nature genetics

Common variants in the *GDF5-UQCC* region are associated with variation in human height

Serena Sanna[1,2,19], Anne U Jackson[1,19], Ramaiah Nagaraja[3], Cristen J Willer[1], Wei-Min Chen[1,4], Lori L Bonnycastle[5], Haiqing Shen[6], Nicholas Timpson[7,8], Guillaume Lettre[9], Gianluca Usala[2], Peter S Chines[5], Heather M Stringham[1], Laura J Scott[1], Mariano Dei[2], Sandra Lai[2], Giuseppe Albai[2], Laura Crisponi[2], Silvia Naitza[2], Kimberly F Doheny[10], Elizabeth W Pugh[10], Yoav Ben-Shlomo[7], Shah Ebrahim[11], Debbie A Lawlor[7,8], Richard N Bergman[12], Richard M Watanabe[12,13], Manuela Uda[2], Jaakko Tuomilehto[14], Josef Coresh[15], Joel N Hirschhorn[9], Alan R Shuldiner[6,16], David Schlessinger[3], Francis S Collins[5], George Davey Smith[7,8], Eric Boerwinkle[17], Antonio Cao[2], Michael Boehnke[1], Gonçalo R Abecasis[1] & Karen L Mohlke[18]

Identifying genetic variants that influence human height will advance our understanding of skeletal growth and development. Several rare genetic variants have been convincingly and reproducibly associated with height in mendelian syndromes, and common variants in the transcription factor gene *HMGA2* are associated with variation in height in the general population[1]. Here we report genome-wide association analyses, using genotyped and imputed markers, of 6,669 individuals from Finland and Sardinia, and follow-up analyses in an additional 28,801 individuals. We show that common variants in the osteoarthritis-associated locus[2] *GDF5-UQCC* contribute to variation in height with an estimated additive effect of 0.44 cm (overall $P < 10^{-15}$). Our results indicate that there may be a link between the genetic basis of height and osteoarthritis, potentially mediated through alterations in bone growth and development.

Height is a complex trait that is influenced by genes and various environmental factors, including diet and the prenatal environment[3]. Heritability estimates suggest that ≥80% of variation in height may be genetically determined[3,4]. Rare mutations with large effects on height in mendelian syndromes have been identified in the genes *FBN1*, *FGFR3*, *GH1*, *EVC*, *EVC2* and *GPC3* (MIM 154700, 100800, 262400, 225500 and 312870 (corresponding syndromes, respectively)). Despite the high heritability, numerous candidate gene and linkage studies to identify loci influencing height in individuals of 'normal stature' have been inconclusive[5]. Overall, variation in human height is likely to be polygenic and heterogeneous. The first genome-wide association study (GWAS) of height[1] identified common variants in *HMGA2* that are associated with normal variation in height both in adults ($P = 4 \times 10^{-16}$) and in children ($P = 1 \times 10^{-6}$). These variants account for a small fraction (~0.3%) of the overall variation in height.

To identify additional genetic variants associated with height, we analyzed genome-wide SNP data on 2,371 Finns from the Finland–United States Investigation of NIDDM Genetics (FUSION) study[6] and on 4,298 Sardinians from the SardiNIA study[7] (**Table 1**), a longitudinal study on aging-related conditions in Sardinia. The two samples were originally genotyped with distinct sets of markers. We used genotype imputation methods[6] to facilitate comparison of the two studies and to evaluate an association between height and

[1]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. [2]Istituto di Neurogenetica e Neurofarmacologia (INN), Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari 09042, Italy. [3]Gerontology Research Center, National Institute on Aging, Baltimore, Maryland 21224, USA. [4]Division of Biostatistics and Epidemiology, Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA. [5]Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland 20892, USA. [6]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [7]Department of Social Medicine, University of Bristol, Canynge Hall, Bristol BS8 2PR, UK. [8]Medical Research Council (MRC) Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Canynge Hall, Bristol BS8 2PR, UK. [9]Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Seven Cambridge Center, Cambridge, Massachusetts 02142, USA. [10]Center for Inherited Disease Research (CIDR), Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21224, USA. [11]London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, UK. [12]Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [13]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. [14]Diabetes Unit, Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, 00300 Helsinki, and Department of Public Health, University of Helsinki, 00014 Helsinki, Finland. [15]Welch Center for Prevention, Epidemiology and Clinical Research, The Johns Hopkins Medical Institutions, Baltimore, Maryland 21205, USA. [16]Geriatric Research and Education Center, Veterans Administration Medical Center, Baltimore, Maryland 21201, USA. [17]Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas 77225, USA. [18]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [19]These authors contributed equally to this work. Correspondence should be addressed to K.L.M. (mohlke@med.unc.edu) or G.R.A. (goncalo@umich.edu).

