considered relatively modest, an increasing number of OA associated SNP alleles

235 appear to mark ASE by comparable amount[19,36–38]. From a more biological perspective, one could

236 consider a prolonged, albeit slight, deviation from homeostasis due to modest ASE of cartilage relevant

237 genes to be of substantial influence over time. This latter hypothesis could contain the molecular basis

238 for increased risk towards developing OA among the ageing population. Additionally, we observed that

239 the rs1800801 alleles also affected expression of MGP in subchondral bone samples. This could imply

240 that, in parallel to an effect in cartilage, he presumed disturbed cartilage homeostasis is further affected

241 by the underlying bone. Further enabling the view that OA is a pathology of the entire joint.

242   Our findings may give an explanation for the known vitamin K association with OA: MGP

243 mediated calcification inhibition is dependent on γ-carboxylation by vitamin K[39]. It has been shown

244 that low vitamin K intake is correlated with OA[40]. Thus vitamin K intake may be a potential therapeutic

245 treatment in OA. Recently, a first randomized control trial testing the effects of vitamin K on OA was

246 published, which reported no overall effect of vitamin K on hand OA[41]. Despite the low power of the

247   trial, there was a significant beneficial effect on joint space narrowing (cartilage degradation) among

248   those individuals that were VitK deficient at the start of the trial[41]. Thus, an adequately powered study

249   of vitamin K may be justified based on the found MGP association. Furthermore, genetic predisposition

250   for hand OA, was not taken into account in the trial. Perhaps, genetic predisposition for hand OA (*MGP*-

251   risk variants) in combination with insufficient vitamin K intake might potentiate cartilage calcification

252   and subsequent risk for developing hand OA. Therefore, future OA trails, therapeutic and preventive

253   treatments might benefit from taking a personalized medicine approach since genetically supported

254   drug targets double the success rate of therapeutics in clinical development[31].

255   Styrkarsdottir *et al.*(2014) reported on common genetic variants that associate with severe hand

256   OA, among the replication cohorts were the Leiden and Rotterdam cohorts[19]. Although we observe

257   suggestive signals at the reported locus (*ALDH1A2* gene, 1p31) the respective variants did not meet the

258   genome-wide significance threshold in our analyses (Supplementary Table S5). This difference is likely

259   caused by the markedly different phenotypes that were used for either analyses. Where Styrkarsdottir

260   *et al.* studied a dichotomous severe hand OA phenotype, our phenotype was semi-quantitatively

261   phenotype.

262   To conclude, we here present coding variants in *MGP*, that are associated with radiographic

263   hand OA, and the hand OA risk allele marks lower expression of *MGP* in articular cartilage. Our findings

264   suggest that *MGP* might play an important role in hand OA pathogenesis through pathways related to

265   articular cartilage calcification and vitamin K. Better understanding of MGP gene and protein regulation

266   and its relation to vitamin K intake and OA, may reveal novel therapeutic drug targets for hand OA.

267

268

308

309   **AUTHOR CONTRIBUTIONS**

310   W.H. and C.G.B. contributed equally to this work. D.H., M.S.Y., Y.F.M.R. and S.M. performed replication

311   analysis for this work, L.B. provided analysis help. L.S.C. and F.R. provided data. M.K. provided

312   phenotypic contribution to the GARP study. M.P. provided data and analyses. T.D.S. contributed data for

313   replication. A.H. contributed data of the RS cohorts. J.D., M.B. and P.E.S contributed to genotyping data

314   and analyses of LLS cohort. R.G.H.H.N. provided contribution to the RAAK study. A.G.U. contributed

315   genotype data of RS cohorts , D.T.F. and A.V. contributed replication data for this work. I.M., and J.B.J.M.

316   jointly supervised this work.

317

Competing Interests

319   The Authors declare no competing interests.

320

Materials & Correspondence

322   GWAS summary statistics can be requested from J.B.J.M. All correspondence should be addressed to
323   J.B.J.M. and I.M.

324

325

326

References

327

328 1   CDC. Prevalence of Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation — United
329   States, 2010–2012. *MMWR Morb Mortal Wkly Rep* 2013;**62**:869–73. doi:10.1001/jama.296.22.2671

330 2   Patel N, Buckland-Wright C. Advancement in the zone of calcified cartilage in osteoarthritic hands of
331   patients detected by high definition macroradiography. *Osteoarthr Cartil* 1999;**7**:520–
332   5.http://www.idealibrary.com (accessed 2 Nov 2016).

