Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31.

Styrkarsdottir, Unnur; Thorleifsson, Gudmar; Helgadottir, Hafdis T; Bomer, Nils; Metrustry, Sarah; Bierma-Zeinstra, S; Strijbosch, Annelieke M; Evangelou, Evangelos; Hart, Deborah; Beekman, Marian; Jonasdottir, Aslaug; Sigurdsson, Asgeir; Eiriksson, Finnur F; Thorsteinsdottir, Margret; Frigge, Michael L; Kong, Augustine; Gudjonsson, Sigurjon A; Magnusson, Olafur T; Masson, Gisli; Hofman, Albert; Arden, Nigel K; Ingvarsson, Thorvaldur; Lohmander, Stefan; Kloppenburg, Margreet; Rivadeneira, Fernando; Nelissen, Rob G H H; Spector, Tim; Uitterlinden, Andre; Slagboom, P Eline; Thorsteinsdottir, Unnur; Jonsdottir, Ingileif; Valdes, Ana M; Meulenbelt, Ingrid; van Meurs, Joyce; Jonsson, Helgi; Stefansson, Kari

Published in:
Nature Genetics

DOI:
10.1038/ng.2957

2014

Severe osteoarthritis of the hand associates with common variants within the *ALDH1A2* gene and with rare variants at 1p31

Unnur Styrkarsdottir[1], Gudmar Thorleifsson[1], Hafdis Helgadottir[1], Nils Bomer[2], Sarah Metrustry[3], S Bierma-Zeinstra[4], Annelieke M Strijbosch[2], Evangelos Evangelou[5], Deborah Hart[3], Marian Beekman[2,6,7], Aslaug Jonasdottir[1], Asgeir Sigurdsson[1], Finnur F Eiriksson[8], Margret Thorsteinsdottir[8, 9], Michael L Frigge[1], Augustine Kong[1], Sigurjon A Gudjonsson[1], Olafur T Magnusson[1], Gisli Masson[1], The TREAT-OA consortium, arcOGEN consortium, Albert Hofman[10], Nigel K Arden[11], Thorvaldur Ingvarsson[12], Stefan Lohmander[13], Margreet Kloppenburg[14], Fernando Rivadeneira[4,10], Rob G H H Nelissen[15], Tim Spector[3], Andre Uitterlinden[4,10], P Eline Slagboom[2,6,7], Unnur Thorsteinsdottir[1,9], Ingileif Jonsdottir[1,9], Ana M Valdes[3,16], Ingrid Meulenbelt[2,7], Joyce van Meurs[4], Helgi Jonsson[9,17], Kari Stefansson[1,9]

*Corresponding author: Kari Stefansson, deCODE genetics/Amgen, Sturlugata 8, 101 Reykjavik, Iceland, kstefans@decode.is, phone: 354-5701931, fax: 354-570-2850.

1. deCODE Genetics/Amgen, Sturlugata 8, Reykjavik, Iceland.
2. Dept. Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.
3. Dept of Twin Research, King's College London, St Thomas' Hospital, London, United Kingdom.
4. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
5. Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece.
6. IDEAL The Netherlands.
7. Genomics Initiative, sponsored by the NCHA, Leiden, The Netherlands.
8. Arctic Mass, Sturlugata 8, Reykjavik, Iceland.
9. University of Iceland, Faculty of Medicine, Reykjavik, Iceland.
10. Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.
11. NIHR Biomedical Research Unit, University of Oxford, England.
12. Department of Orthopedic Surgery, Akureyri Hospital, Akureyri, Iceland and University of Akureyri, Institution of Health Science, Akureyri, Iceland.
13. Department of Orthopedics, Clinical Sciences Lund, Lund University, Lund, Sweden.
14. Dept. Clinical Epidemiology and Dept. Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.
15. Dept. of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands.
16. Academic Rheumatology, University of Nottingham, Nottingham City Hospital, Nottingham, United Kingdom.
17. Department of Medicine, Landspitali,The National University Hospital of Iceland, Reykjavik, Iceland.

Osteoarthritis is the most common form of arthritis and a major cause of pain and disability in the elderly. Genome-wide association (GWA) studies and meta-analyses efforts have yielded several significant loci for osteoarthritis of the hip and of the knee[1-7]. To search for sequence variants that confer risk of osteoarthritis of the hand, we carried out a genome wide association study in subjects with severe hand osteoarthritis, using variants identified through a whole genome sequencing effort of 2,230 Icelanders. We found two significant loci in the Icelandic discovery set; at 15q22 (freq. 50.7%, OR 1.51, $P = 3.99 \times 10^{-10}$) in the *ALDH1A2* gene, and at 1p31 (freq. 0.02%, OR 50.6, $P = 9.8 \times 10^{-10}$). Among the carriers of the variant at 1p31 is a family with several members in whom the risk allele segregates with osteoarthritis. The variants within the *ALDH1A2* gene were confirmed in replication sets from the Netherlands and the United Kingdom, yielding an overall association of OR = 1.46 and $P = 1.1 \times 10^{-11}$ (rs3204689).

Osteoarthritis has a great impact on quality of life due to pain and the loss of joint function. Osteoarthritis is characterized by a dynamic process of destruction and repair of joint tissues. It is a disease of the entire joint, affecting cartilage, synovium, subchondral bone, ligaments and the joint capsule[8]. The clinical presentation of osteoarthritis is heterogeneous. Heritability estimates of osteoarthritis are in the range of 40-65%, depending on the joint[9-11]. There is a clear predilection for certain joints; most common joints to be affected are the knees, hips, hands, spine, neck and the big toe. However, patients with osteoarthritis may have one, a few, or most of these sites affected. A recent report found the prevalence of symptomatic hand osteoarthritis in the general population to be 15.9% in women and 8.2 % in men[12].

GWA studies have yielded several significant loci for osteoarthritis of the hip and osteoarthritis of the knee[1-7]. Hitherto, no study has reported genome wide significant association with hand osteoarthritis.

With the aim of discovering sequence variants that confer risk of osteoarthritis of the hand, we carried out a GWA study in subjects with severe hand osteoarthritis. We included 623 Icelanders with severe osteoarthritis of the hand (severe osteoarthritis at the thumb base and severe finger osteoarthritis in the same person) as cases and 69,153 individuals as population controls (Online **Methods** and **Supplementary Information** for detailed description of phenotype and sample set). We then tested for association between 34.2 million sequence variants, that were identified through whole-genome sequencing of 2,230 Icelanders and subsequently imputed into the cases and controls (Online **Methods**), and severe osteoarthritis of the hand. The most significant associations were with several common variants at 15q22 (**Supplementary Figure 1**), represented by rs12907038[G] (frequency 50.7%, OR 1.51, $P = 3.99 \times 10^{-10}$), and with very rare variants at 1p31 (NCBI_build36/hg18_chr1: 63807756[A] frequency 0.022%, OR 47.7, $P = 1.53 \times 10^{-9}$).

The group of variants at 15q22 includes 55 variants with $P < 5 \times 10^{-8}$, with frequency of the risk alleles between 41% and 54% (**Supplementary Table 1**). These variants are all located within a single linkage disequilibrium block that contains one gene, *ALDH1A2* (aldehyde dehydrogenase 1 family, member A2) (**Figure 1**). We note that the top SNPs fall mainly within two frequency groups, around 41% and 52% for the risk allele, and are therefore not fully correlated with one another (e.g. $r^2 = 0.62$ between rs4238326 (41%) and rs3204689 (52%)). Furthermore, the most strongly associated SNP, rs12907038, is tri-allelic and is likewise not fully

correlated with the other two markers. Conditional analysis shows that markers in either frequency group alone could not account for the association signal, however, the tri-allelic rs12907038 SNP can fully account for the signal (Online **Methods, Supplementary Table 1**).

We tested SNPs from the two frequency groups, rs4238326 (41%) and rs3204689 (52%), in five additional European sample sets; the Dutch GARP[13]/LLS[14] and Rotterdam-I and II[15] cohorts and the UK Twins[16] and Chingford[17] cohorts since the tri-allelic marker rs12907038 is not present in these *in-silico* replication datasets. Both markers associate with severe hand osteoarthritis in the replication samples with $P = 0.0075$ for rs4238326 and $P = 0.0011$ for rs3204689 (**Table 1**, **Supplementary Table 2**), yielding an overall association of OR = 1.44 and $P = 8.6\times10^{-11}$ and OR = 1.46 and $P = 1.1\times10^{-11}$ for rs4238326 and rs3204689, respectively, in the combined analysis of all samples (**Table 1**). Conditional analysis in the replication samples showed that rs3204689 is nominally significant after accounting for rs4238326 (**Supplementary Table 3**). We next examined the association separately with severe finger osteoarthritis and with severe thumb osteoarthritis in all sample-sets. The effect of the variant on severe finger osteoarthritis is comparable to that on severe thumb osteoarthritis (OR between 1.26 and 1.14), although, overall, the association is somewhat more significant for the severe fingers phenotype, or $P = 2.0\times10^{-11}$ and $P = 5.6\times10^{-12}$, compared to $P = 1.5\times10^{-8}$ and $P = 5.1\times10^{-7}$ for association with severe thumb osteoarthritis for rs4238326 and rs3204689, respectively (**Table 1**). To determine whether this locus also associates with osteoarthritis of the larger joints (hip and knee) we examined the results for these markers in large meta-analyses of hip osteoarthritis (4,349 cases and 46,903 controls) and knee osteoarthritis (5,224 cases and 48,571 controls) conducted by the TreatOA (Translational Research in Europe Applied Technologies for

OsteoArthritis) consortium[18] (**Table 2**) (Online **Methods, Supplementary Table 4**). No association is observed with hip osteoarthritis (OR = 1.04, *P* = 0.1 for both markers). The effect of these two markers was nominally protective in the knee osteoarthritis meta-analysis results, although only for rs3204689 (OR = 0.95, *P* = 0.044). The lack of association with knee osteoarthritis is surprising in light of previous findings of correlations between hand osteoarthritis and knee osteoarthritis[19-21].

The *ALDH1A2* gene encodes the retinaldehyde dehydrogenase 2 (RALDH2), an enzyme that catalyzes the synthesis of retinoic acid from retinaldehyde (retinal). Retinoic acid, an active derivative of vitamin A (retinol), is a hormonal signaling molecule that plays an essential role in embryonic development[22] and in maintenance of adult tissues[23], including the cartilage and skeleton[24-26]. The spectrum of activities of retinoic acid is broad and complex, controlling expression of an array of target genes through binding nuclear receptors (retinoic acid (RAR) and retinoid X (RXR) receptors), PPAR (peroxisome proliferator-activated receptor) β/δ, or through direct activation of kinase cascades[27,28]. A tight control of both spatial and temporal expression of all players in the retinoid metabolism pathway, and retinoic acid concentration in particular, is observed during development[22,29], indicating a role for retinoic acid as a morphogen, supported by the observations that both an excess and deficiency of retinoic acid at different time points can cause developmental anomalies[22,30]. *Raldh2*$^{-/-}$ mice display numerous abnormalities and die before midgestation. Limb bud patterning is one of their many abnormalities presented in hypoplastic or absent forelimb buds[31]. However, retinoic acid does not seem necessary for hindlimb development[32].

We explored mRNA expression profile of the *ALDH1A2* gene in white blood cells, adipose tissue, and in articular cartilage harvested during hip or knee joint replacements (Online **Methods and Supplementary Information**). The gene is very weakly expressed in white blood cells and at a low level in adipose tissue. Correlation is observed between the risk variants and decreased expression level of the *ALDH1A2* transcript in adipose tissue (e.g. $P = 0.0017$, 4.1% decrease in expression for rs12907038[G]). *ALDH1A2* is highly expressed in the cartilage from hip and knee. Comparison of the level of expression in cartilage that displays signs of osteoarthritis to expression levels in cartilage that appeared macroscopically normal but isolated from the same joint (preserved cartilage) suggests lowered expression in the osteoarthritic cartilage ($P = 0.025$, **Supplementary Figure 2**). We also tested if there is an allelic imbalance in the expression of the gene in articular cartilage dependent on the genotypes of the rs3204689 SNP located in the 3'UTR of the *ALDH1A2* transcript; the at-risk C allele is expressed at a lower level than the corresponding normal transcript. Overall there is 13.6% ($P = 1.1\times10^{-30}$) more of the G allele than the at-risk allele C (**Supplementary Figure 3)**.

