SUPPLEMENTARY METHODS

Description of Datasets

GIANT study
The consortium consists of KORA (Cooperative Research in the Region of Augsburg, Germany), DGI (Diabetes Genetics Initiative), FUSION (The Finland-United States Investigation of Non-Insulin-Dependent Diabetes Mellitus Genetics), PLCO (Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial) and NHS (Nurses Health Study), WTCCC T2D, HT, CHD, UKBS (Wellcome Trust Case-Control Consortium for Type 2 Diabetes, Hypertension, Coronary Heart Disease, UK Blood Service) and EPIC (European Prospective Investigation of Cancer and Nutrition), Twins UK, deCODE (Iceland), CHARGE (Framingham, Rotterdam), Amish, Family Heart Study, Remaining ENGAGE, EUROSPAN, ADVANCE, PROCARDIS, CAPS, CAHRES, GERMIFS, SEARCH, SHIP, CNRS/ICL, Essen Obesity Study, CoLaus (Cohorte LAUSanneoise, Switzerland) and SardiNIA (Italy).

arcOGEN stage 2 GWAS
Cases. The stage 2 cases were collected prospectively as part of the arcOGEN study at 9 locations across the UK (Edinburgh, London, Newcastle-Upon-Tyne, Nottingham, Oxford, Sheffield, Southampton, Wansbeck, and Worcester). The ascertainment criteria for the majority of cases was primary OA that was severe enough for the individual to require joint replacement of the hip or of the knee; with an additional small number who were collected as part of a randomized, placebo controlled trial of vitamin D replacement (VIDEO Study) in whom the ascertainment criteria was radiographic disease of the knee, with a KL grade of ≥2. All cases were unrelated and of European origin. The exclusion criteria included the need for joint replacement due to fracture, secondary OA of any cause, and developmental, vascular, or infective causes of joint disease. The prospective collections were approved by the National Research Ethics Service in the United Kingdom, and all subjects provided written, informed consent prior to inclusion.

Controls. The controls that were employed in stage 2 were population-based, unrelated UK controls which came from 5 distinct sources: the 1958 Birth Cohort and the UK Blood Donor Service from the Wellcome Trust Case Control Consortium 2 (WTCCC2) study genotyped on the Illumina 1.2M Duo; the 1958 Birth Cohort from the Type 1 Diabetes Genetics Consortium (T1DGC) study genotyped on the Illumina HumanHap550K; the Avon Longitudinal Study of Parents and Children (ALSPAC) genotyped on the Illumina Infinium 670k; and the People of the British Isles (PoBI) study genotyped on the Illumina 1.2M Duo (Illumina, San Diego, USA).

DNA sample preparation and quality control (QC). Genomic DNA from extant cohorts and blood samples from new cases were shipped by the collection centres to the Centre for Integrated Genomic Medical Research (CIGMR), University of Manchester, for processing. DNA extraction was performed as previously described (Panoutsopoulou et al., 2010). 2D bar coded matrix plates and patient data manifests for each set of 96 DNAs were shipped to The Wellcome Trust Sanger Institute, Hinxton. The quality of the DNA and subject identity were validated using a Sequenom iPLEX assay designed to genotype 4 gender-specific SNPs and 26 SNPs present on the Illumina Human 610-Quad array (Illumina, San Diego, USA). Sample preparation (quantification of DNA concentration and QC criteria) was performed as previously described (Panoutsopoulou et al., 2010).

Genotyping and allele calling. DNA samples passing QC were rearrayed into 96-well plates for genotyping using Illumina Human 610-Quad BeadChips (Illumina, San Diego, USA) as previously described (Panoutsopoulou et al., 2010). Normalised bead intensity data obtained for each sample were converted into SNP genotypes using a customised genotype calling algorithm (Teo et al., 2007).
Sample quality control (QC). Sample QC was performed separately for each cohort. We used the same method as previously described (Panoutsopoulou et al., 2010). Briefly, individuals were excluded if they had a call rate <97% and if they had a discrepancy between the supplied gender and the gender that was inferred using X-chromosome heterozygosity rates in PLINK (Purcell et al., 2007). We also excluded samples with excess genome-wide homozygosity and heterozygosity (based on autosomal SNPs only) by examining histograms and removing visual outliers. The thresholds we chose for each particular data set were: arcOGEN cases <30.5% and >34%; ALSPAC <30.5% and >33%; PoBI <26.8% and >28.3%; T1DGC <31% and >32.75%; WTCCC2-58BC <32.5% and >33.74%; WTCCC2-UKBS <26.6% and >28.1%. We removed duplicated/related samples by calculating the genome-wide pair-wise identity by descent (IBD) for each sample; we excluded any sample with a π^ > 0.2 (Purcell et al., 2007). We examined ethnicity by performing multidimensional scaling (MDS) (Purcell et al., 2007) and principle component analysis (PCA) using Goldsurfer 2 (Petersson et al., 2008); we combined autosomal information from each of our populations with 210 founders from the HapMap phase II release 23a data, which included the European (CEU), African (YRI) and Chinese/Japanese (ASN) populations and we excluded samples of non-European descent. For the ALSPAC, PoBI, WTCCC2-58BC and WTCCC2-UKBS cohorts we excluded individuals identified from additional QC analyses performed by the suppliers of the datasets.