333 3   Fuerst M, Bertrand J, Lammers L, *et al.* Calcification of articular cartilage in human osteoarthritis. *Arthritis*
334   *Rheum* 2009;**60**:2694–703. doi:10.1002/art.24774

335 4   Nelson AE, DeVellis RF, Renner JB, *et al.* Quantification of the whole-body burden of radiographic
336   osteoarthritis using factor analysis. *Arthritis Res Ther* 2011;**13**:R176. doi:10.1186/ar3501

337 5   Jonsson H, Helgadottir GP, Aspelund T, *et al.* Hand Osteoarthritis Severity is Associated with Total Knee
338   Joint Replacements Independently of BMI. The Ages-Reykjavik Study. *Open Rheumatol J* 2011;**5**:7–12.
339   doi:10.2174/1874312901105010007

340 6   CDC. Prevalence of Disabilities and Associated Health Conditions Among Adults --- United States, 1999.
341   *MMWR Morb Mortal Wkly Rep* 2009;**58**:421–6.

342 7   Spector TD, Cicuttini F, Baker J, *et al.* Genetic influences on osteoarthritis in women: a twin study. *BMJ*
343   1996;**312**:940–3. doi:10.1136/bmj.312.7036.940

344 8   Felson DT. Risk factors for osteoarthritis: understanding joint vulnerability. *Clin Orthop Relat Res*
345   2004;:S16–21. doi:10.1097/01.blo.0000144971.12731.a2

346 9   Bijkerk C, Houwing-Duistermaat JJ, Valkenburg H a., *et al.* Heritabilities of radiologic osteoarthritis in
347   peripheral joints and of disc degeneration of the spine. *Arthritis Rheum* 1999;**42**:1729–35.
348   doi:10.1002/1529-0131(199908)42:8<1729::AID-ANR23>3.0.CO;2-H

349 10 Ishimori ML, Altman RD, Cohen MJ, *et al.* Heritability patterns in hand osteoarthritis: the role of
350   osteophytes. *Arthritis Res Ther* 2010;**12**:R180. doi:10.1186/ar3144

351 11 Kerkhof HJM, Lories RJ, Meulenbelt I, *et al.* A genome-wide association study identifies an osteoarthritis
352   susceptibility locus on chromosome 7q22. *Arthritis Rheum* 2010;**62**:499–510. doi:10.1002/art.27184

353 12 Nakajima M, Takahashi A, Kou I, *et al.* New sequence variants in HLA class II/III region associated with
354   susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One* 2010;**5**:e9723.
355   doi:10.1371/journal.pone.0009723

356 13 Castaño Betancourt MC, Cailotto F, Kerkhof HJ, *et al.* Genome-wide association and functional studies
357   identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Natl Acad Sci U S*
358   *A* 2012;**109**:8218–23. doi:10.1073/pnas.1119899109

359 14 Consortium A, Collaborators A. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a
360   genome-wide association study. *Lancet* 2012;**380**:815–23. doi:10.1016/S0140-6736(12)60681-3

361 15 Day-Williams AG, Southam L, Panoutsopoulou K, *et al.* A variant in MCF2L is associated with osteoarthritis.
362   *Am J Hum Genet* 2011;**89**:446–50. doi:10.1016/j.ajhg.2011.08.001

363 16 Evangelou E, Kerkhof HJ, Styrkarsdottir U, *et al.* A meta-analysis of genome-wide association studies
364   identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis* 2013;:1–7.

365   doi:10.1136/annrheumdis-2012-203114

366   17   Miyamoto Y, Shi D, Nakajima M, *et al.* Common variants in DVWA on chromosome 3p24.3 are associated
367   with susceptibility to knee osteoarthritis. *Nat Genet* 2008;**40**:994–8. doi:10.1038/ng.176

368   18   Castaño-Betancourt MC, Evans DS, Ramos YFM, *et al.* Novel Genetic Variants for Cartilage Thickness and
369   Hip Osteoarthritis. *PLOS Genet* 2016;**12**:e1006260. doi:10.1371/journal.pgen.1006260

370   19   Styrkarsdottir U, Thorleifsson G, Helgadottir HT, *et al.* Severe osteoarthritis of the hand associates with
371   common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;**46**:498–502.
372   doi:10.1038/ng.2957

373   20   Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. *Ann Rheum Dis* 1957;**16**:494–502.
374   doi:10.1136/ard.16.4.494