We improved the imputation of the rare variants (0.02%) that associate at the 1p31 locus by direct genotyping (**Online Methods**); this resulted in slightly stronger association with severe hand osteoarthritis (OR = 50.6 and $P = 9.8 \times 10^{-10}$ for chr1:63807756) (**Supplementary Table 5)** and with polyarthritis (OR = 33.0 and $P = 8.2 \times 10^{-8}$). Among the 53 carriers found in the Icelandic dataset is a family with several members in whom extremely severe hand osteoarthritis segregates with the risk allele. They also display osteoarthritis at multiple joints (**Figure 2**). All the affected carriers in the dataset share ancestors, a couple born in 1550, and all share a 912 kb region on 1p31 that encompass the three associated variants, seven genes and

one pseudogene (**Supplementary Figure 5**). We could not find any variant within these genes that the three associated intergenic variants tag even by carefully examining the WGS bam-files or by re-sequencing the 13 exons that were not fully covered in the two carriers that are WGS sequenced. Neither did exome sequencing of 9 carriers and 9 non-carriers of chr1:63807756 reveal any tagging variant within this shared region. Of the three associated variants, chr1:63807756, is predicted by ESPERR[33] to be within a functional element or an unknown exon. The SNP is within intron 12 of *EFCAB7*, and 46 and 53.7 kb upstream of *ITGB3BP* and *PGM1*, respectively (**Supplementary Figure 4**). Furthermore, intron 12 of *EFCAB7* is not spliced out in one transcript (Ensemble ENST00000461039) and RNA sequencing also indicates that chr1:63807756 is within a potential exon in undifferentiated chondrocytes from knee joint [34]. Additionally, the variant overlaps MeDIP-seq methylated CpG from brain sample (**Online Methods and Supplementary Information**). In light of the re-sequencing results and bioinformatics indication, this variant may be the functional variant that drives the association signal. However, this remains to be examined further. The chr1:63807756 marker is present in samples from Sweden, the United Kingdom and the Netherlands in a slightly higher frequency than in Iceland, or 0.1%. The hand osteoarthritis status is not known for all the fifteen carriers found in these samples, although two are known to have hand osteoarthritis and four have osteoarthritis of the knee. See **Supplementary Information** for more data on the 1p31 locus.

We report here the results from our genome wide association study on severe hand osteoarthritis, using both common and rare sequence variants in the genome. Variants within the *ALDH1A2* gene associate significantly with severe osteoarthritis of the hand, and yield the first genome wide significant locus for hand osteoarthritis that replicates consistently; in five

European sample sets. These variants confer relatively high odds ratios for common variants; 1.43-1.46. The at-risk allelic variant of the gene is expressed at a lower level in cartilage than its counterpart. We also describe a rare variant at 1p31 (0.02%, OR 50) that segregates with severe hand osteoarthritis, and generalized osteoarthritis, in one Icelandic family with large number of cases. The variant is found in other population where six out of fifteen carriers are known to have osteoarthritis of the hands or the knees.

Author contributions. The study was designed and results were interpreted by U.S., G.T., I.M., U.T. and K.S. Phenotype data and Icelandic subject recruitment was coordinated and managed by T.I. and H.J. D.H., S.M., N.K.A., T.S and A.V coordinated, managed, genotyped and analyzed the Twins UK and Chingford cohort samples. M.B., M.K., P.E.S. and I.M coordinated, managed, genotyped and analyzed the samples from the GARP, LLS, PAPRIKA and RAAK studies. B.H., F.R., A.U., S.B-Z. and J. v M. coordinated, managed, genotyped and analyzed the samples from the Rotterdam cohorts. S.L. coordinated and managed the samples from the MDC study. E.E. undertook the meta-analysis for the TreatOA consortium. G.M., U.T., G.T., and U.S. coordinated sequence data analysis, imputation and association analysis in Icelandic dataset. O.T.M. and S.A. performed exome sequencing and analysis. S.A., A.J. and A.S. carried out bioinformatics analysis at 1p31. Shared haplotype analysis was performed by M.L.F. and A.K. Sanger sequencing of Icelandic samples and Centaurus genotyping of Icelandic and Swedish samples was carried out and analyzed by H.T.H and U.S. Expression experiments were carried out and analyzed by N.B., A.M.S., R.G.H.H.N., A.J., A.S., F.F.E. and M.T. The paper was drafted by U.S., G.T. and K.S. All authors contributed to the final version of the paper**.**

Acknowledgements: We thank the subjects of the deCODE study, the Rotterdam study, the GARP, LLS, RAAK and PAPRIKA studies, the TwinsUK and Chingford studies, and the MDC study for their valuable participation. This work was supported by EC framework 7 programme grant 200800 TREAT-OA. Full acknowledgements are given in Supplementary Material.

Conflict of Interest: U.S., G.T., H.H., A.J., A.S., S.A.J., M.L.F., A.K., O.T.M., G.M., U.T., I.J. and K.S. are employed by deCODE Genetics/Amgen.

References:

1. Kerkhof, H.J. *et al.* A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis Rheum* **62**, 499-510 (2010).
2. Evangelou, E. *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* **70**, 349-55 (2011).
3. Day-Williams, A.G. *et al.* A variant in MCF2L is associated with osteoarthritis. *Am J Hum Genet* **89**, 446-50 (2011).
4. Castano Betancourt, M.C. *et al.* Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Natl Acad Sci U S A* **109**, 8218-23 (2012).
5. Miyamoto, Y. *et al.* Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat Genet* **40**, 994 (2008).
6. Nakajima, M. *et al.* New Sequence Variants in HLA Class II/III Region Associated with Susceptibility to Knee Osteoarthritis Identified by Genome-Wide Association Study. *PLoS ONE* **5**, e9723 (2010).
7. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815-23 (2012).
8. Loeser, R.F., Goldring, S.R., Scanzello, C.R. & Goldring, M.B. Osteoarthritis: A disease of the joint as an organ. *Arthritis Rheum* **64**, 1697-707 (2012).
9. Jonsson, H. *et al.* The inheritance of hand osteoarthritis in Iceland. *Arthritis Rheum* **48**, 391-5 (2003).
10. Spector, T.D. & MacGregor, A.J. Risk factors for osteoarthritis: genetics. *Osteoarthritis Cartilage* **12 Suppl A**, S39-44 (2004).
11. Kraus, V.B. *et al.* The Genetics of Generalized Osteoarthritis (GOGO) study: study design and evaluation of osteoarthritis phenotypes. *Osteoarthritis Cartilage* **15**, 120-7 (2007).
12. Haugen, I.K. *et al.* Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. *Annals of the Rheumatic Diseases* **70**, 1581-1586 (2011).
13. Riyazi, N. *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* **64**, 438-43 (2005).
14. Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84 (2006).
15. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *European Journal of Epidemiology* **26**, 657-686 (2011).
16. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899-906 (2006).
17. Hart, D.J. & Spector, T.D. Cigarette smoking and risk of osteoarthritis in women in the general population: the Chingford study. *Ann Rheum Dis* **52**, 93-6 (1993).
18. Evangelou, E. *et al.* A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Annals of the Rheumatic Diseases* (2013).

19. Dahaghin, S. *et al.* Does hand osteoarthritis predict future hip or knee osteoarthritis? *Arthritis Rheum* **52**, 3520-7 (2005).
20. Jonsson, H. *et al.* Hand Osteoarthritis Severity is Associated with Total Knee Joint Replacements Independently of BMI. The Ages-Reykjavik Study. *Open Rheumatol J* **5**, 7-12 (2011).
21. Valdes, A.M. *et al.* Involvement of different risk factors in clinically severe large joint osteoarthritis according to the presence of hand interphalangeal nodes. *Arthritis Rheum* **62**, 2688-95 (2010).
22. Rhinn, M. & Dolle, P. Retinoic acid signalling during development. *Development* **139**, 843-58 (2012).
23. Gudas, L.J. Emerging roles for retinoids in regeneration and differentiation in normal and disease states. *Biochim Biophys Acta* **1821**, 213-21 (2012).
24. Underhill, T.M. & Weston, A.D. Retinoids and their receptors in skeletal development. *Microsc Res Tech* **43**, 137-55 (1998).
25. Weston, A.D., Hoffman, L.M. & Underhill, T.M. Revisiting the role of retinoid signaling in skeletal development. *Birth Defects Res C Embryo Today* **69**, 156-73 (2003).
26. Laue, K. *et al.* Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. *Am J Hum Genet* **89**, 595-606 (2011).
27. Theodosiou, M., Laudet, V. & Schubert, M. From carrot to clinic: an overview of the retinoic acid signaling pathway. *Cellular and Molecular Life Sciences* **67**, 1423-1445 (2010).
28. Al Tanoury, Z., Piskunov, A. & Rochette-Egly, C. Vitamin A and retinoid signaling: genomic and non-genomic effects. *Journal of Lipid Research* (2013).
29. Pittlik, S. & Begemann, G. New sources of retinoic acid synthesis revealed by live imaging of an Aldh1a2-GFP reporter fusion protein throughout zebrafish development. *Dev Dyn* **241**, 1205-16 (2012).
30. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134**, 921-31 (2008).
31. Niederreither, K., Subbarayan, V., Dolle, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* **21**, 444-8 (1999).
32. Zhao, X. *et al.* Retinoic acid promotes limb induction through effects on body axis extension but is unnecessary for limb patterning. *Curr Biol* **19**, 1050-7 (2009).
33. Taylor, J. *et al.* ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. *Genome Res* **16**, 1596-604 (2006).
34. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046 (2011).
35. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* **22**, 719-48 (1959).
36. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099-103 (2010).
37. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).

Tables:

Table 1. Markers at 15q22, rs4238326-C (freq. 41%) and rs3204689-C (freq. 52%), associate with severe hand osteoarthritis

Hand OA phenotype	marker	Discovery			Replication			Overall	
		OR	*P* value	N cases /controls	OR	*P* value	N cases /controls	OR (95% CI)	*P* value
Severe thumbs and severe fingers	rs4238326	1.48	2.3×10^{-9}	623 / 69,153	1.33	0.0075	214 / 8,172	1.44 (1.29-1.60)	8.6×10^{-11}
	rs3204689	1.48	2.5×10^{-9}		1.41	0.0011		1.46 (1.31-1.63)	1.1×10^{-11}
Severe Fingers	rs4238326	1.27	2.4×10^{-9}	1,935 / 71,595	1.22	0.0017	713 / 8,095	1.25 (1.17-1.34)	2.0×10^{-11}
	rs3204689	1.26	7.7×10^{-9}		1.26	0.000017		1.26 (1.18-1.34)	5.6×10^{-12}
Severe Thumbs	rs4238326	1.23	1.8×10^{-6}	1,610 / 74,060	1.19	0.0021	927 / 7,913	1.21 (1.14-1.30)	1.5×10^{-8}
	rs3204689	1.21	8.4×10^{-6}		1.14	0.014		1.18 (1.11-1.27)	5.1×10^{-7}

The phenotypes of associations are shown, number of cases and controls, along with their respective *P* values and odds ratios (OR) and the 95% confidence interval (CI). The *P* values are corrected for genomic controls. The results are shown for the Icelandic discovery set, the replication sets combined (Rotterdam I and II, Garp, TwinsUK and Chingford) and the overall results for the discovery and replication combined. Results from multiple case-control groups were combined using a Mantel-Haenszel model[35].