Received 19 September 2007; accepted 29 November 2007; published online 13 January 2008; doi:10.1038/ng.74

Table 1 Subject characteristics

Study population	Gender (M/F)	Study age (years)[a]	Standing height (cm)[a]	Sitting height (cm)[a]	Body mass index (kg/m^2)[a]
FUSION T2D stage 1	617/467	62.7 (7.6)	167.3 (9.0)	–	30.2 (4.7)
FUSION NGT stage 1	640/647	60.9 (11.2)	167.4 (9.3)	–	27.0 (3.9)
SardiNIA	1,883/2,415	43.6 (17.5)	159.9 (9.0)	–	25.3 (4.7)
FUSION T2D stage 2	718/490	59.4 (8.7)	168.9 (9.6)	–	30.9 (5.4)
FUSION NGT stage 2	768/490	58.4 (7.7)	169.1 (9.3)	–	26.9 (3.9)
DGI T2D	772/745	58.8 (10.0)	167.7 (9.1)	–	28.5 (4.4)
DGI controls	709/759	64.2 (10.0)	168.9 (8.9)	–	26.7 (3.7)
Old Order Amish	1,253/1,458	49.6 (16.9)	164.9 (8.8)	–	27.2 (5.1)
ARIC European Americans	5,374/5,996	54.4 (5.7)	168.6 (9.4)	89.4 (4.7)	27.0 (4.9)
ARIC African Americans	1,600/2,595	53.6 (5.8)	168.1 (8.9)	86.2 (4.4)	29.6 (6.2)
Caerphilly	1,389/0	56.7 (4.3)	171.2 (6.4)	91.1 (3.3)	26.3 (3.5)
BWHHS	0/3,685	68.9 (5.5)	158.7 (6.1)	83.0 (3.6)	27.6 (5.0)
Total samples	15,723/19,747				

[a]Values are the mean (s.d.).

~2.28 million common genetic variants. After verifying the overall accuracy of imputed genotypes in a few markers, we conducted within-study analyses with a rapid variance components-based association test[8] and then carried out a meta-analysis of the two studies (**Supplementary Fig. 1** online).

Our results provided confirmatory evidence of the association of height with rs1042725 and rs7968682, two common variants in *HMGA2* ($P = 0.031$ and 0.0093, respectively, both in the same direction as the original report[1]; **Supplementary Table 1** online). The five loci showing the most significant evidence of association in our study are listed in **Supplementary Table 2** online. To our knowledge, common variants in these loci have not previously been associated with height.

The genes located on chromosome 20 near our strongest signal ($P < 2 \times 10^{-7}$) have a plausible biological role in human height. Rare variants in growth differentiation factor 5 (*GDF5*) have been associated with disorders of skeletal development (see below), and *UQCC* (also known as *BFZB* or *C20orf44*) encodes a ZIC-binding

Table 2 Association between rs6060369 and height[a]

Study population	*n*	C/C	C/T	T/T	Allele freq. (C)	Effect (s.e.m.) cm	Effect (s.e.m.) standardized[b]	*P* value
FUSION T2D stage 1[c]	1,084	167.4 (8.8)	167.5 (8.9)	167.0 (9.2)	0.426	0.382 (0.259)	0.081 (0.045)	0.072
FUSION NGT stage 1[c]	1,287	167.5 (9.7)	167.7 (9.2)	166.7 (9.0)	0.449	0.634 (0.241)	0.124 (0.041)	0.0024
SardiNIA[d]	4,298	158.8 (8.2)[d]	158.5 (8.3)[d]	158.0 (8.7)[d]	0.384	0.700 (0.186)	0.083 (0.021)	4.35×10^{-4}
	6,669	Stage 1 meta-analysis						9.73×10^{-7}
FUSION T2D stage 2	1,154	168.8 (9.3)	169.4 (9.8)	168.5 (9.5)	0.458	0.477 (0.262)	0.071 (0.041)	0.084
FUSION NGT stage 2	1,203	170.3 (9.4)	168.6 (9.2)	168.9 (9.3)	0.451	0.456 (0.244)	0.103 (0.040)	0.0090
DGI T2D[c]	1,517	168.4 (9.3)	167.6 (9.1)	167.4 (9.1)	0.447	0.449 (0.234)	0.062 (0.038)	0.044
DGI controls[c]	1,468	169.0 (8.6)	169.1 (9.0)	168.3 (8.8)	0.425	0.323 (0.249)	0.069 (0.042)	0.038
Old Order Amish	2,711	165.4 (9.1)	165.3 (9.0)	164.2 (8.5)	0.383	0.424 (0.178)	0.044 (0.020)	0.028
ARIC European Americans	10,882	168.9 (9.3)	168.7 (9.4)	168.3 (9.5)	0.398	0.252 (0.085)	0.029 (0.009)	0.0020
ARIC African Americans	3,860	168.4 (9.1)	167.9 (8.7)	167.3 (8.5)	0.710	0.254 (0.160)	0.025 (0.019)	0.169
Caerphilly	1,097	171.6 (6.6)	171.3 (6.5)	170.8 (6.2)	0.370	0.522 (0.270)	0.083 (0.042)	0.055
BWHHS	3,652	159.1 (6.0)	159.0 (6.1)	158.2 (6.2)	0.362	0.560 (0.147)	0.093 (0.240)	9.71×10^{-5}
	27,544	Stage 2 meta-analysis						1.05×10^{-11}
	34,213	Standing height overall meta-analysis						2.22×10^{-16}
Sitting height								
ARIC European Americans	10,863	89.6 (4.6)	89.4 (4.6)	89.3 (4.6)	0.398	0.137 (0.046)	0.029 (0.009)	0.0036
ARIC African Americans	3,857	86.3 (4.5)	86.3 (4.2)	86.1 (4.1)	0.710	−0.043 (0.085)	−0.008 (0.019)	0.679
Caerphilly	1,092	91.3 (3.4)	91.2 (3.4)	90.9 (3.2)	0.370	0.275 (0.138)	0.087 (0.041)	0.038
BWHHS	3,655	83.3 (3.6)	83.2 (3.6)	82.7 (3.4)	0.362	0.345 (0.083)	0.100 (0.023)	1.73×10^{-5}
	19,467	Sitting height meta-analysis						1.40×10^{-5}

