

A common variant of *HMGA2* is associated with adult and childhood height in the general population

Michael N Weedon^{1,2,21}, Guillaume Lettre^{3,4,21}, Rachel M Freathy^{1,2,21}, Cecilia M Lindgren^{5,6,21}, Benjamin F Voight^{3,7}, John R B Perry^{1,2}, Katherine S Elliott⁵, Rachel Hackett³, Candace Guiducci³, Beverley Shields², Eleftheria Zeggini⁵, Hana Lango^{1,2}, Valeriya Lyssenko^{8,9}, Nicholas J Timpson^{5,10}, Noel P Burt³, Nigel W Rayner⁶, Richa Saxena^{3,7,11}, Kristin Ardlie³, Jonathan H Tobias¹², Andrew R Ness¹³, Susan M Ring¹⁴, Colin N A Palmer¹⁵, Andrew D Morris¹⁶, Leena Peltonen^{3,17,18}, Veikko Salomaa¹⁹, The Diabetes Genetics Initiative, The Wellcome Trust Case Control Consortium, George Davey Smith¹⁰, Leif C Groop^{8,9}, Andrew T Hattersley^{1,2}, Mark I McCarthy^{5,6,21}, Joel N Hirschhorn^{3,4,20,21} & Timothy M Frayling^{1,2,21}

Human height is a classic, highly heritable quantitative trait. To begin to identify genetic variants influencing height, we examined genome-wide association data from 4,921 individuals. Common variants in the *HMGA2* oncogene, exemplified by rs1042725, were associated with height ($P = 4 \times 10^{-8}$). *HMGA2* is also a strong biological candidate for height, as rare, severe mutations in this gene alter body size in mice and humans, so we tested rs1042725 in additional samples. We confirmed the association in 19,064 adults from four further studies ($P = 3 \times 10^{-11}$, overall $P = 4 \times 10^{-16}$, including the genome-wide association data). We also observed the association in children ($P = 1 \times 10^{-6}$, $N = 6,827$) and a tall/short case-control study ($P = 4 \times 10^{-6}$, $N = 3,207$). We estimate that rs1042725 explains $\sim 0.3\%$ of population variation in height (~ 0.4 cm increased adult height per C allele). There are few examples of common genetic variants reproducibly associated with human quantitative traits; these results represent, to our knowledge, the first consistently replicated association with adult and childhood height.

Adult height is a classic polygenic trait. The genetics of height were central to the mendelian versus biometrician debate in the early part of the twentieth century that was resolved by Fisher, who proposed that height and other human phenotypes showed multifactorial inheritance¹. Twin, family and adoption studies suggest that up to 90% of normal variation in human height within populations is due to genetic variation²⁻⁶. Severe mutations in several genes cause rare syndromes with extreme stature; however, these cannot explain normal population height variation⁷. Many regions of the genome have been linked with height based on numerous genome-wide linkage scans, with some overlap between studies⁶, but thus far there have not been any examples of gene variants that are reproducibly associated with height variation in the general population.

The recent flood of data from many genome-wide association (GWA) studies offers new opportunities to identify genes influencing adult height. The identification of such genes will probably provide important insights into how best to dissect the genetics of polygenic quantitative traits. The identification of genes influencing growth may also have important medical implications. Height is associated with several common disorders, including a number of cancers^{8,9}.

¹Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU, UK. ²Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DW, UK. ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁴Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA. ⁵Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. ⁶Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK. ⁷Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁸Department of Clinical Sciences, Diabetes and Endocrinology, University Hospital Malmö, Lund University, 205 02 Malmö, Sweden. ⁹Department of Medicine, Helsinki University Central Hospital, Folkhalsan Genetic Institute, Folkhalsan Research Center and Research Program for Molecular Medicine, University of Helsinki, Helsinki, Finland. ¹⁰The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK. ¹¹Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹²Clinical Science at South Bristol, University of Bristol, Bristol BS2 8AE, UK. ¹³Department of Oral and Dental Science, Bristol Dental School, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, UK. ¹⁴Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK. ¹⁵Population Pharmacogenetics Group, Biomedical Research Centre, and ¹⁶Diabetes Research Group, Division of Medicine and Therapeutics, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, UK. ¹⁷Department of Molecular Medicine, National Public Health Institute, Biomedicum, FI-00290 Helsinki, Finland. ¹⁸Department of Medical Genetics, University of Helsinki, FI-00014 Helsinki, Finland. ¹⁹Department of Epidemiology and Health Promotion, National Public Health Institute, FI-00300 Helsinki, Finland. ²⁰Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²¹These authors contributed equally to this work. Correspondence should be addressed to T.M.F. (Tim.Frayling@pms.ac.uk).