Table 2. rs4238326 and rs3204689 are not associated with hip or knee osteoarthritis

Osteoarthritis phenotype	Marker	OR (95% CI)	*P* value	Cases	Controls
Hip osteoarthritis	rs4238326-C	1.04 (0.99-1.09)	0.10	4,349	46,903
	rs3204689-C	1.04 (0.99-1.09)	0.11		
Knee osteoarthritis	rs4238326-C	0.96 (0.91-1.01)	0.10	5,224	48,571
	rs3204689-C	0.95 (0.90-1.00)	0.044		

The results from meta-analyses of hip osteoarthritis and knee-osteoarthritis conducted by the TreatOA consortium. The *P* values are corrected for the overall meta-analyses genomic controls, λ = 1.028 for hip and λ = 1.036 for knee.

Figure 1. Regional association plot for the 15q22 *ALDH1A2* locus

The *P* values ($-\log_{10}$) of SNP association with severe hand osteoarthritis in the Icelandic discovery samples are plotted against their positions at the 15q22 locus. The purple circle highlights the most significant SNP, rs12907038, in the discovery analysis. The SNPs are color coded to reflect their LD with rs12907038 in the Icelandic dataset. Red line indicates recombination rates, based on the Icelandic recombination map for male and females combined[36], with the peaks indicating recombination hotspots defining LD blocks in Icelanders. Known genes in the region are shown underneath the plot, taken from the UCSC genes track in the UCSC genome browser. All positions are in NCBI Build 36 coordinates. The plot was created using an standalone version of the LocusZoom software[37] (http://csg.sph.umich.edu/locuszoom/).

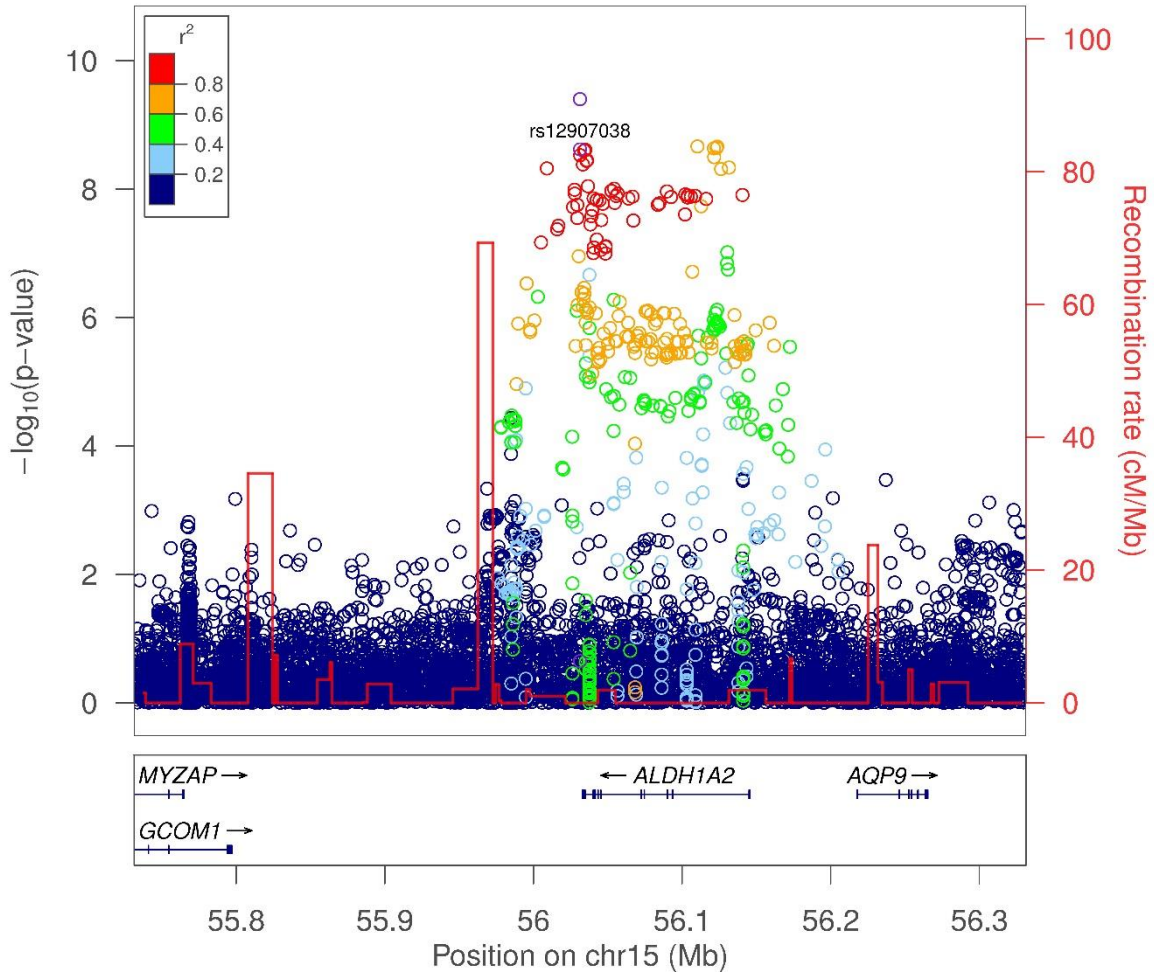
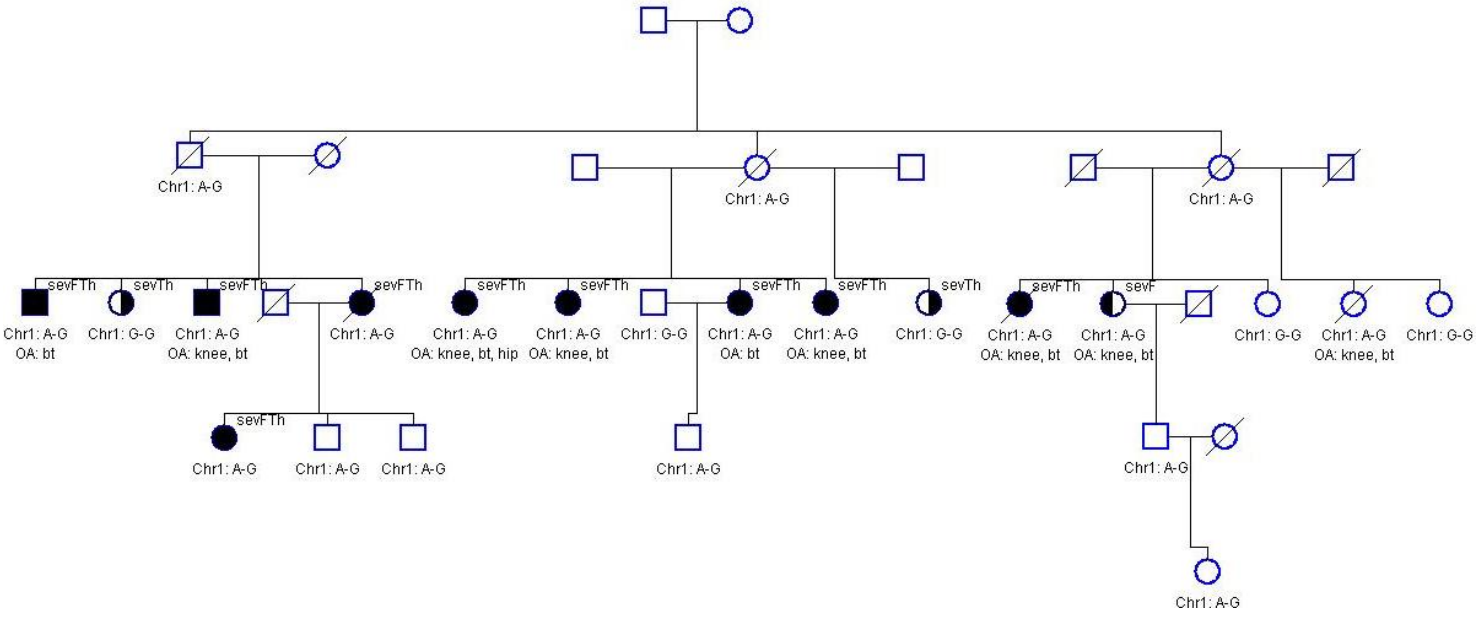


Figure 2. Pedigree segregating the rare variants at 1p31.

Family segregating the rare 1p31 variants. Chr1:A-G and Chr1:G-G refer to the genotypes of chr1:63807756; Chr1:A-G are carriers and Chr1:G-G are non-carriers. Severe fingers and thumbs phenotype (sevFTh) is presented by filled symbols, severe thumbs (sevTh) and severe fingers (sevF) by vertical half-filled symbols. Osteoarthritis at other joints is shown below the symbols, at knee, hip and the big toe (bt). All individuals in generation III that have phenotype or genotype information are shown.



ONLINE METHODS:

Study populations: We used six study populations for the *ALDH1A2* gene study: a discovery set of 69,786 Icelanders and five *in-silico* replication sets; two sample sets from the United Kingdom (Twins UK study[16] of 1,557 women and Chingford study[17] of 578 women), three Dutch samples (the Garp[13]/LLS studies[14] of 2,475 subjects, the Rotterdam-I study[15] of 3,137 persons, and Rotterdam-II[15] study of 1,184 persons). For the 1p31 locus we used nine study populations; the same discovery set of 69,786 Icelanders and eight additional samples to investigate if the marker exists in other populations, some with limited information on severe hand osteoarthritis; Twins UK cohort (n=526), the Dutch Rotterdam III cohort (n=3,068), the GARP[13] (n=490), LLS[14] (n=2,456), RAAK (n=150) and PAPRIKA (n=971) studies, and Swedish samples of 1,787 individuals (MDC study).

Hand osteoarthritis phenotype: The hand osteoarthritis patients included in the study were confined to those considered to have severe hand osteoarthritis. In the scoring of the fingers in the Icelandic discovery set, main emphasis was on affection on distal interphalangeal joints (DIP) 2 and 3 bilaterally with proximal interphalangeal joints (PIP) 2 and 3 contributing to the assessment and less emphasis on other joints. Joint count, severity and bilaterality contribute to severity. In the scoring of the thumb base, similar principles apply but bilaterality was not a requisite for diagnosis. Prior surgery of the thumb base was automatically registered as severe thumb base hand osteoarthritis. The clinical severity scoring was adjusted for age and thus, slightly different comparison criteria had to be used according to the age at examination (**Supplementary Information**). The severe hand osteoarthritis cases in the replication sets had Kellgren Lawrence score ≥ 3 in at least 1 DIP bilateral and Kellgren Lawrence score ≥ 3 in at least 1 carpometacarpal joint of the thumb. Further description of characteristics of the study populations is provided in the **Supplementary Information**. All participants provided written informed consent, and we obtained approval from all institutional review board to carry out the study.

Genotyping and association analysis of Icelandic samples: The genotyping, imputation methods and association analysis method in the Icelandic samples were as described

previously[38]. In short, about 34.2 million sequence variants (30.6 million SNPs and 3.6 million indels) were identified through whole-genome sequencing of 2,230 Icelanders (to an average sequencing depth of > 10X) using Illumina GAIIx and HiSeq2000 instruments and imputed into 95,085 Icelanders genotyped with Illumina SNP chips and phased using long-range phasing[39,40]. With the use of the Icelandic genealogy the probabilities of genotypes were furthermore predicted for 296,526 close relatives of the 95,085 chip-typed individuals using the nationwide Icelandic genealogy database. Association testing for case–control analysis was performed using logistic regression, matching controls to cases based on how informative the imputed genotypes were. In order to account for relatedness and stratification within the case and control sample sets we applied the method of genomic control[41]. The inflation λ_g in the chi^2 statistic in each genome-wide analysis was estimated based on a subset of about 300,000 common variants and the *P* values adjusted by dividing the corresponding chi^2 values by this factor. For the traits reported in this paper the estimated inflation factors were: 1.17 for severe fingers and severe thumbs, 1.36 for severe fingers, and 1.23 for severe thumbs.

The threshold for genome-wide significance was set at a *P* value less than 1.5×10^{-9} (= 0.05/34.2 million variants), a conservative estimate as many of the variants tested are highly correlated.

Single SNP genotyping of markers at 1p31 was carried out by the Centaurus (Nanogen) platform and by Sanger sequencing.