[a]Association results are shown for an additive genetic model. Height means (s.d.) in cm are shown for each genotype class. Effect sizes are annotated with the corresponding standard error (s.e.m.). [b]Standardized effects are relative to the regression model when using the normalized trait and represent the increase in height in s.d. units, on average, for each additional copy of the C allele. *P* values correspond to standardized effects. [c]Genotypes for individuals not successfully genotyped for rs6060369 were imputed to increase the call rate to 100%. [d]SardiNIA genotype means (s.d.) are given for the 1,412 individuals genotyped with the Affymetrix Mapping 500K Array; the effect size estimates and *P* values represent the analysis of 4,298 individuals including those with either experimentally derived or imputed genotypes.

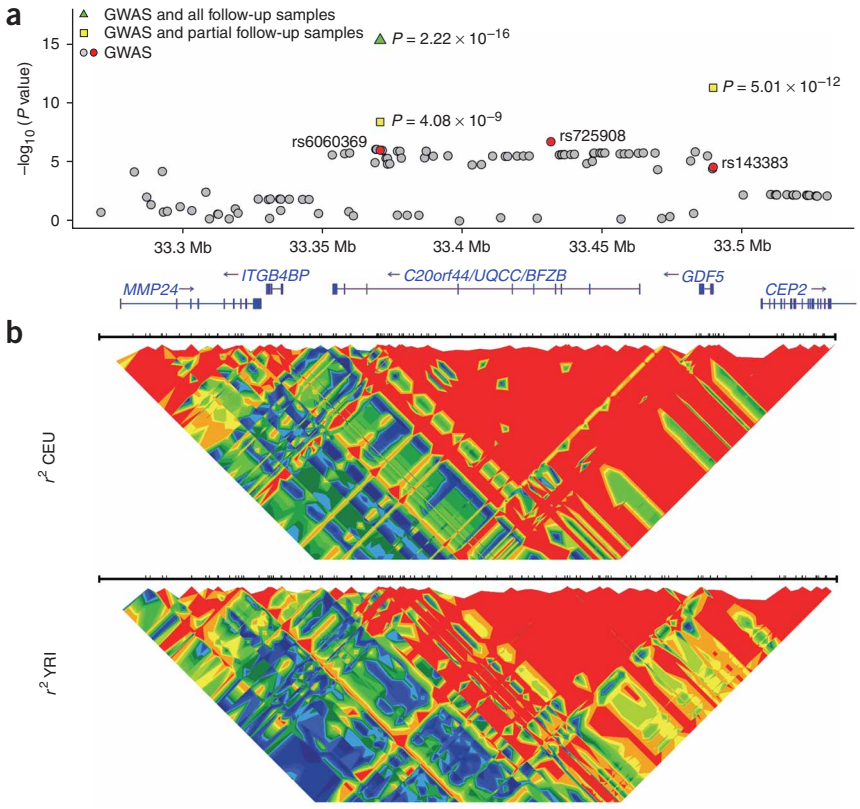

Figure 1 Evidence of association with height and linkage disequilibrium around *GDF5* and *UQCC*. (**a**) Plot of all genotyped or imputed SNPs in the SardiNIA and FUSION GWASs with association *P* values (additive test) against genomic position in NCBI Build 35 (gray circles; red circles indicate labeled SNPs). Yellow squares indicate *P* values for SNPs analyzed in a portion of follow-up samples (FUSION stage 1 and 2, SardiNIA, DGI and ARIC studies). Green triangle indicates the rs6060369 SNP that was analyzed in all GWAS and follow-up samples. (**b**) Patterns of linkage disequilibrium (r^2) in two of the HapMap populations (CEPH (Utah residents with northern and western European ancestry)(CEU) and Yoruba in Ibadan, Nigeria (YRI))[29] plotted and colored with a 15-color scale representing r^2 values ranging from high (red), to intermediate (green), to low (blue).

protein repressed by basic fibroblast growth factor[9] (bFGF; also known as FGF2).

