Osteoarthritis (OA) is a prevalent, heritable degenerative joint disease with a substantial public health impact. We used a 1000-Genomes-Project-based imputation in a genome-wide association scan for osteoarthritis (3177 OA cases and 4894 controls) to detect a previously unidentified risk locus. We discovered a small disease-associated set of variants on chromosome 13. Through large-scale replication, we establish a robust association with SNPs in MCF2L (rs11842874, combined odds ratio [95% confidence interval] 1.17 [1.11–1.23], p = 2.1 × 10⁻¹⁰) across a total of 19,041 OA cases and 24,504 controls of European descent. This risk locus represents the third established signal for OA overall. MCF2L regulates a nerve growth factor (NGF), and treatment with a humanized monoclonal antibody against NGF is associated with reduction in pain and improvement in function for knee OA patients.
Birth Cohort (58BC) and the UK National Blood Donor Service (UKBS) and were obtained from an early release of the Wellcome Trust Case Control Consortium 2 (WTCCC2) data. The genotyping and quality control (QC) of these individuals and their genotype data were described previously in the initial arcOGEN GWAS.8 Our primary 1KGP imputation was based on the April 2009 release of haplotypes for 57 individuals. After removing rare variants (with minor allele frequency [MAF] < 0.01) and SNPs with low imputation quality (r² < 0.3), 7,258,070 variants were tested for association with OA. Figure 1. Overview of Study Design

- Impute on the basis of 1000 Genomes Project pilot 1 data
- Test for association in 3,177 cases and 4,894 UK controls across ~7.2M variants with MAF>0.01
- Closely examine signals with p<10⁻⁵ and remove low quality variants
- Reimpute and retest for association
- Validate 6 promising imputed signals in the original GWAS samples
- Replicate 6 promising signals in 5,165 cases and 6,155 UK controls
- Replicate chr13 signal in 2 further sets of 4,095 cases and 3,062 UK controls
- Replicate chr13 signal in 4 further European populations of 6,604 cases and 10,393 controls

Closely examine signals with p<10⁻⁵ and remove low quality variants

As part of our follow-up, we first genotyped an independent set of 5165 arcOGEN-collected cases and 6155 population-based controls from the 58BC and UKBS cohorts. Seven out of the eight SNPs were successfully typed with a Sequenom MassArray iPLEX Gold assay (Table S1) and one SNP, rs11842874 on 13q34, replicated with p = 2.60 × 10⁻³ (allelic odds ratio [OR] 1.17 [1.06–1.30]) and with consistent effect direction as the original scan (Table 1). We subsequently took this signal forward to de novo genotyping in two further sample sets from the UK: the Genetics of Osteoarthritis and Lifestyle (GOAL) study16,17 (1686 total joint replacement cases, 743 non-OA controls) and an additional independent set of 2409 newly recruited arcOGEN cases and 2319 population-based controls from the 58BC and UKBS cohorts. The combined UK meta-analysis (n = 12,437 cases, 14,111 controls) allelic OR was 1.22 [1.14–1.30], p = 2.24 × 10⁻⁸. We further investigated association with this variant in four non-UK OA sample sets: two from the Netherlands (Rotterdam Study I [RSI], 1950 cases and 3243 controls, and Rotterdam Study II [RSII], 485 cases and 1460 controls, both in silico),18–20 one sample set from Estonia (Estonian Genome Center, University of Tartu [EGCUT], 2617 cases and 2619 controls, de novo genotyping),21,22 and one from Iceland (deCODE, 1552 cases and 3071 controls, de novo genotyping). We used a meta-analysis framework to combine results across the follow-up studies only and across all data. We obtained the combined estimates of ORs for reference alleles by weighting the logORs of each study by the inverse of their variance via a fixed effects model. We investigated evidence of heterogeneity of ORs by using the Cochran’s Q and I² statistics. The meta-analysis was performed with the GWAMA software package.23 In all seven follow-up datasets combined, rs11842874 was associated with OA with p = 3.0 × 10⁻⁶ (allelic OR 1.13 [1.07–1.20]). Combined with the discovery sample set, the overall fixed effects meta-analysis (across 19,041 cases and 24,504 controls) established association at this variant with p = 2.07 × 10⁻⁸ (allelic OR 1.17 [1.11–1.23]; Figure 2, Table 1). The variant appears to be more strongly associated with knee OA (allelic OR 1.17 [1.10–1.25], p = 2.52 × 10⁻⁶, effective sample size of 28,987) than with hip OA (allelic OR 1.11 [1.03–1.19], p = 3.54 × 10⁻³, effective sample size of 27,452). Studies contributing data to this manuscript acquired informed consent from all participants and were approved by the appropriate ethics committee(s) for the respective institutions and countries.

rs11842874 is one of several highly correlated SNPs at 13q34 and constitutes the observed association signal, which spans 12.7 kb (Figure 3). The surrounding 1 Mb region is characterized by low levels of linkage disequilibrium and contains only nine SNPs correlated with rs11842874 at r² > 0.7. rs11842874 was selected for replication because it is the only variant present on some GWAS platforms. The OA risk-increasing allele is the major allele with population frequency of 0.927 (mean over the UK control data), and as a common variant with low OR, it contributes little to the sibling recurrence risk (estimated λ = 1.001). All SNPs that comprise the signal reside in intron 4 of MCF2L (MCF2 cell-line-derived transforming sequence-like [MIM 609499], encoding the guanine nucleotide exchange factor). MCF2L studies in rat models of OA have shown expression in articular chondrocytes.24,25 In
human cells MCF2L regulates neurotrophin-3-induced cell migration in Schwann cells.