375   21   Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif
376   alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;**40**:930–4.
377   doi:10.1093/nar/gkr917

378   22   Chadwick LH. The NIH Roadmap Epigenomics Program data resource. *Epigenomics* 2012;**4**:317–24.
379   doi:10.2217/epi.12.18

380   23   Kundaje A, Meuleman W, Ernst J, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*
381   2015;**518**:317–30. doi:10.1038/nature14248

382   24   Kent WJ, Sugnet CW, Furey TS, *et al.* The Human Genome Browser at UCSC. *Genome Res* 2002;**12**:996–
383   1006. doi:10.1101/gr.229102

384   25   Goya R, Sun MGF, Morin RD, *et al.* SNVMix: predicting single nucleotide variants from next-generation
385   sequencing of tumors. *Bioinformatics* 2010;**26**:730–6. doi:10.1093/bioinformatics/btq040

386   26   Bos SD, Bovée JVMG, Duijnisveld BJ, *et al.* Increased type II deiodinase protein in OA-affected cartilage and
387   allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Ann Rheum Dis*
388   2012;**71**:1254–8. doi:10.1136/annrheumdis-2011-200981

389   27   Kerkhof HJM, Doherty M, Arden NK, *et al.* Large-scale meta-analysis of interleukin-1 beta and interleukin-1
390   receptor antagonist polymorphisms on risk of radiographic hip and knee osteoarthritis and severity of knee
391   osteoarthritis. *Osteoarthr Cartil* 2011;**19**:265–71. doi:10.1016/j.joca.2010.12.003

392   28   Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat Rev Genet*
393   2010;**11**:533–8. doi:10.1038/nrg2815

394   29   Luo G, Ducy P, McKee MD, *et al.* Spontaneous calcification of arteries and cartilage in mice lacking matrix
395   GLA protein. *Nature* 1997;**386**:78–81. doi:10.1038/386078a0

396   30   Yagami K, Suh JY, Enomoto-Iwamoto M, *et al.* Matrix GLA protein is a developmental regulator of
397   chondrocyte mineralization and, when constitutively expressed, blocks endochondral and
398   intramembranous ossification in the limb. *J Cell Biol* 1999;**147**:1097–
399   108.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2169349&tool=pmcentrez&rendertype=a
400   bstract

401   31   Nelson MR, Tipney H, Painter JL, *et al.* The support of human genetic evidence for approved drug
402   indications. *Nat Genet* 2015;**47**:856–60. doi:10.1038/ng.3314

403   32   Munroe PB, Olgunturk RO, Fryns JP, *et al.* Mutations in the gene encoding the human matrix Gla protein

404   cause Keutel syndrome. *Nat Genet* 1999;**21**:142–4. doi:10.1038/5102

405 33 Misra D, Booth SL, Crosier MD, *et al.* Matrix Gla protein polymorphism, but not concentrations, is
406   associated with radiographic hand osteoarthritis. *J Rheumatol* 2011;**38**:1960–5.
407   doi:10.3899/jrheum.100985

408 34 Wallin R, Schurgers LJ, Loeser RF. Biosynthesis of the vitamin K-dependent matrix Gla protein (MGP) in
409   chondrocytes: a fetuin-MGP protein complex is assembled in vesicles shed from normal but not from
410   osteoarthritic chondrocytes. *Osteoarthritis Cartilage* 2010;**18**:1096–103. doi:10.1016/j.joca.2010.05.013

411 35 Hangoc G, Williams DE, Falkenburg JH, *et al.* Influence of IL-1 alpha and -1 beta on the survival of human
412   bone marrow cells responding to hematopoietic colony-stimulating factors. *J Immunol* 1989;**142**:4329–
413   34.http://www.ncbi.nlm.nih.gov/pubmed/2656864 (accessed 22 Aug 2016).

414 36 Gee F, Clubbs CF, Raine EVA, *et al.* Allelic expression analysis of the osteoarthritis susceptibility locus that
415   maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC Med Genet* 2014;**15**:53.
416   doi:10.1186/1471-2350-15-53

417 37 Raine EVA, Dodd AW, Reynard LN, *et al.* Allelic expression analysis of the osteoarthritis susceptibility gene
418   COL11A1 in human joint tissues. *BMC Musculoskelet Disord* 2013;**14**:85. doi:10.1186/1471-2474-14-85