Received 17 April; accepted 3 August; published online 2 September 2007; doi:10.1038/ng2121

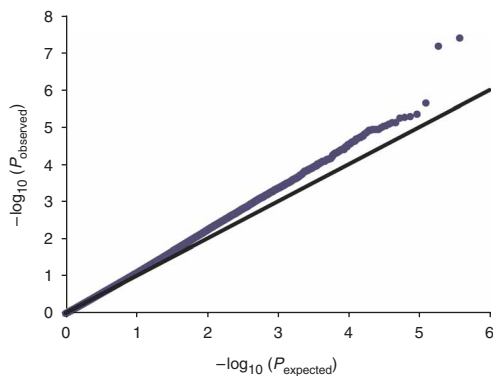


Figure 1 Quantile-quantile plot of 364,301 SNPs from the meta-analysis of DGI and WTCCC genome-wide association statistics. Blue dots represent observed statistics, and black line represents expected statistics.

To begin to identify gene variants influencing adult height, we analyzed GWA data from a total of 4,921 participants. These included 1,896 UK individuals with type 2 diabetes from the Wellcome Trust Case Control Consortium (WTCCC¹⁰; height data on the controls was unavailable) and 3,025 Swedish or Finnish participants (1,496 individuals with type 2 diabetes and 1,529 nondiabetic controls from the Diabetes Genetics Initiative (DGI); ref. 11 and **Supplementary Table 1** online). All participants were of self-reported European ancestry. All DNA samples were genotyped using the Affymetrix GeneChip Human Mapping 500K platform. After performing quality control to exclude poorly performing SNPs, we conducted a meta-analysis of sex- and age-adjusted height Z scores for the 364,301 autosomal SNPs common across the data sets. These SNPs provided 64% coverage of CEU HapMap SNPs (minor allele frequency (MAF) > 5% and $r^2 > 0.8$; see **Supplementary Methods** online).

We created a quantile-quantile (QQ) plot for the meta-analysis (**Fig. 1**). In the individual GWA studies, the most associated SNPs were rs4552313 ($P = 2 \times 10^{-6}$), rs7316173 ($P = 3 \times 10^{-6}$) and rs10804515 ($P = 1 \times 10^{-6}$) in the WTCCC participants, the DGI affected individuals and the DGI controls, respectively, but these did not replicate ($P < 0.05$) across studies, and the combined P was $> 4 \times 10^{-5}$ for all three SNPs. The two SNPs most strongly associated with height in our meta-analysis (**Table 1** and **Fig. 2**), rs1042725 ($P = 4 \times 10^{-8}$) and rs7968682 ($P = 7 \times 10^{-8}$), were in linkage disequilibrium (LD) with each other (r^2 of 0.87 and 0.92 for WTCCC and DGI, respectively) and were the only SNPs to reach a level of statistical significance strongly suggestive of true association ($P < 5 \times 10^{-7}$)¹⁰, after allowing for multiple testing. Association studies of height and other traits may be susceptible to false positive results from genotyping artifacts or population substructure^{12,13}. However, several lines of evidence suggest that the observed association is not artifactual. First, the similar results obtained with two highly correlated SNPs suggest that technical problems in genotyping are unlikely to explain our results. Second, the availability of methods that use dense genome-wide SNP data to estimate and account for ancestry allows us to be confident that the associations are not explained by population stratification. Adjusting for residual population structure using EIGENSTRAT¹⁴ did not substantially alter the strength of the association (WTCCC: unadjusted $P = 9 \times 10^{-4}$; adjusted $P = 1 \times 10^{-3}$; DGI, unrelated individuals only: unadjusted $P = 1 \times 10^{-3}$; adjusted $P = 2 \times 10^{-3}$ for rs1042725), and the genomic control inflation factor¹⁵ was only 1.11 from the meta-analysis. Third, stratification by geographical