We pursued the chromosome 20 signal because it was the single best result in our initial scan, the surrounding region accounts for 40 of the 50 lowest *P* values in our meta-analysis, and it overlaps a locus associated with osteoarthritis susceptibility[2]. We focused our follow-up efforts for this locus on SNP rs6060369 ($P = 9.7 \times 10^{-7}$ in the initial meta-analysis; **Table 2**) because it showed the strongest evidence for association among all of the Affymetrix Mapping 500K SNPs that had been assessed in our GWAS samples; we favored genotyped over imputed SNPs for initial follow up. The SNP that was most significantly associated with height in the meta-analysis of imputed HapMap SNPs was rs725908, which is in strong linkage disequilibrium (LD) with rs6060369 (square of correlation coefficient (r^2) = 0.96, **Supplementary Table 2**). The rs6060369 SNP was initially imputed in the FUSION GWAS, but direct genotyping sustained strong evidence of association (meta-analysis, $P = 1.5 \times 10^{-6}$). In the GWAS samples, each copy of the C allele at rs6060369 was associated with an increase in height of 0.3–0.7 cm (accounting for 0.3–0.6% of the variance in height, after adjustment for age and gender).

Motivated by previous reports of gender differences at osteoarthritis-associated loci[10], we carried out an analysis stratified by gender. The stratified analyses show strong evidence of association in both males and females, and no evidence of heterogeneity (**Supplementary Table 3** online). We did not detect significant association of the SNP with other anthropometric measures ($P > 0.05$ for weight, body mass index, waist circumference, hip circumference and waist-to-hip ratio).

We tested the association of rs6060369 with height in nine follow-up samples, comprising 23,684 individuals of European ancestry and 3,860 African American individuals (**Table 2**). Six of the samples provided significant evidence of association ($P < 0.05$); the other three showed a trend ($P < 0.20$) in the same direction (**Table 2**). The *P* value for association in all 27,544 follow-up samples was 1.05×10^{-11}, and in all 34,213 GWAS and follow-up samples combined was 2.22×10^{-16} (**Fig. 1**). In the follow-up samples, each copy of the C allele at rs6060369 was associated with an increase in height of 0.2–0.6 cm (**Table 2**); overall, we estimate that this SNP explains 0.3–0.5% of the variance in height in both males and females (**Supplementary Table 3**). Further independent evidence for association between rs6060369 and adult height comes from the 1958 British Birth Cohort, in which rs6060369 is associated with height measured at 44–45 years of age ($P = 0.0046$, explaining 0.5% of the variance; URL accessed October 2007).

Our association signal lies in a 136-kb stretch of LD that contains two genes, *GDF5* and *UQCC* (**Fig. 1** and **Supplementary Table 4** online). *UQCC* is present in differentiating chondrocytes[11], and is first expressed at early stages of mesenchymal cell proliferation[12]. Studies in mouse embryonic stem cells have shown that *UQCC* is downregulated on addition of FGF2 (ref. 9), which functions in concert with bone morphogenic proteins and several Hox gene products to initiate and to promote morphogenesis and growth of the skeleton. Thus, *UQCC* seems to be involved in a network of *FGF2*-regulated growth control. *GDF5* is a member of the transforming growth factor-β superfamily and is involved in bone growth and differentiation in both adult and embryonic tissues[13,14]. *GDF5* is typically expressed in the primordial cartilage of long bones, and shows little expression in the vertebrae and ribs[13]. Mutations in *GDF5* are associated with several disorders of skeletal development (MIM 201250, 200700, 112600, 113100, 228900, 185800 and 601146). Recombinant human GDF5 has been shown to restore vertebral disc height, perhaps through enhanced production of extracellular matrix, in a rabbit model of disc degeneration[14]. Other nearby genes do not seem to be involved in chondrocyte differentiation, bone growth or development, but notably the locus that includes *GDF5* and *UQCC* has been found to show strong evidence for selection in a genome-wide search for regions that have undergone recent selection[15]. The target of selection is unknown at present.

An SNP located in the 5′ untranslated region of *GDF5*, rs143383, is strongly associated with osteoarthritis[2,16] and is estimated to be in very strong LD with rs6060369 in the HapMap, FUSION and SardiNIA samples (r^2 = 0.83–0.90). The SNP seems to influence *GDF5*

Table 3 Association between rs143383 and height[a]