Neurotrophin-3 is a member of the nerve growth factor (NGF) family. Treatment of knee OA patients with a humanized monoclonal antibody that inhibits NGF was found to be associated with joint pain reduction and an improvement in function.

The MCF2L OA locus was taken forward on the basis of evidence accrued through 1KGP-based imputation. Direct typing of ~600,000 SNPs through GWAS resulted in modest (p > 10^{-5}) evidence for the association of a single variant, rs11842874, with OA in this region and HapMap-based imputation left the picture unchanged (Figures 3A and 3B). Prioritization strategies for follow-up in our published GWAS down-weighted lone variants with no corroboration of association from neighboring SNPs. The denser 1KGP reference set empowered the association of multiple additional correlated (i.e., non-independent), imputed variants, several of which showed stronger evidence for association with OA, thus highlighting this region for validation and replication genotyping (Figure 3C, Table S1).

The associated variants are common, but their minor allele frequencies are toward the lower end of the frequency spectrum (at ~0.07). The identification of similar variants with modest effect sizes (OR 1.17) at genome-wide significance levels will require sample sizes in the order of ~23,000 cases and an equal number of controls. Through several rounds of cluster-plot inspection, removal of poor quality SNPs, and reimputation, we observed that the

<table>
<thead>
<tr>
<th>Study</th>
<th>OR (95% CI)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>arcOGEN GWAS</td>
<td>1.32 (1.16-1.50)</td>
<td>1.67 x 10^{-5}</td>
</tr>
<tr>
<td>arcOGEN replication set 1</td>
<td>1.17 (1.06-1.30)</td>
<td>2.60 x 10^{-3}</td>
</tr>
<tr>
<td>GOAL</td>
<td>1.24 (0.99-1.56)</td>
<td>7.20 x 10^{-2}</td>
</tr>
<tr>
<td>arcOGEN replication set 2</td>
<td>1.16 (0.98-1.37)</td>
<td>7.86 x 10^{-2}</td>
</tr>
<tr>
<td>RSI</td>
<td>0.98 (0.83-1.17)</td>
<td>9.61 x 10^{-1}</td>
</tr>
<tr>
<td>RSII</td>
<td>1.46 (1.07-2.00)</td>
<td>1.68 x 10^{-2}</td>
</tr>
<tr>
<td>EGCUT</td>
<td>1.16 (1.01-1.34)</td>
<td>4.01 x 10^{-2}</td>
</tr>
<tr>
<td>deCODE</td>
<td>1.03 (0.98-1.20)</td>
<td>7.31 x 10^{-1}</td>
</tr>
<tr>
<td>UK meta-analysis</td>
<td>1.17 (1.11-1.22)</td>
<td>2.07 x 10^{-8}</td>
</tr>
<tr>
<td>deCODE</td>
<td>1.09 (1.00-1.19)</td>
<td>4.70 x 10^{-2}</td>
</tr>
<tr>
<td>Combined meta-analysis</td>
<td>1.17 (1.11-1.23)</td>
<td>2.07 x 10^{-8}</td>
</tr>
</tbody>
</table>

Meta-analysis results are denoted in bold.
majority of signals were caused by genotyping and imputation artifacts. This fact highlights the need for imputed signals to be scrutinized postimputation before follow-up studies are deployed. As the field of complex trait association studies shifts its focus toward low-frequency and rare variants, thorough quality control of signals becomes highly relevant. In this study, we have restricted 1KGP-based imputation to variants with an MAF > 0.01. As reference panel sizes become larger, imputation of lower frequency variants will become more feasible, empowering the examination of rare variation in next generation association studies.

The genetic architecture of OA has not been elucidated yet. By identifying this susceptibility locus, the third one discovered for OA, our study now provides a foundation on which functional studies can be based. New additions to the genetic study toolset, including larger sample sets, well-characterized phenotypes, and resequenced reference-panel-based imputation approaches hold the promise of providing insights into the etiology of this common degenerative joint disease.

Supplemental Data
Supplemental Data include one table and the Acknowledgments and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments
The collection of GOAL samples was funded by Astra Zeneca UK. The author affiliated with AstraZeneca is an employee of AstraZeneca, a global research-based biopharmaceutical company focused on discovering, developing and marketing medicines for some of the world’s most serious illnesses, and owns stocks or stock options and has a pending patent application in the company. The authors that are affiliated with deCODE genetics are all employees of deCODE, a biotechnology company that provides genetic testing services, and some own stocks or stock options in the company. Additional acknowledgments are in the Supplemental Data.

Received: June 15, 2011
Revised: July 28, 2011
Accepted: August 1, 2011
Published online: August 25, 2011
References