region-of-ascertainment did not reduce the strength of the association (rs1042725 meta-analysis $P = 1 \times 10^{-6}$ with and without stratification, using only unrelated subjects from DGI). Fourth, a purely family-based analysis of the sibship-based portion of the DGI sample also contributed evidence for association (982 individuals in 406 sibships; $P = 0.06$, with the same direction of effect). Finally, 13 SNPs that are strongly correlated with the major axis of population differentiation in UK samples did not show association with height (all $P > 0.05$)¹⁰.

In addition to showing strong statistical evidence for association, the two SNPs lie in (rs1042725) and 12 kb downstream (rs7968682) of the 3'-UTR region of the high mobility group-A2 (*HMGA2*) gene, which is a strong biological candidate for influencing height. *Pygmy* mutant mice, which bear homozygous deletions of the orthologous *Hmgic* gene, are short in length¹⁶, whereas mice expressing truncated *Hmgic*, including only the first three exons under the regulatory control of the cytomegalovirus (CMV) promoter, develop gigantism and lipomatosis¹⁷. Furthermore, the autosomal dwarf (adw) phenotype has been mapped to the syntenic region in the chicken¹⁸. In humans, an individual with a severe overgrowth syndrome (stature of 169 cm at age 8 years, > 7 s.d. above the mean) carries a chromosomal inversion that truncates the *HMGA2* gene product¹⁹.

A combination of this strong statistical and biological evidence led us to investigate the role of these SNPs in additional population-based studies. We used primarily rs1042725 in the additional studies because it had a marginally stronger significance level in the initial analysis. We genotyped an additional 29,098 individuals of European ancestry from five studies, including three population-based studies (15,167 adults and 6,827 children), a type 2 diabetes case-control set (3,897 adults) and a collection of 3,207 adults sampled from the near-extremes of the height distribution (the 5th to 10th and 90th to 95th percentiles; **Supplementary Table 1**).

In the replication studies of adults sampled from across the height distribution, each copy of the C allele at rs1042725 was associated with an increase of 0.07 in the adult height Z score (95% confidence interval (c.i.) 0.05–0.09), equivalent to ~ 0.4 cm ($P = 3 \times 10^{-11}$; **Table 1** and **Supplementary Table 2** online). This result provides a strong replication of the original association. When we combined these replication data with the initial results from the genome-wide studies, the statistical evidence in favor of association increased ($P = 4 \times 10^{-16}$). There was little evidence of heterogeneity ($I^2 = 0\%$)²⁰ across all adult studies, no evidence that the effect size was different between males and females ($P = 0.63$) and no evidence for deviation from an additive inheritance model ($P = 0.93$ for likelihood ratio test of additive versus full two-degree of freedom (2-d.f.) model). The C-allele frequency was similar across the different studies (0.48–0.54), further adding to the evidence that population stratification does not explain the observed association. When we used the adults sampled from the near-extremes of the height distribution, we found that each copy of the rs1042725 C allele increased the odds of being in the group of tall individuals (odds ratio = 1.27, 95% c.i. 1.15–1.40, $P = 4 \times 10^{-6}$).

As an initial search for other variants in this region that might be associated with adult height, we genotyped 11 additional SNPs in the population-based FINRISK97 sample ($N = 6,533$) and in the European American adult height panel ($N = 3,207$) drawn from the near-extremes of the height distribution. These 11 SNPs and 20 specific multimarker haplotypes were selected to capture the 42 variants in HapMap phase II build 21a (CEU population) with frequency > 1% that lie within the region of strong LD surrounding rs1042725 (**Supplementary Table 3** online, **Fig. 2** and **Supplementary Methods**). None of these single markers or multimarker haplotypes was

Table 1 Association of adult height with rs1042725 genotypes in participants used for genome wide association studies and replication samples