SNP QC. SNP QC was conducted on autosomal SNPs using thresholds that have been previously described (Panoutsopoulou et al, 2010). In short, SNPs were excluded if they had a minor allele frequency (MAF) ≥5% and a call rate <95%, or a MAF <5% and call rate <99%. All monomorphic SNPs were excluded and also those SNPs with an exact Hardy Weinberg Equilibrium (HWE) p value<0.0001. The number of SNPs that passed standard QC in each dataset was: for arcOGEN cases 558,546; for ALSPAC 542,060; for PoBI 965,407; for T1DGC 535,266; for WTCCC2-58BC 1,005,923; for WTCCC2-NBS 996,623; for TwinsUK 542,037. We extracted 492,811 overlapping SNPs between all population-based control datasets and performed pairwise association tests for all possible pairs. Upon inspection of qq plots 866 SNPs with multiplicative p<10^-4 were removed from all population-based control datasets prior to pooling them into one common control group. 490,130 SNPs from the common control group overlapped with the clean arcOGEN cases SNPs. Finally we removed 70 SNPs following inspection of intensity plots of the top signals (p<10^-5) that were identified in our case/control association analyses and 4,569 SNPs with MAF <1% (estimated separately for each cohort). The resulting stage 2 GWAS OA cases vs. population-based controls dataset contained 4,324 cases, 6,518 controls (samples that had not been included in arcOGEN stage 1 GWAS) and 485,491 autosomal SNPs.

TwinsUK controls
We performed BMI adjusted analyses of stage 1 arcOGEN cases against TwinsUK, OA-free controls for the signals of interest (for females only). Genotyping and QC of the TwinsUK dataset has been previously described [Panoutsopoulou et al., 2010]. We also excluded 47 samples that failed QC when the normalised intensity data for these samples were pooled with data from a further 3 genotype platforms HumanHap300, 1M-Duo and 1.2MDuo (Illumina, San Diego, USA) (Illumina, San Diego, USA) using the following criteria (i) sample call rate <98%; (ii) heterozygosity across all SNPs ≥2 s.d. from the sample mean; (iii) evidence of non-European ancestry, assessed by PCA comparison with HapMap phase III populations; (iv) observed pairwise IBD probabilities suggestive of sample identity errors.

De novo replication samples and genotyping.
The case set comprised 5,165 subjects that had not been included in the arcOGEN stage 1 GWAS. The control set comprised 6,155 individuals from the WTCCC2-58BC and WTCCC2-UKBS cohorts that were not part of the arcOGEN stage 1 GWAS. Genotyping was performed on the Sequenom MassArray iPLEX Gold assay at the Wellcome Trust Sanger Institute. The genotypes were assigned (and manually confirmed) using the MassArray TyperAnalyzer software version 4.0 (Sequenom). All standard QC checks were performed.
Analysis Methods

Sign test
Starting with the LD-thinned list of SNPs ordered on p value for association with OA, we determined whether each SNP was concordant or discordant for risk allele in the two phenotypes under study (OA and height or OA and BMI) and assigned a score of +1 or -1 respectively. Starting from the top of the list, these scores were added and deviation from the null (50:50 distribution of concordant and discordant SNPs) was tested every 20th SNP using the binomial test.

SUPPLEMENTARY RESULTS

Sign Test
The results for the cumulative scores for concordant/discordant alleles for OA-height and OA-BMI are shown in Supplementary Figure 1a and 1b, respectively. For the overlap of OA and height (Supplementary Figure 1a) the graph shows that there is a trend for discordance of OA and height risk alleles (-1). This is demonstrated by the cumulative score decreasing as the proportion of discordant alleles (-1) increases (the null hypothesis of no relationship between the diseases would show a flat line where there is a 50:50 mix of concordant (+1) and discordant (-1) alleles). The binomial test was used to test the deviation of the scores from the null hypothesis. Across all 24,000 independent SNPs, an excess of (12,735) discordant alleles were scored compared to (11,265) concordant alleles (binomial p=1.29x10^{-21}). For the OA-BMI comparison the picture is less clear (Supplementary Figure 1b). There is no obvious trend for OA risk alleles to overlap with either risk or non-risk alleles for BMI. Also the negative log of the p value never exceeds 1.4 (1 x 10^{-1}).

SUPPLEMENTARY REFERENCES