Study population	*n*	G/G	G/A	A/A	Allele freq (G)	Effect (s.e.m.) cm	Effect (s.e.m.) standardized[b]	*P* value
FUSION T2D stage 1[c]	1,084	167.3 (8.8)	167.8 (9.0)	166.7 (9.0)	0.406	0.461 (0.281)	0.088 (0.049)	0.071
FUSION NGT stage 1[c]	1,287	167.8 (9.8)	167.7 (9.3)	166.7 (9.0)	0.425	0.697 (0.263)	0.129 (0.044)	0.0037
SardiNIA[d]	4,298	158.9 (8.2)	158.5 (8.4)	157.9 (8.6)	0.403	0.546 (0.189)	0.065 (0.021)	6.73×10^{-3}
	6,669	Stage 1 meta-analysis						2.70×10^{-5}
FUSION T2D stage 2	1,167	168.9 (9.2)	169.1 (9.8)	168.5 (9.6)	0.436	0.417 (0.262)	0.057 (0.041)	0.166
FUSION NGT stage 2	1,216	170.2 (9.4)	168.8 (9.4)	168.8 (9.2)	0.432	0.460 (0.247)	0.098 (0.040)	0.014
DGI T2D[d]	1,517	168.4 (9.3)	167.7 (9.1)	167.3 (9.1)	0.422	0.550 (0.237)	0.080 (0.039)	0.018
DGI controls[d]	1,468	169.1 (8.5)	169.1 (8.9)	168.3 (8.9)	0.392	0.359 (0.255)	0.070 (0.043)	0.036
Old Order Amish	–	–	–	–	–	–	–	–
ARIC European Americans	10,857	168.9 (9.3)	168.8 (9.5)	168.3 (9.5)	0.387	0.257 (0.086)	0.029 (0.009)	0.0019
ARIC African Americans	3,881	168.2 (9.0)	167.4 (8.5)	167.3 (8.4)	0.879	0.608 (0.222)	0.065 (0.025)	0.011
Caerphilly	–	–	–	–	–	–	–	–
BWHHS	–	–	–	–	–	–	–	–
	20,106	Stage 2 meta-analysis						8.48×10^{-8}
	26,775	Standing height overall meta-analysis						5.01×10^{-12}
Sitting height								
ARIC European Americans	10,838	98.6 (4.5)	89.5 (4.7)	89.3 (4.6)	0.387	0.122 (0.046)	0.026 (.010)	0.0098
ARIC African Americans	3,878	86.3 (4.3)	86.3 (4.1)	86.5 (4.5)	0.879	−0.100 (0.119)	−0.027 (.027)	0.318
Caerphilly	–	–	–	–	–	–	–	–
BWHHS	–	–	–	–	–	–	–	–
	14,716	Sitting height meta-analysis						0.088

[a]Association results are shown for an additive genetic model. Height means (s.d.) in cm are shown for each genotype class. Effect sizes are annotated with the corresponding standard error (s.e.m.). [b]Standardized effects are relative to the regression model when using the normalized trait and represent the increase in height in s.d. units, on average, for each additional copy of the G allele. *P* values correspond to standardized effects. [c]Genotypes for individuals not successfully genotyped for rs143383 were imputed to increase the call rate to 100%. [d]The rs143383 marker was not genotyped in SardiNIA or DGI samples; the data are based on the most likely genotypes from imputation.

expression[2,16] and, we reasoned, might be a causal variant. We therefore analyzed this SNP in our screening samples and a few of our follow-up samples. The rs143383 A allele, which was previously associated with increased risk of osteoarthritis, was associated with a decrease in height in our studies ($P = 5.01 \times 10^{-12}$ versus $P = 4.08 \times 10^{-9}$ for rs6060369 in the same subset of samples; **Fig. 1** and **Table 3**). Analysis stratified by gender showed strong association in both males and females (**Supplementary Table 5** online).

The ARIC African American samples, which had only a trend toward association with rs6060369 ($P = 0.17$), showed significant evidence of association with rs143383 ($P = 0.011$), illustrating the utility of studying different ancestral groups in the fine-mapping of complex disease genes[17,18]. In the ARIC African American samples, rs143383 remained marginally associated with height even when rs6060369 was included in a regression model ($P = 0.034$, estimated additive effect = 0.579 cm), and the association of rs6060369 disappeared ($P = 0.92$, estimated additive effect = −0.019 cm) when conditioning on rs143383. These results suggest that *GDF5* is more likely to influence height, although other non-synonymous SNPs present in *GDF5* and *UQCC* may affect function instead or in addition.

The A allele of rs143383 is associated with decreased transcriptional activity of *GDF5* in chondrogenic cells[2]. Lower expression of *GDF5* would logically lead to a reduction in limb bone growth, consistent with decreased height, as we observed. Lower transcription of *GDF5* may influence the amount of cartilage in the vertebrae, limb proportions or joint angles, resulting in both a modest decrease in stature and susceptibility to osteoarthritis.

To evaluate the impact of osteoarthritis as a confounding factor, we repeated the association analysis on younger individuals (<40 y). In the SardiNIA set, we analyzed 1,964 individuals and confirmed the association ($P = 0.0018$ for rs6060369; $P = 0.015$ for rs143383), with an effect size similar to estimates for the combined sample (0.70 cm per copy of the C allele for rs6060369). In the Old Order Amish set, the younger subgroup of 891 individuals showed a trend toward the association of allele C with increased height (0.60 cm per copy of the C allele at rs6060369), but this trend was not significant ($P = 0.86$), probably owing to low statistical power.