Genotyping and association analysis of replication samples: The replication sets were genotyped with Illumina's HumanHap 300, HumanHap550 or HumanHap610 SNP chips and the two replication SNPs on 15q22 were looked-up in this data (**Supplementary Information**). Association testing for case–control analysis was performed using logistic regression. The marker chr1:63807756 was genotyped with Centaurus assay in the MDC study, and with TaqMan assay (Applied Biosystems) in the GARP, PAPRIKA, LLS, RAAK and Rotterdam studies. A look-up was performed in 526 subjects in the TwinsUK cohort that had been whole-genome-sequenced.

***Meta-analysis*:** Results from multiple case-control groups were combined using a Mantel-Haenszel model[35] in which the groups were allowed to have different population frequencies for

alleles and genotypes but were assumed to have common relative risks (a fixed-effect model). Heterogeneity in the effect estimate was tested assuming that the estimated ORs for different groups follow a log-normal distribution and using a likelihood ratio chi^2 test with degrees of freedom equal to the number of groups compared minus 1.

Conditional analysis at 15q22: When doing the conditional analysis, testing a variant conditional on the expected genotypes of another variant, for the Icelandic dataset this was done using the same software that was used in the GWAs analysis[38], including both the chip-typed individuals and those with genotype probabilities predicted based on genotyped relatives. For the replication sample sets - Rotterdam I and II, The GARP study, The Twins UK and the Chingford studies - the conditional analysis was done using the R software (http://www.r-project.org/).

TreatOA meta-analysis: The hip osteoarthritis meta-analysis is as described in Evangelou[42] and the knee osteoarthritis meta-analysis is an extension to a previously published meta-analysis by the TreatOA consortium[2]. Definition of the hip osteoarthritis and knee osteoarthritis phenotypes[43], the study samples and whole genome genotyping platforms, quality control measures and imputation methods have been described previously[1,2,7,42,44]. The studies included and their sample sizes are shown in **Supplementary Table 4**. The hip osteoarthritis meta-analysis included 4,349 cases and 46,903 controls from 8 sample-sets from 4 populations, and the knee osteoarthritis meta-analysis included 5,224 cases and 48,571 controls from 9 sample-sets from 5 populations. All samples are of Northern European descent. Genomic control was applied to each study before meta-analysis. The effect estimates of each study were synthesized using an additive model and meta-analyzed using inverse variance fixed-effect models. The *P* values reported for the hand osteoarthritis SNPs at 15q22 have been corrected for the meta-analyses overall genomic controls, $\lambda = 1.028$ for hip osteoarthritis and $\lambda =$ for 1.036 for knee osteoarthritis. No between-study heterogeneity was observed for either SNP ($I^2 = 0$).

Expression analysis of ALDH1A2 gene in cartilage: The cartilage samples were derived from the ongoing Research Articular osteoArthritis Cartilage (RAAK) study (**Supplementary Information**). In the current study, we used paired preserved and osteoarthritis affected

cartilage samples of donors undergoing joint replacement surgery for primary osteoarthritis in either knee or hip (**Supplementary Table 6 and Supplementary Information**). DNA was isolated using the Promega Wizard Genomic DNA Purification kit according to the manufacturer's protocol. The RNA was collected using Qiagen mini columns according to the manufacturer's protocol. RNA was processed with the First Strand cDNA Synthesis Kit according to the manufacturer's protocol (Roche Applied Science, Almere, The Netherlands). RT-qPCR measurements were performed on the Roche Lightcycler 480 II, using Fast Start Sybr Green Master reaction mix according to the manufacturer's protocol (Roche Applied Science). An allele-specific realtime Taqman assay (C___3232487_10, Life technologies) was used to quantify the allelic ratio of *ALDH1A2* (rs3204689) in heterozygous samples. The allelic ratios were calculated using the formula ($2^{-FAM\ Ct}$ / $2^{-VIC\ Ct}$)[45,46]. The ratios between the amounts of each allele in every sample were calculated for genomic DNA and cDNA. For each sample the average allelic ratio for genomic DNA, which represents the 1:1 ratio, was used to normalize the cDNA ratio to generate a corrected allelic ratio. To determine if there was an overall difference in expression between alleles for a particular tissue across all patients the mean allelic ratios for the patient cDNAs were compared to the mean allelic ratios for the patient genomic DNAs using a Student T-test. Overall expression of *ALDH1A2* expression was determined by Illumina HT-12 V3 microarrays using standard methods. There were 2 probes, approximately 499 base pairs apart, on the array used for *ALDH1A2* (**Supplementary Information and Supplementary Figure 2**).

Expression analysis of ALDH1A2 gene in white blood cells and in adipose tissue: We investigated expression of *ALDH1A2* in a dataset that includes RNA samples from white blood cells of 1,002 Icelandic individuals and from adipose tissues of 673 individuals[47]. Most of those individuals, 973 with white blood samples and 646 with adipose tissue samples, had imputed genotypes for the 34.2 million variants identified in the WGS. The correlation between expression and genotypes of the variants was tested by regressing the measured MLRs (mean log expression ratio) on the number of copies of the risk allele an individual carries. The effects of age and gender were taken into account by including those variables as explanatory variables. For white blood we adjusted also for differential blood cell count as those variables

correlated strongly with the expression of a large fraction of the genes measured[47]. All *P* values were adjusted for the relatedness of the individuals by simulating genotypes through the Icelandic genealogy as previously described[48]. The resulting adjustment factors for the chi^2 statistic were 1.08 and 1.06 for adipose and whole blood, respectively. The gene is expressed below reliable detection limits in white blood cells. In adipose tissue the gene is expressed at moderate level and rs12907038[G] is correlated with the expression level, with each risk allele decreasing the expression by about 4% ($P = 0.0017$). Although this correlation is not very strong, no other common variant located in or within 100 kb away from *ALDH1A2,* and not highly correlated with rs12907038[G]), is more strongly correlated with its expression.

Re-imputation of variants at the 1p31 locus: We validated the imputation of the rare variants at the 1p31 locus that associated with severe hand osteoarthritis with $P < 1\times10^{-7}$ in the initial genome-wide association scan by directly genotyping likely and possible carriers and predicted non-carriers by Sanger sequencing or by Centaurus assays. The directly assessed genotypes were added to the genotypes for the 2,230 sequenced individuals and the combined set used as a training set for re-imputing the variants. This improved the imputation information for all four markers and the association based on re-imputed genotypes with severe hand osteoarthritis is slightly stronger for three of them, whereas one is no longer associated (**Supplementary Table 5**).

Sanger sequencing of exons in 1p31 shared region. Dye-terminator Sanger sequencing was performed with the Applied Biosystems BigDye Terminator v3.1 Cycle Sequencing Kit with Agencourt Ampure XP and Agencourt CleanSeq for cleanup of the PCR and cycle sequencing product, respectively. AmpureXP and CleanSeq bead cleaning was performed on Zymark Sciclone ALH-500 liquid handling robot system. Tray dilutions for PCR setup and cycle sequencing setup were prepared on Packard MultiprobeII HTEX liquid handling robot system, and the genomic DNA and PCR product were transferred into the respective trays using Zymark Sciclone ALH-500. PCR and cycle sequencing reactions were performed on MJ Research PTC-225 thermal cyclers. For signal detection, Applied Biosystems 3730xl DNA analyzers were used.

Exome Sequencing of chr1:63807756 carriers: DNA samples isolated from blood were prepared for exome sequencing using the Nextera Exome Enrichment method from Illumina. In short, 50ng of gDNA is „tagmented" using the Nextera transposome complex, generating adaptor ligated DNA fragments ready for enrichment. All samples were barcode/index labelled as a part of the tagmentation process. Samples were quantified using Picogreen measurements and the quality of each sample was assessed using the Agilent BioAnalyzer. Up to twelve samples were pooled together in equal quantities before capture. A bait library of >340,000, 95-mer probes was used to enrich a 62Mb region of the genome containing coding exons (>200,000 exons) of 20,794 genes, UTR´s and some ncRNA. The biotinylated probes were used to hybridize to the target sequences, following capture using streptavidin beads. The captured DNA was further amplified using PCR and the quality of the pooled, enriched sequencing libraries was measured on the BioAnalyzer. Further quality assessment was done by sequencing each twelve sample pool on the Illumina MiSeq instrument. This was done to optimize cluster densities for further sequencing and assess the proportion of reads from each sample within the pool. Sequencing of pooled Nextera libraries was done on the Illumina HiSeq 2000 instrument. In general, each pool (12 samples) were sequenced on four lanes on the HiSeq, generating approximately 150 Gb (10-15 Gb per sample) of raw data.

Bioinformatics analysis of 1p31: We carried out a search for overlaps between variant locations and known bioinformatics features. We retrieved data from University of California Santa Cruz (UCSC) test browser (HG19 build 37)[49] and UCSC main browser (HG18 build 36). We accessed feature tracks from tissues that may be relevant to hand osteoarthritis, and other tissues, containing genome positional information and identified those features that overlapped with the SNP[33,50].

References:

38. Styrkarsdottir, U. *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497**, 517–520 (2013).
39. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* (2008).

40. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868-74 (2009).
41. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
42. Evangelou, E. *et al.* A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Annals of the Rheumatic Diseases* **submitted**(2013).
43. Kerkhof, H.J.M. *et al.* Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. *Osteoarthritis and Cartilage* **19**, 254-264 (2011).
44. Panoutsopoulou, K. *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Annals of the Rheumatic Diseases* **70**, 864-867 (2011).
45. Raine, E.V.A., Dodd, A., Reynard, L. & Loughlin, J. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC Musculoskeletal Disorders* **14**, 85 (2013).
46. Bos, S.D. *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Annals of the Rheumatic Diseases* **71**, 1254-1258 (2012).
47. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423-8 (2008).
48. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat Genet* **37**, 129-137 (2005).
49. Meyer, L.R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Research* **41**, D64-D69 (2013).
50. Rosenbloom, K.R. *et al.* ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research* **41**, D56-D63 (2013).

Supplementary information: Severe osteoarthritis of the hand associates with markers within the *ALDH1A2* gene

Study subjects:

Icelandic samples:

A database registering hand osteoarthritis (HOA) in Iceland was started in 1992. There are now approximately 8,500 individuals registered with HOA in this database. Most of these represent people seeking medical service for HOA through the socialized medical system (doctors' visits, radiographs, physiotherapists, ergotherapists) but in cases with family history, discovered through medical history taking or the Icelandic Genealogy Database, a number of family members have also been recruited. Approximately 65% of this group has been examined by one of two examiners (HJ, TI) and more than 75% have hand radiographs. In cases with multiple recordings, the most recent one was use1063666d in this study.

The same severity scoring of HOA has been used from the beginning, grading HOA as mild and severe at two sites, fingers and thumb base[1]. In the scoring of the fingers, main emphasis has been on affection on distal interphalangeal joints (DIP) 2 and 3 bilaterally with proximal interphalangeal joints (PIP) 2 and 3 contributing to the assessment and less emphasis on other joints. Joint count and severity and bilaterality affected the mild versus severe estimate. In the scoring of the thumb base, similar principles applied but bilaterality was not a requisite for diagnosis. Prior surgery of the thumb base was automatically registered as severe thumb base HOA.

To enable comparison between the clinical and radiographic HOA data, certain generalizations had to be made based on experience, and on previous studies where both clinical examination and radiographs were available[2,3]. The clinical severity scoring was adjusted for age and thus, slightly different comparison criteria had to be used according to the age at examination (Table 1).

Table1. Age adjustment and comparison between clinical and radiographic assessment

Fingers		
Age	*Mild OA*	*Severe OA*
≤45	Clear affection (KL≥2) of >1 joint	bilateral KL≥2 (in dip23 mainly but others add)
46-55	bilateral KL≥2	bilateral KL≥2+ (many joints or KL≥3, EHOA)
56+	bilateral KL≥2	bilateral KL≥2+ (many joints or KL≥34, EHOA)
Thumb base		
Age	*Mild OA*	*Severe OA*
≤45	Clear affection(KL≥2 in one joint)	KL≥3 or bilateral KL 1 and 2
46-55	KL≥2 (one joint)	KL≥3 or bilateral KL≥2
56+	KL≥2 (one joint)	KL≥4 (one joint) or KL≥3 plus other side

Patients who had undergone total hip replacement (THR) or total knee replacement (TKR) for primary OA of the hip/knee were identified through a computer-aided search of hospital records from all 6 orthopedic clinics in Iceland. The patients' records were reviewed by a clinician to ascertain a correct diagnosis, excluding fracture and rheumatoid arthritis.