419 38 Bos SD, Bovée JVMG, Duijnisveld BJ, *et al.* Increased type II deiodinase protein in OA-affected cartilage and
420   allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Ann Rheum Dis*
421   2012;**71**:1254–8. doi:10.1136/annrheumdis-2011-200981

422 39 Schurgers LJ, Uitto J, Reutelingsperger CP. Vitamin K-dependent carboxylation of matrix Gla-protein: a
423   crucial switch to control ectopic mineralization. *Trends Mol Med* 2013;**19**:217–26.
424   doi:10.1016/j.molmed.2012.12.008

425 40 Shea MK, Kritchevsky SB, Hsu F-C, *et al.* The association between vitamin K status and knee osteoarthritis
426   features in older adults: the Health, Aging and Body Composition Study. *Osteoarthritis Cartilage*
427   2015;**23**:370–8. doi:10.1016/j.joca.2014.12.008

428 41 Neogi T, Felson DT, Sarno R, *et al.* Vitamin K in hand osteoarthritis: results from a randomised clinical trial.
429   *Ann Rheum Dis* 2008;**67**:1570–3. doi:10.1136/ard.2008.094771

430

431

432 **Figures**

433 **Figure 1.** GWAS results for association with the KLsum score in the discovery phase. Manhattan plot for association

434 with the KLsum-score, adjusted for age and sex, in the discovery cohorts of RSI, RSII and RSIII. The -log10 P-values,

435 for each of the ~11 million SNPs analyzed (remaining after EASYQC quality control) as part of the genome wide

436 association with the KLsum-score, plotted against their position per chromosome. The red dotted horizontal line

437 corresponds to the genome-wide significant threshold (P = $5x10^{-8}$). The dotted grey line corresponds to the

438 selection for replication threshold (P = $5x10^{-6}$). SNP location represented by [], if the SNP is localized intergenic the

439 dashes denotes the distance, -≤10 kb, --≤100kb, ---≤1000kb, ----≤1Mkb, -----≥1Mkb.

440

441 **Figure 2.** Locus zoom plot for rs4764133. locus zoom plot for rs4764133, 150 kb upstream and downstream of

442 rs4764133 has been taken as plotted region (**A**). Zoom in on *MGP* and three SNPs in high LD with top SNP that are

443 located in the MGP mRNA transcript (**B**). Also represented is ROADMAP chromatin 18-state data of two tissue

444 types: human Mesenchymal Stem Cell (hMSC) derived cultured chondrogenic cells and primary osteoblasts. In

445 both these cell types the chromatin contains active marks surrounding the *MGP* promoter.

446

447 **Figure 3.** Allelic imbalanced expression of *MGP* marked by the alleles among heterozygotes of rs1800801 (**A**), in

448 the assessed cartilage RNA sequencing dataset. Validation of selected rs1800801 using a custom TaqMan assay

449 confirmed the imbalance (**B**). Allelic imbalance was also assed in with a custom TagMan assay in subchondral bone

450 samples (**C**). Preserved (P) and OA lesioned (OAL) samples are shown respectively in blue and red, and genomic

451 DNA (TaqMan control) in black (G). For ASE results for rs4236 and rs1049879, see Supplementary Figure S3 and for

452 information on the samples see Supplementary Table S2.

453

454

455   **Table 1.** Results GWAS quantitative bilateral phenotype of osteoarthritis of the hand (KLsum-score), discovery, replication and meta-analysis

SNP	Chr	Position (hg19)	Effect Allele	Other Allele	EAF†	Discovery*			Replication**			Combined			Genomic location††
						Beta	SE	P-value	Beta	SE	P-value	Beta	SE	Pval	
rs1494593	5	23574856	T	C	0.88	-0.83	0.18	4.46E-06	-0.15	0.32	6.31E-01	-0.67	0.16	2.28E-05	*PRDM9--[]---C5orf17*
rs114370021	5	167398535	A	G	0.27	-0.84	0.18	2.12E-06	0.48	0.32	1.28E-01	-0.53	0.16	6.96E-04	*[TENM2]*
rs7770034	6	44447004	A	G	0.48	-0.54	0.12	3.42E-06	-0.36	0.21	8.71E-02	-0.50	0.10	1.01E-06	*CDC5L --[]---SUPT3H*
6:132063842:D‡	*6*	*132063842*	*D*	*I*	*0.27*	*0.64*	*0.13*	*1.28E-06*	*0.41*	*0.23*	*7.82E-02*	*0.58*	*0.11*	*3.79E-07*	*[ENPP3]*
11:90657297:D	11	90657297	D	I	0.11	0.93	0.19	1.35E-06	0.33	0.33	3.11E-01	0.78	0.17	2.91E-06	*DISC1FP1-[]----FAT3*
rs4764133	**12**	**15064363**	**T**	**C**	**0.39**	**0.75**	**0.12**	**3.45E-10**	**1.11**	**0.22**	**3.34E-07**	**0.83**	**0.10**	**1.80E-15**	***MGP--[]- ERP27***
rs7139060	12	28693144	A	G	0.67	-0.73	0.12	6.12E-09	0.11	0.22	6.16E-01	-0.52	0.11	1.47E-06	*[CCDC91]*
rs1950427	14	25955502	T	C	0.12	0.86	0.18	1.09E-06	-0.28	0.30	3.56E-01	0.57	0.15	1.80E-04	*STXBP6---[]---NOVA1*
rs6108226	20	8960884	T	C	0.77	0.70	0.15	3.94E-06	0.20	0.27	4.43E-01	0.58	0.13	1.14E-05	*PLCB1--[]--PLCB4*