To compare the evidence of association with length of long bones as compared with that of vertebrae and skull, we tested rs6060369 and rs143383 for evidence of association with sitting height, which was measured in the ARIC and BWHHS studies. In ARIC European Americans and the BWHHS British sample, similar evidence of association was observed for both standing and sitting height. In ARIC African Americans, only rs143383 was significantly associated ($P < 0.05$) with height, and it was associated only with standing height and not with sitting height (**Table 3**), perhaps suggesting that the association has a stronger effect on long bones than on vertebrae.

Multiple regression analysis of our data suggests that a single common variant in the region may underlie the evidence of association. Specifically, multiple regression analysis in the GWAS samples showed that, after including rs6060369, rs143383 or rs725908 as a covariate, other association signals in the region become nonsignificant. One of these common variants, or another nearby unmeasured variant in LD, may influence height through effects on *GDF5* expression[2,16]; however, *GDF5*, *UQCC* or both could be affected. Thoroughly

evaluating the contribution of this locus to human height will require resequencing the region to catalog all genetic variants and genotyping to evaluate their effects.

Combined, the variants identified here and those previously reported in *HMGA2* account for <1% of the variance in height; thus, most of the 80% of variation in height that is estimated to be under genetic control remains unexplained. Our GWAS provides evidence to suggest that several other loci influence height. For example, after excluding SNPs within 250 kb of *GDF5*, we observed a slight excess of SNPs with $P < 10^{-5}$ (38 observed versus 23 expected; **Supplementary Fig. 2** online). It seems likely that many of the common variants influencing height will have only small effects. Follow-up of additional SNPs in larger meta-analyses will be necessary to define these variants[19], which may be relevant not only to normal variation in height but also to musculoskeletal or other diseases.

METHODS

Study subjects. Informed consent was obtained from all study participants and ethics approval was obtained from the participating institutions.

FUSION GWAS. The FUSION study GWAS included 1,161 Finns with type 2 diabetes (T2D), 1,174 normal glucose tolerant (NGT) controls and 122 offspring of case-control pairs[6]. Cases and controls were approximately frequency-matched, by taking into account age, gender and birth province within Finland[6]. For the height analysis, our sample consisted of 1,084 individuals with T2D and 1,287 NGT controls with height measurements from clinical exams. Samples were genotyped with an Infinium II HumanHap300 BeadChip[6] (Illumina) and with a GoldenGate Custom Panel (Illumina) designed to improve genomic coverage around T2D candidate genes. The two imputed SNPs selected for additional follow up were subsequently genotyped by using a TaqMan allelic discrimination assay (Applied Biosystems).

SardiNIA GWAS. The SardiNIA GWAS examined a total of 4,305 related individuals participating in a longitudinal study of aging-related quantitative traits in the Ogliastra region of Sardinia, Italy. These individuals are distributed across 1,133 inter-related sibships, each with an average of 3.9 phenotyped individuals. For this study, we analyzed phenotypes for 4,298 individuals, excluding four with short stature due to Morquio syndrome (MIM 253000) and three for whom height measurements were not available. Among the individuals examined, 1,412 were genotyped with the Affymetrix Mapping 500K Array Set. Taking advantage of the relatedness among individuals in the SardiNIA sample, we conducted a second round of computational analysis to impute genotypes for analysis in an additional 2,893 individuals who were genotyped only with the Affymetrix Mapping 10K Array Set. In this second round, we identified large stretches of chromosome shared within each family, and we probabilistically 'filled-in' genotypes within each stretch whenever one or more of its carriers was genotyped with the 500K Array Set[8,20]. These 2,893 individuals were mostly offspring and siblings of the 1,412 individuals genotyped at high density. For computational efficiency, the second round of imputation was applied to subpedigrees, which each included no more than 20–25 individuals.

Follow-up samples and genotyping. From the results of the combined SardiNIA and FUSION results, SNPs rs6060369 and rs143383 were each examined in up to 28,801 additional individuals. These included individuals of European ancestry from the FUSION stage 2 study ($n = 2,466$), the DGI (Diabetes Genetics Initiative[21]; $n = 2,985$), the Old Order Amish[22,23] ($n = 2,711$), the ARIC (Atherosclerosis Risk in Communities) Study[24] ($n = 11,370$), the Caerphilly Study[25,26] (1,389 men) and the BWHHS (British Women's Heart and Health Study)[27] (3,685 women). We also examined 4,195 African American individuals from the ARIC study. Additional details of follow-up samples and genotyping methods are included in the **Supplementary Methods** online. Within each follow-up sample, SNP genotype success rates were >90% and genotype counts were consistent with Hardy-Weinberg equilibrium ($P > 0.05$).

Genotype imputation. Our initial screen was based on the meta-analysis of two genome-wide association studies. Because the studies used two different genotyping platforms, we imputed genotypes for all polymorphic HapMap SNPs in each study by using a hidden Markov model programmed in MACH[6]. Details are provided in the **Supplementary Methods**.