The study was approved by the Data Protection Authority of Iceland and the National Bioethics Committee of Iceland. Informed consent was obtained from all participants.

The Rotterdam Study: the study population comprises men and women aged 45 years and older of the Rotterdam Study, which is a prospective population-based study on determinants of chronic disabling diseases. The rationale and study design have been described previously[4]. The severe hand OA cases had KL ≥3 in at least 1 DIP bilateral and KL≥3 in at least 1 carpometacarpal joint of the thumb (CMC). Control individuals had no bilateral KL≥2 in DIP AND no KL≥2 in CMC joint. The medical ethics committee of Erasmus University Medical School approved the study and written informed consent was obtained from each participant.

TwinsUK: the study participants were white monozygotic and dizygotic twin pairs from the TwinsUK adult twin registry, a group used to study the heritability and genetics of age-related diseases[5]. These unselected twins were recruited from the general population through national media campaigns in the United Kingdom. Ethics approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant. The severe hand OA cases had KL ≥3 in at least 1 DIP bilateral and KL≥3 in at least 1 carpometacarpal joint of the thumb (CMC). Control individuals had no bilateral KL≥2 in DIP AND no KL≥2 in CMC joint.

The Chingford Study: is a prospective population-based longitudinal cohort, which includes women derived from the age/sex register of a large general practice in North London. The study design and rationale have been described elsewhere in detail[6]. The Guy's St. Thomas' Trust and the Waltham Forest Trust ethics committees approved the study protocol. After study procedures were explained to participants, written informed consent was given by each participant. The severe hand OA cases had KL ≥3 in at least 1 DIP bilateral and KL≥3 in at least 1 carpometacarpal joint of the thumb (CMC). Control individuals had no bilateral KL≥2 in DIP AND no KL≥2 in CMC joint.

The Genetics OsteoArthritis and Progression (GARP) study from Leiden, the Netherlands, consists of 192 sibling pairs concordant for clinical and radiographically (K/L score) confirmed OA at two or more joint sites among hand, spine (cervical or lumbar), knee or hip. Details of the ascertainment have been described elsewhere[7]. Written informed consent was obtained from each subject as approved by the ethical committees of the Leiden University Medical Center. The GARP study consists of 382 subjects (312 females and 70 men) who met all inclusion criteria; in this paper we included subjects with OA hand. The severe hand OA cases had KL ≥3 in at least 1 DIP bilateral and KL≥3 in at least 1 carpometacarpal joint of the thumb (CMC). As controls we used ~2374 subjects of the Leiden Longevity study (LLS)[8].

The Leiden Longevity Study: The Leiden Longevity Study is a longitudinal cohort consisting of 421 families of long-lived Caucasian siblings of Dutch descent together with their offspring and the partners

thereof[9]. The partners of the offspring were included as controls being of comparable age and sharing the same socioeconomic and geographical background as the offspring. Families were recruited if at least two long-lived siblings were alive and fulfilled the age-criterion of 89 years for males and 91 years for females. Sex-specific age-criteria were used due to the higher life-expectancy of females compared to males. No selection criteria on health or demographic characteristics were applied.

RAAK study: The ongoing Research Arthritis and Articular Cartilage (RAAK) study is approved by the ethical committee and is aimed at the biobanking of joint materials (cartilage, bone and where available ligaments) and mesenchimal stem cells (hip joints only) and primary chondrocytes of patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. In the current study we collected such sample pairs for 33 donors undergoing joint replacement surgery for primary OA, 22 hips and 11 knees were included in the study. Joints were assessed macroscopically for OA related damage such as the loss of articular cartilage, fibrillation or crack formation and whiteness of the cartilage.

Paprika study: The Paprika study is performed at the Leiden University Medical Center (Dept. Orthopedics) and consists in a long-term follow-up study of patients that have undergone total joint replacement (TJR) at hip or knee[10] and has been approved by the medical ethical committee. Patients of Caucasian descent were included when they were diagnosed with primary osteoarthritis based on radiographs and the ACR rheumatology classification criteria (mean age males-hip: 66; years males-knee: 68 years; females-hip: 66 years; females-knee: 69 years). Patients with secondary OA or requiring a revision were excluded in this study. Written consent was obtained from each participant.

Swedish Malmo Diet and Cancer (MDC) study: All men and women living in the city of Malmö in Sweden, who were born between 1923 and 1945 (men) or between 1923 and 1950 (women), were invited to participate in the Malmö Diet and Cancer (MDC) study. The screening examination was performed during 1991-1996. All participants (n=28,449) were followed until first OA surgery, emigration from Sweden, death or December 31 2005, whichever came first. Information on knee and hip arthroplasty for OA and mortality were based on record linkage with the national Swedish Hospital Discharge Register and the Swedish Causes of Death Register. Cases were defined as those who during the follow-up time were treated with knee or hip arthroplasty (421 and 551, respectively). Controls were identified in the study population matching each arthroplasty case for age, gender and BMI. The research ethical committee at Lund University approved the MDC study (LU 51–90). Each participant signed a written informed consent.

Genotyping:

The Rotterdam Study: Genotyping of the samples with the Illumina HumanHap550v3 Genotyping BeadChip was carried out at the Genetic Laboratory of the Department of Internal Medicine of Erasmus Medical Center, Rotterdam, the Netherlands. The Beadstudio GenCall algorithm was used for genotype calling and quality control procedures were as described previously[11]. All of the SNPs tested in this report passed quality filtering (SNP call rate >= 95%, minor allele frequency >= 5%, p-value HWE >=1×10^{-6}).

Genomic inflation factors were calculated for all analyses and there was no evidence of population stratification.

GARP: Genotyping of the GARP and LLS study was done by Illumina Infinium HumanHap610. The following sample QC criteria were applied sample call rate >97.5% Hardy-Weinberg p-values<10^{-4} and minor allele frequencies < 5% were excluded. All SNPs analyzed were directly genotyped. We used generalized estimating equations (GEE)[12] for the association analyses between the hand phenotypes and the genotypes as implemented in SPSS 18.0. The GEE methodology provides a method of analyzing correlated data that otherwise could be modeled as a generalized linear model. By applying this method we were able to effectively adjust for the familial dependencies of the sibling pairs[13].

TwinsUK: Samples were genotyped with the Infinium HumanHap 300 assay (Illumina, San Diego, USA) at the Duke University Genotyping Center (NC USA), Helsinki University (Finland) and the Wellcome Trust Sanger Institute. The Illuminus calling algorithm was used for genotype calling. Imputation was performed using the IMPUTE software (v0.2.0)[14].

The Chingford Study: The Chingford samples were genotyped using the Illumina HumanHap610Q array. The normalised intensity data was then used by the Illluminus calling algorithm[15] to assign genotypes. No calls were assigned if an individual's most likely genotyped was called with less than a posterior probability threshold of 0.95. Sample exclusion criteria were: (i) sample call rate <98%, (ii) heterozygosity across all SNPs ≥2 s.d. from the sample mean; (iii) evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations; (iv) observed pair-wise IBD probabilities suggestive of sample identity errors; (v). SNPs. Exclusion criteria were (i) Hardy-Weinberg p-value<10−6, assessed in a set of unrelated samples; (ii) MAF<1%, assessed in a set of unrelated samples; (iii) SNP call rate <97% (SNPs with MAF≥5%) or < 99% (for 1% ≤ MAF < 5%).

Expression analysis of *ALDH1A2* gene in cartilage:

Samples: The cartilage samples were derived from the ongoing Research Articular osteoArthritis Cartilage (RAAK) study (**Supplementary Information**). In the current study, we used paired preserved and OA affected cartilage samples of donors undergoing joint replacement surgery for primary OA in either knee or hip. At the moment of collection (within 2 hours following surgery) tissue was washed extensively with phosphate buffered saline (PBS) to decrease the risk of contamination by blood, and cartilage was collected of the weight-baring area of the joint. Cartilage was classified macroscopically and collected separately for macroscopically OA affected and preserved regions. Classification was done according to predefined features for OA related damage as described previously: [1, 2] based on color/whiteness of the cartilage, based on surface integrity as determined by visible fibrillation/crack formation, and based on depth and hardness of the cartilage upon sampling with a scalpel. During collection with a scalpel, care was taken to avoid contamination with bone or synovium. Collected cartilage was snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. We histologically assessed cartilage samples with the modified Mankin scoring system[3, 4].

Nucleic acid isolation and genotyping: Cartilage samples were pulverized using a Retsch MM200 under cryogenic conditions. On average 150 mg of pulverized cartilage was dissolved in 1 ml of Trizol reagent, and mixed vigorously. After addition of 200µl of chloroform the sample was mixed and centrifuged for 15 minutes at 16.000g. The clear aqueous layer was transferred to a new vial and 1 volume of 70% ethanol/DEPC-treated water was added to precipitate RNA. The RNA was collected using Qiagen mini columns according to the manufacturer's protocol. DNA was isolated using the Promega Wizard Genomic DNA Purification kit according to the manufacturer's protocol. Samples were genotyped for rs3204689 using using the transcript SNPs' quantitative real time Taqman genotyping assays (C___3232487_10, Life technologies) on a Roche LighCycler 480 II system.

Real time quantitative reverse transcription PCR: RNA was processed with the First Strand cDNA Synthesis Kit according to the manufacturer's protocol (Roche Applied Science, Almere, The Netherlands). RT-qPCR measurements were performed on the Roche Lightcycler 480 II, using Fast Start Sybr Green Master reaction mix according to the manufacturer's protocol (Roche Applied Science).

Allelic imbalance analysis: An allele-specific realtime Taqman assay (C___3232487_10, Life technologies) was used to quantify the allelic ratio of ALDH1A2 (rs3204689) in heterozygous samples. Genomic DNA and cDNA samples were diluted 8 times and 2 µl was used as template in the Taqman assay in a final volume of 5 µl. Samples were subjected to 10 minutes of denaturation at 95°C, and 40 cycles of 92°C for 15 seconds and 1 minute at 60°C on a LightCycler 480 (Roche). Reactions were performed with eight pipetting replicates and they were followed real-time and after cycling an end measurement of fluorescence levels was performed. The allelic ratios were calculated using the formula $(2^{-\text{FAM Ct}} / 2^{-\text{VIC Ct}})$[5, 6]. The ratios between the amounts of each allele in every sample were calculated for genomic DNA and cDNA. For each sample the average allelic ratio for genomic DNA, which represents the 1:1 ratio, was used to normalize the cDNA ratio to generate a corrected allelic ratio. To determine if there was an overall difference in expression between alleles for a particular tissue across all patients the mean allelic ratios for the patient cDNAs were compared to the mean allelic ratios for the patient genomic DNAs using a Student T-test. The capacity for each assay to discriminate between SNP alleles was verified using a standard curve performed on cDNA and DNA of varying allelic ratios (data not shown).

Gene expression: Expression was determined by Illumina HT-12 V3 microarrays using standard methods. Using the Beadstudio software the intensity values were normalized using the "rsn" option in the Lumi R-package. The corresponding signals increase exponentially with relative levels and units are light intensity (Illumina provided values). The as obtained raw probe-level data (overall mean normalized probe level value of measured genes in cartilage) were exported for analyses using Limma. As implemented in Limma, a paired t-test was used on all samples. There were 2 probes, approximately 499 bp apart, on the array used for *ALDH1A2* (**Supplementary Figure 2**).