Study	Gender	Age, in years (mean, s.d.)	Total N ^a	Mean height, in cm (95% c.i.), by genotype			Per-C allele effect size (s.e.m.) ^b	P value ^b
				TT	CT	CC		
GWA								
UK WTCCC (T2D)	Male	58.9 (9.9)	1,104	174.9 (174.1, 175.7)	175.4 (174.9, 176.0)	176.2 (175.4, 177.1)	0.109 (0.031)	0.0006
	Female	57.9 (10.5)	792	160.4 (159.6, 161.3)	161.5 (160.9, 162.2)	162.2 (161.3, 163.1)		
DGI (T2D) ^a	Male	63.4 (10.1)	702	174.4 (173.4, 175.3)	174.2 (173.5, 174.8)	175.4 (174.4, 176.3)	0.108 (0.037)	0.003 ^c
	Female	65.2 (10.1)	638	160.0 (159.0, 160.9)	161.3 (160.6, 162.0)	162.1 (161.2, 163.1)		
DGI (Controls) ^a	Male	58.4 (9.9)	562	175.6 (174.5, 176.7)	176.1 (175.5, 176.8)	175.9 (174.2, 176.8)	0.082 (0.042)	0.003 ^c
	Female	58.5 (9.4)	546	162.1 (161.0, 163.1)	162.8 (162.1, 163.5)	163.7 (162.8, 164.5)		
Combined GWA							0.108 (0.020) ^c	4 × 10 ^{-8c}
Replication samples								
UKT2D GCC (T2D)	Male	64.0 (9.1)	1,138	173.4 (172.6, 174.2)	173.6 (173.0, 174.1)	174.2 (173.4, 175.0)	0.067 (0.032)	0.037
	Female	64.4 (9.7)	820	159.0 (158.1, 159.8)	159.3 (158.7, 160.0)	159.9 (159.1, 160.8)		
UKT2D GCC (Controls)	Male	59.6 (11.6)	988	175.2 (174.4, 176.0)	176.6 (176.1, 177.2)	177.1 (176.3, 177.9)	0.088 (0.032)	0.006
	Female	57.7 (12.2)	951	161.9 (161.1, 162.8)	162.8 (162.3, 163.4)	162.4 (161.7, 163.2)		
ALSPAC mothers (Population-based)	Female	28.4 (4.7)	6,780	163.6 (163.3, 163.9)	164.0 (163.8, 164.2)	164.4 (164.1, 164.7)	0.058 (0.017)	0.0006
EFSOCH parents (Population-based)	Male	32.9 (6.0)	920	176.9 (176.0, 177.8)	178.1 (177.4, 178.7)	178.6 (177.8, 179.5)	0.095 (0.033)	0.004
	Female	30.4 (5.2)	936	164.6 (163.9, 165.4)	165.0 (164.4, 165.6)	165.4 (164.6, 166.2)		
FINRISK97 (Population-based)	Male	47.9 (13.1)	3,023	174.8 (174.3, 175.3)	175.7 (175.3, 176.1)	175.9 (175.4, 176.4)	0.064 (0.017)	0.0002
	Female	47.1 (12.7)	3,508	161.9 (161.4, 162.3)	162.4 (162.1, 162.7)	162.5 (162.2, 163.0)		
Combined replication studies							0.067 (0.010)	3 × 10 ⁻¹¹
All studies							0.074 (0.009)	4 × 10 ^{-16d}

GWA, genome-wide association. T2D, individuals with type 2 diabetes. "Controls" indicates participants selected from the general population with exclusion of individuals with T2D. ^aMeans, 95% c.i. values and total numbers shown for the DGI study are based on unrelated individuals only. ^bP values are calculated using linear regression under an additive model, corrected for age and sex. Effect size (regression coefficient) and standard error values are expressed in s.d. units. ^cFor the DGI GWAS, P values are given for the whole DGI data set (N = 3025); a genomic control method was applied to control for relatedness. Meta-analysis results shown here for GWAS also include the whole DGI data set (using only unrelated samples from DGI, meta-analysis effect size = 0.102, s.e.m. = 0.021 and P value = 1 × 10⁻⁶). ^dOverall meta-analysis results do not include the related component of the DGI study. The results for rs7968682 in the original GWAS were very similar to the results for rs1042725 (GWAS effect size per allele = 0.110 (s.e.m. = 0.02); P = 7 × 10⁻⁸).