GWAS analysis. Within the FUSION and SardiNIA study samples, we carried out association analyses to relate observed and estimated genotypes to height. For each SNP, height was related to allele counts for a reference allele in a regression model that also included gender and age^2 as covariates; FUSION covariates also included birth province and study[6]. For SNPs genotyped in the laboratory, allele counts were discrete (0, 1 or 2), whereas, for imputed SNPs, allele counts were fractional (between 0.0 and 2.0, depending on the number of copies of the allele expected for each individual). For the FUSION set, T2D and control individuals were analyzed separately and the results were combined by using the meta-analytic techniques described below. To allow for relatedness, regression coefficients were estimated in the context of a variance components-based model that can handle imputed genotypes and can account for background polygenic effects[8]. When evaluating significance, we applied quantile normalization to trait values (SardiNIA) or to residuals after adjustment for covariates (FUSION), by ranking all height values and then converting them to z scores according to quantiles of the standard normal distribution. In parallel to the analysis of quantile normalized data, we also analyzed untransformed height (in centimeters) to estimate effect sizes.

Meta-analysis. To summarize results for the three initial scans (1,084 T2D cases, 1,287 controls from FUSION and 4,298 individuals from Sardinia), we carried out a meta-analysis. We used a meta-analysis rather than an analysis of pooled data to avoid an increase in false-positive rates owing to population stratification. The Sardinian and Finnish populations are genetically and geographically distinct, with an average Fst of 0.025 among the 45,284 autosomal SNPs genotyped in both samples, and with clear differences in height. Genetic differentiation was estimated using Weir's unbiased estimator of Fst, calculated using the variance in allele frequencies between samples and standardized according to the mean allele frequency in the combined sample. The genomic control parameter for our meta-analysis, which estimates inflation in test statistics owing to the combined effects of population stratification, cryptic relatedness and genotyping error[28], was 1.02, suggesting both that population stratification and unmodeled relatedness had a negligible impact on our association results and that our meta-analysis of imputed genotypes resulted in appropriate control of false-positive rates.

For each marker, we selected an arbitrary reference allele and calculated a z statistic characterizing the evidence for association in each study (summarizing both the P value, in its magnitude, and the direction of effect, in its sign). We then calculated an overall z statistic as a weighted average of the three individual statistics and calculated the corresponding P value[19]. Weights were proportional to the square root of the number of individuals examined in each sample and were selected such that the squared weights summed to 1.0. An analogous strategy was used to summarize results of follow-up genotyping.

Accession numbers. GenBank mRNA sequences: GDF5, NM_000557; FGF2-repressed ZIC-binding protein (*UQCC*), NM_018244.

URLs. Genetic information from the British 1958 Birth Cohort, http://www.b58cgene.sgul.ac.uk/; MACH, http://www.sph.umich.edu/csg/abecasis/MACH/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the individuals who participated in this study; Y. Li for computational analysis; K. Gaulton, A. Swift, M. Erdos and N. Narisu for FUSION genotyping and technical expertise; investigators at the Center for Inherited Disease Research for FUSION GWAS genotyping; staff and participants of the ARIC study for contributions; and the Amish Research Clinic Staff for study subject recruitment and characterization. The SardiNIA team thanks Monsignore Piseddu (Bishop of Ogliastra), the mayors and citizens of the four Sardinian towns (Lanusei, Ilbono, Arzana and Elini), and the head of the Public Health Unit ASL4 for cooperation; A. Maschio, F. Busonero, A. Mulas, N. Sestu and M. Grazia Piras for genotyping and technical expertise; and all of the physicians, nurses and recruitment

personnel of the ProgeNIA group in Lanusei. This research was supported (in part) by the intramural Research Program of the US National Institutes of Health (NIH) National Institute on Aging (contracts NO1-AG-1-2109 to the SardiNIA ('Progenia') team and 263-MA-410953 to the University of Michigan (G.R.A.)); NIH grants DK072193 (K.L.M.), HL084729 (G.R.A.), HG002651 (G.R.A.), DK062370 (M.B.), DK54361 (A.R.S.), HL72515 (A.R.S.), AG18728 (A.R.S.) and AR046838 (A.R.S.); the National Human Genome Research Institute (intramural project number 1 Z01 HG000024 (F.S.C.)); the American Diabetes Association, including a postdoctoral fellowship award (C.J.W.); and March of Dimes (research grant 6-FY04-61 (J.N.H.)). K.L.M. and G.R.A are Pew Scholars in the Biomedical Sciences. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021 and N01-HC-55022). FUSION GWAS genotyping was performed with support from CIDR NIH (contract N01-HG-65403) and the John Hopkins University Genetic Resources Core Facility SNP Center. The whole-genome genotyping and analysis in the Diabetic Genetics Initiative genome scan was supported by Novartis Institutes for BioMedical Research (to D. Altshuler), with additional support from The Richard and Susan Smith Family Foundation/American Diabetes Association Pinnacle Program Project Award (to D. Altshuler, J.N.H. and M.J. Daly). Funding and support were also provided by the University of Maryland General Clinical Research Center (M01 RR 16500), the National Institute of Diabetes and Digestive and Kidney Diseases Clinical Nutrition Research Unit of Maryland (NIH P30 DK072488), the Department of Veterans Affairs, and Veterans Affairs Medical Center Baltimore Geriatric Research, Education and Clinical Center. The Caerphilly study was funded by the UK MRC. The Caerphilly study was undertaken by the former MRC Epidemiology Unit in South Wales and was funded by the UK MRC. The data archive is maintained by the Department of Social Medicine, University of Bristol. We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the MRC (grant G0000934) and the Wellcome Trust (grant 068545/Z/02).