The 1p31 locus

We validated the imputation of the four rare variants at the 1p31 locus that associated with severe hand osteoarthritis with $P < 1\times10^{-7}$ in the initial genome-wide association scan by directly genotyping likely and possible carriers and predicted non-carriers by Sanger sequencing or by Centaurus single-track genotyping assays. The agreement between the imputed and the directly assessed genotypes varied from $r^2 = 0.14$ to $r^2 = 0.86$. The directly assessed genotypes were added to the genotypes for the 2,230 sequenced individuals and the combined set used as a training set for re-imputing the variants. This improved the imputation information for all four markers and the association based on re-imputed genotypes with severe hand osteoarthritis is slightly stronger for three of them, whereas one marker is no longer associated (**Supplementary Table 5**).

Genealogical analysis revealed that all 53 carriers found in the Icelandic dataset share common ancestors 10 generations back, a couple that were born 1549 and 1550 in Southern Iceland. Furthermore, 14 of the 53 carriers cluster in one dense family. The risk alleles segregate with extremely severe hand osteoarthritis in this family. Additionally, many carriers in this family show signs of generalized osteoarthritis; displaying symptoms of osteoarthritis of the knees and big toes as well as the hands. We identified two additional carriers in the family by direct genotyping, 38 years and 53 years old, whom had not been diagnosed with hand osteoarthritis (unknown hand osteoarthritis status), resulting in 5 unknown alive carriers in the family (**Figure 2**).

The affected carriers all share a 912 kb region on 1p31 that encompass the three associated variants (**Supplementary Figure 4**). There are seven genes in this region: *LINC00466* (long intergenic non-protein coding RNA 466), *FOXD3* (forkhead box D3), *ALG6* (alpha-1,3-glucosyltransferase) *ITGB3BP* (integrin beta 3 binding protein), *EFCAB7* (EF-hand calcium binding domain 7), *PGM1* (phosphoglucomutase 1), *ROR1* (receptor tyrosine kinase-like orphan receptor 1), and one pseudogene *DLEU2L* (deleted in lymphocytic leukemia 2-like). We could not find any functional variant within these genes that the three associated variants could tag by carefully eyeballing the sequencing bam-files of the two sequenced carriers. Nor did re-sequencing of the 13 exons that were not fully covered in the two carriers that are WGS sequenced reveal any new exonic sequence variant tagging the signal. We re-sequenced 27 carriers and 59 non-carriers, 7.6 kb in total, for this purpose. We also exome sequenced 9 carriers of chr1:63807756 and 9 non-carriers for additional investigation of functional variant tagging the signal. No functional variant tagging the signal was revealed by this method either. Of the three associated variants, chr1:63807756, which is within intron 12 of *EFCAB7*, and 46 and 53.7 kb upstream of *ITGB3BP* and *PGM1,* respectively (**Supplementary Figure 4**), is predicted to be within potential functional elements by ENCODE[16] and ESPERR[17]. Therefore, and in light of the re-sequencing results and the bioinformatics analysis, this variant may be the functional variant that drives the association signal. However, this remains to be supported with experimental data.

We directly genotyped the chr1:63807756 SNP in additional European sample sets; the Swedish Malmo Diet and Cancer (MDC) study (N=1,787), the Dutch GARP[7] (N=490), LLS[8] (N=2,456), RAAK (N=150) and PAPRIKA (N=971) studies and the Rotterdam Study III[4] (N=3,068), and looked it up in WGS data from the UK Twins[5] cohort (N=526).

The chr1:63807756 SNP is present in these samples in slightly higher frequency than in Iceland, or around 0.1%. The hand osteoarthritis status of the four identified carriers in the Swedish osteoarthritis sample-set is unknown, but two of them had undergone total knee replacements. The single carrier in the Twins UK cohort from the United Kingdom did not have hand osteoarthritis at the age of 51 when she was examined (X-rays). The average age of cases with the severe hand OA in the Twins UK cohort is 63 years old, with age range between 55 and 72 years. This carrier is therefore 4 years younger than the youngest case with severe hand osteoarthritis in Twins UK. She may have developed hand osteoarthritis later on. A single carrier who had undergone knee replacement due to OA was found in the PAPRIKA study, however, no information was available on hand osteoarthritis status. Nine carriers were identified in the Rotterdam III study of relatively young people (mean age 56 years). Information on osteoarthritis was available for six carriers; one had knee osteoarthritis and one had hand osteoarthritis.

Bioinformatics analysis: We cross referenced chr1:63807756 against potential biological functional features in the UCSC genome browser[18]. The variant doesn't overlap any reported open chromatin or transcription factor binding sites in ENCODE project[19]. The variant overlaps region of increased density of mapped reads in Long RNA-seq from ENCODE/Cold Spring Harbor Lab[19] in cell line from undifferentiated Chondrocytes from knee joint (HCH). The variant overlaps MeDIP-seq methylated CpG from brain sample[20]. Furthermore the variant is in high scoring region of ESPERR Regulatory Potential (7 Species) track marking potential unknown exon or regulatory region[17,21].

Full acknowledgements

We thank all the study subjects for their valuable participation, the staff from all studies and the participating physicians. The study was supported by deCODE Genetics/Amgen. TreatOA is funded by the European Commission framework 7 programme (grant 200800). The GARP study was supported the Leiden University Medical Centre and the Dutch Arthritis Association. Pfizer Inc., Groton, CT, USA supported the inclusion of the GARP study. The genotypic work was supported by the Netherlands Organization of Scientific Research (MW 904-61-095, 911-03-016, 917 66344 and 911-03-012), Leiden University Medical Centre and the Centre of Medical System Biology and Netherlands Consortium for Healthy Aging both in the framework of the Netherlands Genomics Initiative (NGI). The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under Grant Agreement No. 259679. The Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) (project nr. 050-060-810) and the Erasmus Medical Center and Erasmus University, Rotterdam. The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust

and King's College London. Tim Spector is holder of an ERC Advanced Principal Investigator award*. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR for TwinsUK. We acknowledge the help and support of the twin volunteers. The Chingford Study is funded by a grant from ARUK and genotyping was provided by a grant from Pfizer Inc. With help from Alan Hakim and Maxine Daniels and the Chingford volunteers.

References:

1. Jonsson, H. *et al.* The inheritance of hand osteoarthritis in Iceland. *Arthritis Rheum* **48**, 391-5 (2003).
2. Eliasson, G.J., Verbruggen, G., Stefansson, S.E., Ingvarsson, T. & Jonsson, H. Hand radiology characteristics of patients carrying the T(303)M mutation in the gene for matrilin-3. *Scand J Rheumatol* **35**, 138-42 (2006).
3. Jonsson, H. *et al.* High hand joint mobility is associated with radiological CMC1 osteoarthritis: the AGES-Reykjavik study. *Osteoarthritis Cartilage* **17**, 592-5 (2009).
4. Hofman, A. *et al.* The Rotterdam Study: 2012 objectives and design update. *European Journal of Epidemiology* **26**, 657-686 (2011).
5. Spector, T.D. & Williams, F.M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899-906 (2006).
6. Hart, D.J. & Spector, T.D. Cigarette smoking and risk of osteoarthritis in women in the general population: the Chingford study. *Ann Rheum Dis* **52**, 93-6 (1993).
7. Riyazi, N. *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* **64**, 438-43 (2005).
8. Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84 (2006).
9. Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84 (2005).
10. Keurentjes, J.C. *et al.* Socio-Economic Position Has No Effect on Improvement in Health-Related Quality of Life and Patient Satisfaction in Total Hip and Knee Replacement: A Cohort Study. *PLoS ONE* **8**, e56785 (2013).
11. Richards, J.B. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* **371**, 1505-12 (2008).
12. Zeger, S.L. & Liang, K.Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-30 (1986).
13. Diggle, P.J., Liang, K.Y. & Zeger, S.L. Analysis of longitudinal data. *Oxford University Press* (1994).
14. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
15. Teo, Y.Y. *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**, 2741-6 (2007).
16. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **9**, e1001046 (2011).
17. Taylor, J. *et al.* ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. *Genome Res* **16**, 1596-604 (2006).
18. Meyer, L.R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Research* **41**, D64-D69 (2013).

19. Rosenbloom, K.R. *et al.* ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research* **41**, D56-D63 (2013).
20. Maunakea, A.K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-257 (2010).
21. King, D.C. *et al.* Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. *Genome Research* **15**, 1051-1060 (2005).
22. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).

Supplementary Table 1. Correlation between rs12907038 and markers belonging to the two frequency groups in the 15q22 region

					Allele		rs12907038 G		rs4238326		rs3204689	
P	OR	SNP	Freq.	Position	EA	OA	r^2	P_{adj}	r^2	P_{adj}	r^2	P_{adj}
4.0E-10	1.51	rs12907038	50.73	56031324	G	-	1	0.22	0.65	0.013	0.92	0.026
2.2E-09	1.48	rs10851633	41.31	56110357	T	C	0.66	0.070	0.99	0.22	0.61	0.022
2.2E-09	1.48	rs35246600	41.16	56123718	A	T	0.65	0.068	1	0.22	0.60	0.021
2.3E-09	1.48	rs4238326	41.15	56123292	C	T	0.65	0.070	1	0.23	0.60	0.022
2.3E-09	1.48	rs11071365	41.16	56121418	A	G	0.65	0.070	1	0.24	0.60	0.022
2.4E-09	1.48	rs12907038	52.45	56031324	-	C	0.92	0.48	0.60	0.026	1	0.15
2.5E-09	1.48	rs4646640	52.38	56034855	G	C	0.92	0.47	0.60	0.026	1	0.15
2.5E-09	1.48	rs3204689	52.45	56034094	C	G	0.92	0.49	0.60	0.027	1	0.16
3.0E-09	1.48	rs7495968	52.43	56031395	G	C	0.92	0.54	0.60	0.030	1	0.20
3.2E-09	1.47	rs11071366	41.36	56121536	T	A	0.65	0.080	0.99	0.36	0.6	0.03
3.6E-09	1.48	rs4646638	52.42	56036075	G	A	0.92	0.60	0.60	0.034	1	0.25
3.6E-09	1.48	rs12910113	52.41	56035665	C	A	0.92	0.60	0.61	0.035	1	0.25
4.1E-09	1.47	rs9325	52.23	56033278	A	T	0.91	0.61	0.61	0.039	0.99	0.28
4.6E-09	1.47	rs4646568	40.82	56131582	T	C	0.65	0.092	0.97	0.48	0.60	0.03
4.8E-09	1.47	rs8025493	48.98	56009019	G	A	0.84	0.34	0.65	0.055	0.77	0.10
4.9E-09	1.46	rs4646576	41.09	56126290	T	A	0.65	0.11	1	0.59	0.60	0.03
9.1E-09	1.46	rs11855259	51.60	56036669	G	C	0.87	0.70	0.63	0.075	0.95	0.46
9.9E-09	1.46	rs35511675	51.37	56054270	A	T	0.86	0.70	0.63	0.079	0.94	0.46
1.0E-08	1.46	rs7178497	51.24	56027625	T	C	0.85	0.67	0.63	0.076	0.94	0.45
1.1E-08	1.46	rs4646611	51.29	56052375	C	T	0.86	0.72	0.64	0.083	0.94	0.49
1.1E-08	1.46	rs4646612	51.29	56052369	T	C	0.86	0.73	0.64	0.083	0.94	0.49
1.1E-08	1.46	rs4646593	51.46	56089759	G	A	0.85	0.69	0.64	0.087	0.93	0.47
1.2E-08	1.45	rs11630835	51.31	56056754	T	C	0.86	0.76	0.64	0.089	0.95	0.53
1.2E-08	1.45	rs11071356	51.16	56027810	G	T	0.85	0.72	0.63	0.084	0.94	0.50
1.2E-08	1.45	rs12911071	51.48	56102140	A	G	0.85	0.74	0.65	0.094	0.93	0.51
1.2E-08	1.45	rs4646563	51.44	56140627	C	T	0.85	0.70	0.64	0.095	0.93	0.51
1.3E-08	1.45	rs12901462	51.52	56103779	G	A	0.85	0.75	0.65	0.10	0.93	0.52
1.3E-08	1.45	rs11852540	51.49	56108756	G	A	0.85	0.76	0.65	0.10	0.93	0.53
1.3E-08	1.45	rs12903551	51.31	56055664	T	C	0.86	0.78	0.64	0.093	0.94	0.55
1.3E-08	1.45	rs4369598	51.47	56105895	T	G	0.85	0.76	0.64	0.10	0.93	0.54
1.3E-08	1.45	rs12915901	51.26	56066724	A	G	0.86	0.77	0.63	0.092	0.94	0.54
1.3E-08	1.45	rs4646586	51.43	56092636	A	C	0.85	0.76	0.64	0.10	0.93	0.54
1.4E-08	1.45	chr15:56103658	51.40	56103658	G	GGAAGA	0.85	0.76	0.65	0.10	0.93	0.54
1.4E-08	1.45	rs3784262	51.28	56040398	C	T	0.86	0.82	0.64	0.10	0.95	0.59