more significantly associated with height than rs1042725, and none remained associated after conditioning on rs1042725. Therefore, rs1042725 remains the best explanation for the observed association with height at the *HMGA2* locus. However, which precise SNP(s) in *HMGA2* are functional and how these SNP(s) might alter the expression or function of *HMGA2* is not yet known.

To determine the age at which the association with height appears, we also analyzed the longitudinal Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort of 6,079 children, for whom growth measures were available from birth through to early adolescence, and 748 children from the Exeter Family Study of Childhood Health, for whom birth measures were available. There was no evidence that rs1042725 altered birth length (P ≥ 0.37), but there was a strong association with height at age 7 years, with an increased height of 0.5 cm per C allele (95% c.i. 0.3–0.7, P = 1 × 10⁻⁶); the association persisted at ages 9, 10 and 11 years (Table 2). Using all data from the children at ages 7–11 years, and taking into account the correlation between measurements at the varying time points, the overall effect in the children was 0.07 s.d. per allele (95% c.i. 0.04–0.10; P = 3 × 10⁻⁵), equivalent to ~0.4 cm. These results, showing normal birth measures and increased postnatal growth, are consistent with the normal birth length of the individual with a homozygous disruption of *HMGA2* (ref. 19).

The growth of the spine and the long bones in the limbs proceeds by different mechanisms in humans. Therefore, we examined whether

rs1042725 is associated with variation in limb growth, spine growth or both. In 7-year-old children from the ALSPAC study, each C allele of rs1042725 was associated with both an increase of 0.3 cm in leg length (95% c.i. 0.2–0.4, P = 8 × 10⁻⁷) and an increase of 0.2 cm in sitting height, a measure of the spinal portion of the skeleton (95% c.i. 0.1–0.3, P = 0.0002). This was maintained at ages 8, 9, 10 and 11 years (Supplementary Table 4 online), suggesting that the effect is on general longitudinal skeletal growth. As expected, we observed an association of the rs1042725 C allele with lean mass (P = 0.007) and a consistent trend with bone mass (P = 0.092) as assessed by dual-energy X-ray absorptiometry (DXA) in 5,289 9-year-old children.

Mice lacking a copy of the *HMGA2* homolog (*Hmgic*) have greatly reduced fat mass and are resistant to diet-induced obesity¹⁶. This raised the question of whether the SNPs associated with height also affect body mass index (BMI). We did not find any evidence that the height association signal also affects BMI in adults (Supplementary Table 5 online) or BMI or DXA-assessed fat mass (aged 9) in children (Supplementary Table 6 online).

High-mobility group (HMG) proteins are DNA-binding proteins, often with low-affinity binding sites, and are thought to be involved in altering chromatin structure for regulation of gene expression²¹. Rearrangement of *HMGA2* is a common feature of mesenchymal tumors, most notably lipomas¹⁹. It is not known whether variation at SNP rs1042725 is also associated with variation in cancer risk.