AUTHOR CONTRIBUTIONS

D.S., F.S.C., A.C., M.B., G.R.A. and K.L.M. designed the study; Y.B.-S., S.E., D.A.L., J.T., J.C., J.N.H., A.R.S., D.S., G.D.S., E.B. and A.C. provided material and reagents; R.N., L.L.B., G.U., P.S.C., M.D., S.L., L.C., S.N., K.F.D., E.W.P., R.N.B., R.M.W. and M.U. performed the research; S.S., A.U.J., C.J.W., W.-M.C., H.S., N.T., G.L., H.M.S., L.J.S., G.A., M.B. and G.R.A. analyzed the data; and S.S., A.U.J., C.J.W., A.C., M.B., G.R.A. and K.L.M. wrote the paper.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Weedon, M.N. *et al.* A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat. Genet.* **39**, 1245–1250 (2007).
2. Miyamoto, Y. *et al.* A functional polymorphism in the 5′ UTR of *GDF5* is associated with susceptibility to osteoarthritis. *Nat. Genet.* **39**, 529–533 (2007).
3. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).
4. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2**, e132 (2006).
5. Perola, M. *et al.* Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet.* **3**, e97 (2007).
6. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
7. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
8. Chen, W.M. & Abecasis, G.R. Family based association tests for genome wide association scans. *Am. J. Hum. Genet.* **81**, 913–926 (2007).
9. Vetter, K. & Wurst, W. Expression of a novel mouse gene 'mbFZb' in distinct regions of the developing nervous system and the adult brain. *Mech. Dev.* **100**, 123–125 (2001).
10. Valdes, A.M. *et al.* Sex and ethnic differences in the association of *ASPN*, *CALM1*, *COL2A1*, *COMP* and *FRZB* with genetic susceptibility to osteoarthritis of the knee. *Arthritis Rheum.* **56**, 137–146 (2007).
11. Imabayashi, H. *et al.* Redifferentiation of dedifferentiated chondrocytes and chondrogenesis of human bone marrow stromal cells via chondrosphere formation with expression profiling by large-scale cDNA analysis. *Exp. Cell Res.* **288**, 35–50 (2003).
12. Goldring, M.B., Tsuchimochi, K. & Ijiri, K. The control of chondrogenesis. *J. Cell. Biochem.* **97**, 33–44 (2006).
13. Chang, S.C. *et al.* Cartilage-derived morphogenetic proteins. New members of the transforming growth factor-β superfamily predominantly expressed in long bones during human embryonic development. *J. Biol. Chem.* **269**, 28227–28234 (1994).
14. Chujo, T. *et al.* Effects of growth differentiation factor-5 on the intervertebral disc—*in vitro* bovine study and *in vivo* rabbit disc degeneration model study. *Spine* **31**, 2909–2917 (2006).
15. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
16. Southam, L. *et al.* An SNP in the 5′-UTR of *GDF5* is associated with osteoarthritis susceptibility in Europeans and with *in vivo* differences in allelic expression in articular cartilage. *Hum. Mol. Genet.* **16**, 2226–2232 (2007).
17. McKenzie, C.A. *et al.* Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I–converting enzyme (ACE). *Hum. Mol. Genet.* **10**, 1077–1084 (2001).
18. Frere, C. *et al.* Fine mapping of quantitative trait nucleotides underlying thrombin-activatable fibrinolysis inhibitor antigen levels by a transethnic study. *Blood* **108**, 1562–1568 (2006).
19. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
20. Burdick, J.T., Chen, W.M., Abecasis, G.R. & Cheung, V.G. *In silico* method for inferring genotypes in pedigrees. *Nat. Genet.* **38**, 1002–1004 (2006).
21. Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
22. Hsueh, W.C. *et al.* Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* **23**, 595–601 (2000).
23. Streeten, E.A. *et al.* Reduced incidence of hip fracture in the Old Order Amish. *J. Bone Miner. Res.* **19**, 308–313 (2004).
24. ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am. J. Epidemiol.* **129**, 687–702 (1989).
25. The Caerphilly and Speedwell Collaborative Group. Caerphilly and Speedwell collaborative heart disease studies. *J. Epidemiol. Community Health* **38**, 259–262 (1984).
26. Fehily, A.M., Butland, B.K. & Yarnell, J.W. Body fatness and frame size: the Caerphilly study. *Eur. J. Clin. Nutr.* **44**, 107–111 (1990).
27. Lawlor, D.A., Bedford, C., Taylor, M. & Ebrahim, S. Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *J. Epidemiol. Community Health* **57**, 134–140 (2003).
28. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
29. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).