1.4E-08	1.45	rs12903474	51.35	56064039	A	C	0.86	0.81	0.64	0.10	0.94	0.59
1.4E-08	1.45	rs12148093	51.49	56116232	T	A	0.85	0.79	0.65	0.10	0.93	0.57
1.5E-08	1.45	rs4646627	51.29	56042597	G	A	0.86	0.84	0.64	0.10	0.95	0.61
1.5E-08	1.45	rs4646629	51.30	56042507	G	A	0.86	0.84	0.64	0.10	0.95	0.62
1.5E-08	1.45	rs4646622	51.33	56044491	G	A	0.86	0.84	0.63	0.10	0.94	0.61
1.6E-08	1.45	chr15:56054709	51.29	56054709	A	ACATA	0.86	0.87	0.63	0.11	0.94	0.65
1.7E-08	1.45	rs1372368	51.42	56084655	T	C	0.85	0.83	0.64	0.11	0.93	0.62
1.7E-08	1.45	rs7164408	51.31	56046628	G	A	0.86	0.84	0.64	0.11	0.94	0.62
1.8E-08	1.45	rs10851630	51.23	56029440	T	C	0.86	0.87	0.63	0.11	0.94	0.66
1.8E-08	1.45	rs12148907	51.43	56083742	A	T	0.85	0.85	0.64	0.12	0.93	0.64
1.9E-08	1.44	rs74655564	44.62	56112804	C	G	0.65	0.20	0.86	0.49	0.71	0.13
1.9E-08	1.45	rs17820823	52.83	56026650	G	A	0.89	0.97	0.58	0.084	0.97	0.83
2.1E-08	1.44	rs7170896	51.30	56039685	A	T	0.86	0.95	0.64	0.12	0.94	0.75
2.5E-08	1.44	rs12910752	51.48	56101971	A	G	0.85	0.97	0.64	0.14	0.93	0.78
2.6E-08	1.44	rs7165247	51.27	56039078	C	T	0.86	0.97	0.64	0.14	0.94	0.85
2.8E-08	1.44	rs10851631	51.22	56029485	T	C	0.86	0.97	0.63	0.14	0.94	0.86
3.0E-08	1.44	rs4646619	51.22	56045634	G	A	0.86	0.93	0.64	0.16	0.94	0.90
3.1E-08	1.44	rs34871384	51.32	56067166	A	C	0.85	0.96	0.63	0.15	0.94	0.88
3.6E-08	1.44	rs4646636	52.27	56038326	G	A	0.85	0.96	0.61	0.14	0.94	0.88
3.7E-08	1.44	rs1579744	53.83	56016815	G	A	0.87	0.86	0.58	0.12	0.95	0.98
4.2E-08	1.44	rs11852835	53.94	56015891	G	A	0.86	0.84	0.57	0.12	0.94	0.99

Association with severe hand osteoarthritis for variants in the 15q22 region that correlate with the tri-allelic marker rs12907038. The two frequency groups are presented by red (52%) and blue (41%) colors. The table includes for each variant with $P < 5\times10^{-8}$, the *P* value and OR from the association tested in the 623 individuals with severe hand osteoarthritis and 69,153 controls, the SNP, the position in NCBI Build 36 coordinates, the marker alleles where EA stands for effect allele and OA refers to the other allele, and, for each of the three key markers in the paper: rs12907038, rs4238326 and rs3204689, the r^2 with that marker and an adjusted *P* value corresponding from a test of the association conditional on the observed association with that marker. All *P* values have been adjusted for relatedness using the method of genomic control.

Supplementary Table 2. Association of rs4238326 and rs3204689 with severe hand osteoarthritis in individual replication sample sets

Phenotype	Marker	Sample set	N cases / controls	OR	*P* value	*P* het
Severe fingers and severe thumbs	rs4238326-C	Rotterdam I	94/2,652	1.30	0.082	
		Rotterdam II	22/1,080	1.30	0.42	
		GARP	29/2,374	1.49	0.13	
		Twins UK	21/1,536	1.04	0.90	
		Chingford	48/530	1.55	0.14	
		Combined		*1.33 (1.08-1.64)*	*0.0075*	*0.90*
	rs3204689-C	Rotterdam I	94/2,652	1.44	0.015	
		Rotterdam II	22/1,080	1.40	0.28	
		GARP	29/2,374	1.42	0.19	
		Twins UK	21/1,536	1.26	0.46	
		Chingford	48/530	1.48	0.22	
		Combined		*1.41 (1.15-1.74)*	*0.0011*	*1.00*
Severe fingers	rs4238326-C	Rotterdam I	392/2,652	1.11	0.20	
		Rotterdam II	76/1,080	1.18	0.34	
		GARP	99/2,374	1.63	0.00086	
		Twins UK	47/1,510	1.05	0.82	
		Chingford	99/479	1.55	0.090	
		Combined		*1.21 (1.07-1.36)*	*0.0017*	*0.15*
	rs3204689-C	Rotterdam I	392/2,652	1.18	0.029	
		Rotterdam II	76/1,080	1.21	0.27	
		GARP	99/2,374	1.63	0.00090	
		Twins UK	47/1,510	1.10	0.66	
		Chingford	99/479	1.34	0.28	
		Combined		*1.25 (1.11-1.40)*	*0.00017*	*0.37*
Severe thumbs	rs4238326-C	Rotterdam I	461/2,652	1.16	0.046	
		Rotterdam II	93/1,080	0.99	0.97	
		GARP	45/2,374	1.02	0.94	
		Twins UK	136/1,421	1.41	0.012	
		Chingford	192/386	1.50	0.057	
		Combined		*1.19 (1.07-1.33)*	*0.0021*	*0.34*
	rs3204689-C	Rotterdam I	461/2,652	1.16	0.043	
		Rotterdam II	93/1,080	1.14	0.40	
		GARP	45/2,374	1.01	0.95	
		TwinsUK	136/1,421	1.24	0.11	
		Chingford	192/386	0.97	0.87	
		Combined		*1.14 (1.03-1.27)*	*0.014*	*0.85*

The phenotypes of associations are shown, along with their respective *P* values and odds ratios (OR). The 95% confidence interval (CI) of the association in the combined sample sets is shown and the heterogeneity *P* value, *P* het. Numbers of cases and controls for each phenotype is shown.

Supplementary Table 3. Pair-wise conditional analysis of rs4238326 and rs320468 in replication samples

Hand OA phenotype	marker	*P* value	OR_{adj}	P_{adj} value	*P* het
Severe fingers AND severe thumbs	rs4238326-C	0.014	1.02	0.92	0.77
	rs3204689-C	0.0012	1.41	0.037	0.95
Severe Fingers	rs4238326-C	0.0014	1.02	0.88	0.51
	rs3204689-C	0.00011	1.24	0.025	0.95
Severe Thumbs	rs4238326-C	0.051	1.16	0.11	0.05
	rs3204689-C	0.02	1.02	0.79	0.19

Association with severe hand osteoarthritis phenotypes in replication samples done adjusting the association with rs4238326-C or rs320468-C for the other marker, respectively. The table includes the unadjusted *P*-value, the adjusted *P*-value, P_{adj}, and the heterogeneity *P*-value, *P* het, and the adjusted OR, OR_{adj}.

Supplementary Table 4. Studies included in the Treat OA meta-analyses, their country of origin (all of European descent) and sample sizes

Study	**Hip osteoarthritis**		**Knee osteoarthritis**	
	N cases	**N controls**	**N cases**	**N controls**
arcOGEN stage 1, United Kingdom	1,728	4,896	1,643	4,896
TwinsUK, United Kingdom	68	228	113	228
deCODE, Iceland	1,423	31,385	783	31,385
EGCUT, Estonia	64	2531	123	2,531
Framingham, United States America	-	-	417	1,667
GARP, The Netherlands	106	1,671	148	1,671
Rotterdam I, The Netherlands	760	3,233	1,476	3,239
Rotterdam II, The Netherlands	159	1,472	369	1,467
Rotterdam III, The Netherlands	41	1,487	152	1,487
Total	**4,349**	**46,903**	**5,224**	**48,571**

Supplementary Table 5. Associated rare variants at the 1p31 locus

P value	OR	Marker	Freq.	Info	Effect allele	Other allele	# genotyped (prediced carriers)	Freq.[a]	*P* value[a]	Ora	Info[a]
2.8×10^{-9}	42.7	chr1:63488386	0.023	0.99	G	C	1,430 (43)	0.005	0.068	20.3	1
1.5×10^{-9}	47.7	chr1:63724786	0.022	0.97	T	TAAGG	450 (32)	0.02	9.6×10^{-10}	50.7	0.99
1.5×10^{-9}	47.7	chr1:63807756	0.022	0.97	A	G	5,554 (44)	0.02	9.8×10^{-10}	50.6	0.99
2.7×10^{-8}	23.5	chr1:63827357	0.059	0.79	T	C	334 (39)	0.02	3.8×10^{-10}	63.7	0.95

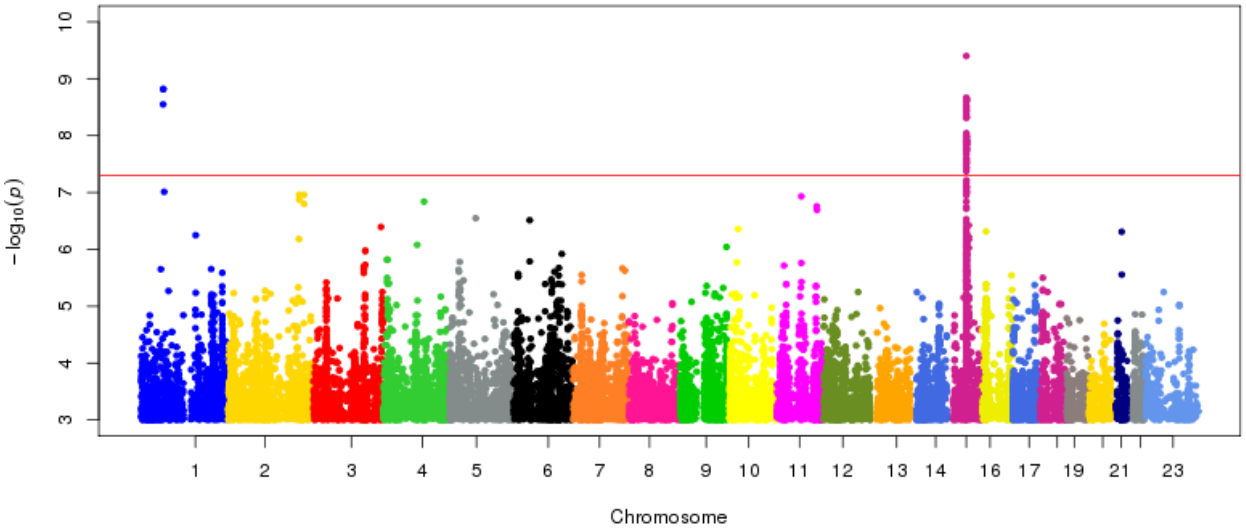
Results for the rare variants at the 1p31 locus. The marker name is given according to the NCBI_build36/hg18 coordinates, and the table includes the frequency (Freq.) in the Icelandic samples, and the *P* values and odds ratios (OR) from logistic regression for a case control analysis in the 623 individuals with severe hand osteoarthritis and 69,153 controls used in the genome wide association analyses. The estimated imputation information is shown. The number of individuals directly genotyped with predicted and likely carriers and in parenthesis. [a]The frequency, *P* values, OR and imputation information based on the re-imputed genotypes. All *P* values have been adjusted for relatedness using the method of genomic control