* Discovery: RS-I, RS-II, RS-III, *n*= 8,743

** Replication : GARP, LSS, TwinsUK & FHS, *n*= 4,011

† EAF: Effect Allele Frequency

††SNP location represented by [], if the SNP is localized intergenic the dashes denotes the distance, -≤10 kb, --≤100kb, ---≤1000kb, ----≤1Mkb, -----≥1Mkb

‡For TwinsUK a proxy SNP was used: rs3850251 r^2=1 D'=1 (as calculated in the RSI, RSII and RSIII cohorts)

Table 2. rs4764133 LD block (r²>0.8) annotation of potential functional elements in osteoblasts and chondrogenic cells, X marks no potential functional annotation i.e. enhancer region, promoter region or altered protein binding motifs.

| SNP | P-value Discovery | r² | Annotation* | Regulatory Chromatin Marks** | | Altered Protein Binding Motifs (Haploreg V3) |
				Chondrogenic cells	Osteoblasts	
rs1049897	3.48E-09	0.88	MGP 3'-UTR	Transcription	X	X
rs4236	4.16E-09	0.86	MGP non-synonymous	Enhancer region	X	HNF4, PLAG1
rs1800801	1.12E-09	0.95	MGP 5'UTR	Promoter region	Promoter region	Zfp410
rs7310951	4.04E-09	0.86	C12orf60	Enhancer region	X	DMRT7, Gfi1, Pax-5
rs12320004	4.04E-09	0.86	C12orf60	Enhancer region	X	BHLHE40, P300, HEN1, LBP-1, RAD21,TATA, Zfx
rs10772814	3.76E-09	0.88	C12orf60	Enhancer region	X	HNF4
rs10492151	1.21E-09	0.95	C12orf60	Enhancer region	X	AIRE, Hoxa13
rs725445	3.58E-08	0.82	C12orf60	Enhancer region	X	Hand1
rs725444	3.92E-09	0.87	C12orf60	Enhancer region	X	Foxf1, Foxi1, Foxo, Foxq1, Mef2
rs4764131	6.31E-10	0.97	C12orf60	Enhancer region	Enhancer region	Myc
rs9668569	5.91E-10	0.97	C12orf60	Promoter region	Promoter region	X
rs2430687	2.44E-09	0.89	C12orf60	Enhancer region	Enhancer region	BHLHE40
rs12311463	6.91E-10	0.97	C12orf60	Enhancer region	Enhancer region	Pou1f1, Pou2f2, TATA
rs67482087	4.61E-10	0.95	C12orf60	Enhancer region	Enhancer region	Foxp1, Irx, Pou1f1, Pou2f2, Pou3f3, TATA
rs67436073	6.76E-10	0.97	C12orf60	Enhancer region	Enhancer region	Foxj2, Foxk1, Foxo, GATA, Mef2, Pou2f2, Pou3f2, Pou6f1, TATA, Zfp
rs11276	8.05E-09	0.96	C12orf60 non-synonymous	X	X	SPIB,NF-AT
rs3088189	9.46E-09	0.96	C12orf60 synonymous	X	X	SPIB
rs1861698	3.56E-09	0.96	C12orf60 synonymous	X	X	Bbx, Pou1f1, TATA

*Gene annotation based on the hg19 release of the UCSC Genome Browser
** Regulatory chromatin marks taken from the ROADMAP Epigenomes project chromatin state learning core 18-state model