Supplementary Table 6. RAAK study sample characteristics

Sample number	Donor	Gender	Joint	BMI	Remarks
1	RAAK-28	Male	Hip	27.2	
2	RAAK-38	Male	Hip	27	
3	RAAK-141	Female	Knee	35.8	
4	RAAK-155	NA	Knee	30	
5	RAAK-158	Female	Hip	25.6	
6	RAAK-161	Female	Hip	30.8	
7	RAAK-163	Male	Hip	25.3	
8	RAAK-180	Female	Knee	35.6	
9	RAAK-183	Female	Knee	NA	
10	RAAK-220	Female	Knee	NA	
11	RAAK-52	Female	Hip	28.8	
12	RAAK-66	Female	Hip	25.1	
13	RAAK-119	Male	Knee	NA	
14	RAAK-145	Male	Hip	NA	Non OA
15	RAAK-225	Female	Hip	26.3	
16	RAAK-41	Male	Hip	28.7	
17	RAAK-51	Male	Knee	29.3	
18	RAAK-56	Male	Hip	20.8	
19	RAAK-134	Female	Hip	26.6	
20	RAAK-166	Female	Hip	24.7	
21	RAAK-172	Female	Hip	NA	
22	RAAK-174	Female	Hip	42	

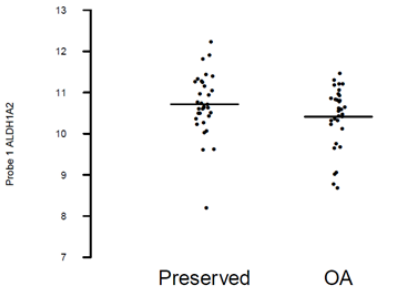
Supplementary Figure 1. Manhattan plot of discovery genome-wide association study.

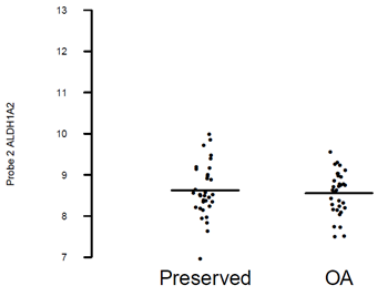
The *P* values ($-\log_{10}$) are plotted against their respective positions on each chromosome. $P = 5\times10^{-8}$ is indicated by the horizontal pink line.



Supplementary Figure 2. Overall expression level of *ALDH1A2* in cartilage

Expression of ALDH1A2 was examined by exploring a micro array mRNA expression dataset generated on Illumina HT-12 V3 chips in cartilage samples of 33 patients (13 males and 20 females of European descent aged from 54 to 80) that underwent joint replacement due to end stage OA disease. The level of expression, as determined by the mean normalized probe level value, for *ALDH1A2* probe 1 (mean level 10.44, chromosome 9, position 58,245,807-58,245,857; Hg19) and probe 2 (mean level 8.51, chromosome 9, position 58246307-58246356; Hg19) which was above the observed average expression of genes in the articular cartilage (mean normalized probe level value of measured genes in cartilage was 7.4 with range 6.6-14.9). When we tested for differential expression of *ALDH1A2* among the pairs of preserved and OA affected cartilage, we observed a significant lower expression of ALDH1A2 OA affected cartilage only for probe 1 (FC=0.82; nominal P-value = 0.0253).





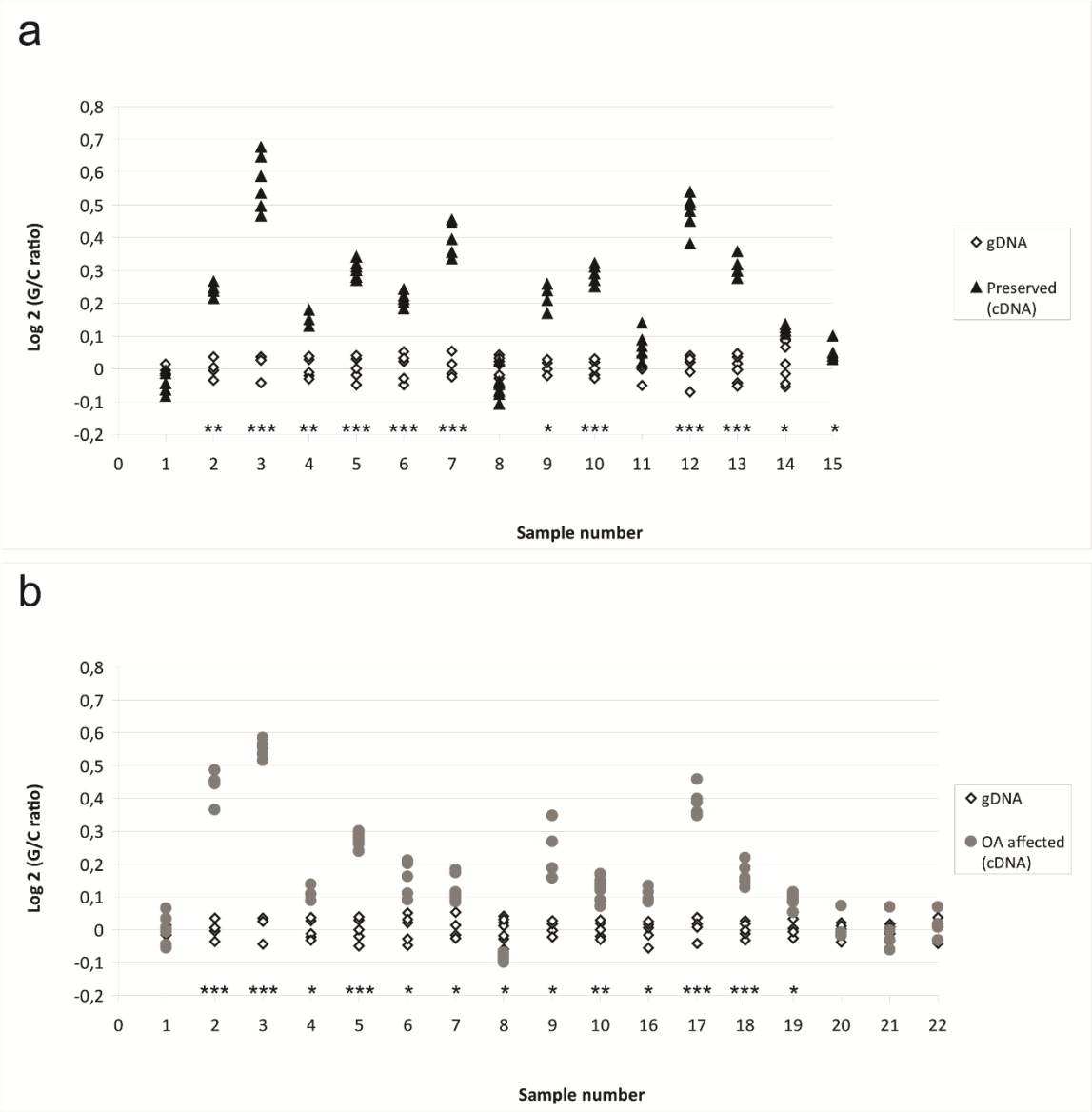
Overall expression probe 1 = 10.44
P – OA FC= 0.82 $P_{nominal}$ = 0.0253

Overall expression probe 2 = 8.51
P – OA FC= 0.96 $P_{nominal}$ = 0.5595

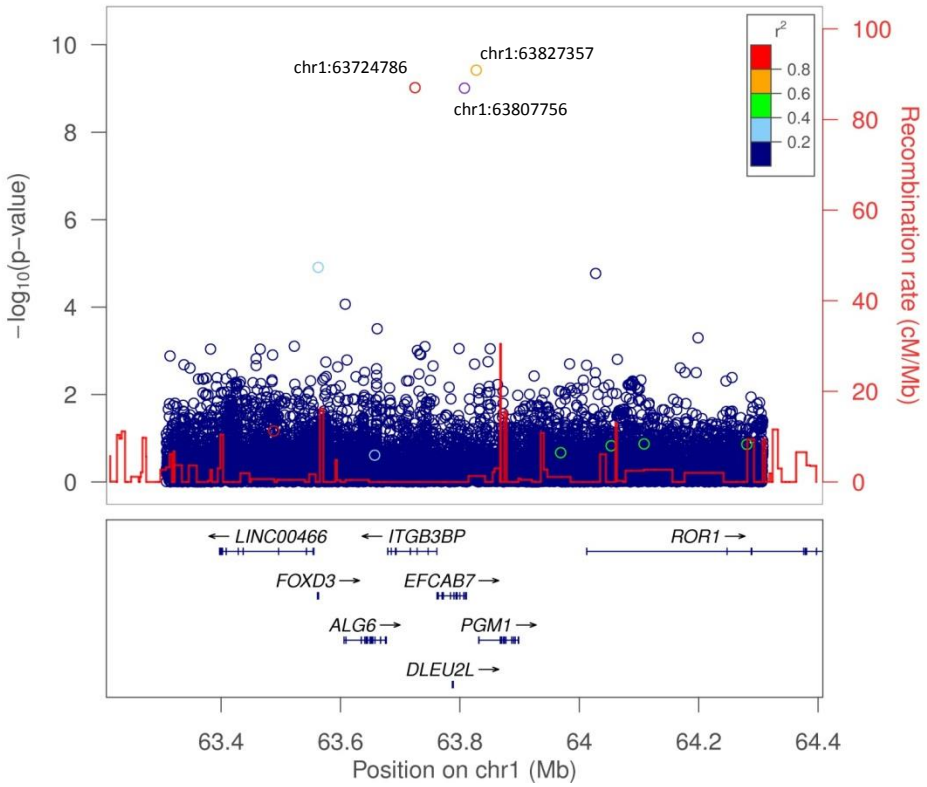
Supplementary Figure 3. Differential allelic expression of *ALDH1A2* transcripts according to genotypes of rs3204689

Allelic imbalance depicted by allele-ratios (G/C) plotted as log2-values for 22 individuals heterozygous for the rs3204689 SNP. (**a**) Preserved articular cartilage samples (N=15) (**b**) Osteoarthritis affected articular cartilage samples (N=17). *$P < 0.0005$, **$P < 0.0001$, ***$P < 0.00001$).

Overall there is 13.6% ($P = 1.09\times10^{-30}$) abundance of the G allele as compared to the risk allele C. When samples were stratified according to affection status, we observed 16.5% ($P = 4.2\times10^{-18}$) and 11.2% ($P = 1.3\times10^{-14}$) abundance of the G allele as compared to the risk-allele C in, preserved and osteoarthritis affected cartilage, respectively. This difference in extend of the allelic imbalance between preserved and OA affected cartilage was significant ($P = 0.012$).



Supplementary Figure 4. Genes in the 1p31 locus region



Location of genes in the region on 1p31 that is shared by all the affected carriers and the location of the three associated markers and the linkage disequilibrium blocks (LD) are shown. The genes are: *LINC00466* (long intergenic non-protein coding RNA 466), *FOXD3* (forkhead box D3), *ALG6* (alpha-1,3-glucosyltransferase) *ITGB3BP* (integrin beta 3 binding protein), *EFCAB7* (EF-hand calcium binding domain 7), *PGM1* (phosphoglucomutase 1), and *ROR1* (receptor tyrosine kinase-like orphan receptor 1), and the pseudogene *DLEU2L* (deleted in lymphocytic leukemia 2-like). All positions are in NCBI Build 36 coordinates. The plot was created using an standalone version of the LocusZoom software[22] (http://csg.sph.umich.edu/locuszoom/).