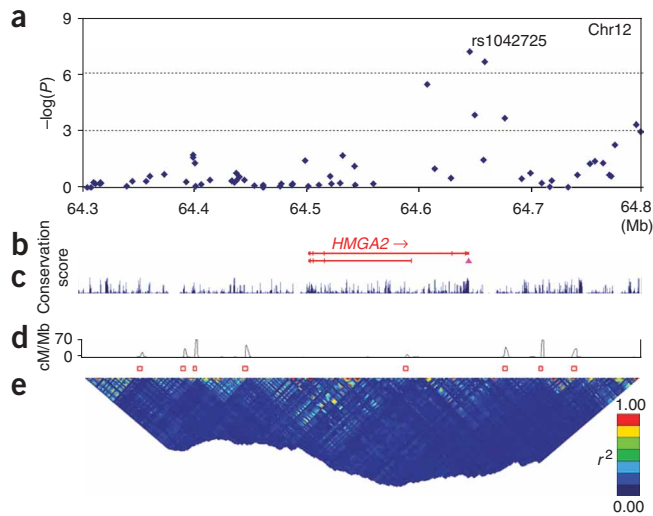


Figure 2 Association, gene structure, conservation and linkage disequilibrium of the *HMGA2* gene region. (a) Plot of $-\log(P)$ versus chromosome position for the WTCCC and DGI meta-analysis. (b) Genomic location of genes showing intron and exon structure (NCBI build 35). (c) Multi-Z vertebrate alignment of 17 species showing evolutionary conservation. (d) Recombination rate given as cM/Mb. Red boxes represent recombination hotspots. (e) GOLDsurfer plot of linkage disequilibrium in HapMap CEU samples (expressed as pairwise r^2).

There are still relatively few examples of common gene variants that influence quantitative population-based traits with convincing evidence in humans. Our data provide an example of a strongly replicated association with a quantitative trait. As the best results from our combined GWAS meta-analysis just reached a P value of only $<1 \times 10^{-7}$, with nearly 5,000 individuals, this study highlights the need for using many thousands of individuals to identify variants underlying polygenic traits with appropriate statistical support. Furthermore, the expected ‘winner’s curse’ phenomenon²²—illustrated in our data by the larger effect size estimates in the GWA data than in the replication samples—aided in our discovery of this variant, suggesting that other variants with similar effect sizes may also be present but with lower levels of significance in the GWA data. Their discovery will require larger sample sizes and/or more aggressive replication efforts. This study also provides additional insights into

the genetic architecture of a classic complex trait. Although height is an accurately measured phenotype with a very strong genetic component, individual common variants have modest effects on this complex trait. We estimate the percentage of the variance of height explained by rs1042725 to be only $\sim 0.3\%$; even taking into account the winner’s curse, the number of as-yet-undiscovered common variants with similar or larger effect sizes must be low, suggesting that hundreds of loci of even smaller effect will ultimately be shown to comprise the genetic basis of height. Fortunately, an increasing amount of high-quality GWA data are becoming available; in combination with well-powered replication efforts, such data should permit the identification of these additional loci for height and other quantitative traits.

In conclusion, common variation in the *HMGA2* gene is associated with adult height in multiple studies at very high levels of statistical confidence. The effect on growth is present in individuals as early as 7 years of age. To our knowledge, these results represent the first reproducible association of a common variant with human stature and suggest that by analyzing data from many thousands of individuals, it will now finally be possible to dissect the genetics of this highly heritable polygenic trait. Insights gained from the study of height are likely to have general implications for the study of other complex traits and common diseases.

METHODS

Genome-wide association samples. As detailed elsewhere¹⁰, samples in the WTCCC genome-wide scan had type 2 diabetes. All had four grandparents of

Table 2 Association of height and birth length with rs1042725 genotypes in the same children at increasing ages

Cohort	Age, in years	Gender	Total N	Mean trait value, in cm (95% c.i.), by genotype			Per-C allele effect size (s.e.m.) ^a	P value ^a
				TT	CT	CC		
Children ^b								
ALSPAC	7	Male	3,119	126.1 (125.7, 126.5)	126.2 (125.9, 126.4)	126.9 (126.5, 127.3)	0.089 (0.018)	1×10^{-6}
		Female	2,948	125.1 (124.7, 125.5)	125.4 (125.1, 125.7)	126.2 (125.8, 126.6)		
	8	Male	2,663	132.8 (132.4, 133.3)	132.8 (132.4, 133.1)	133.3 (132.9, 133.8)	0.067 (0.020)	0.0006
		Female	2,584	131.7 (131.2, 132.1)	132.0 (131.7, 132.3)	132.7 (132.3, 133.1)		
	9	Male	2,792	139.6 (139.1, 140.1)	139.7 (139.4, 140.0)	140.4 (139.9, 140.8)	0.072 (0.019)	0.0002
		Female	2,763	138.9 (138.4, 139.4)	139.4 (139.0, 139.7)	140.0 (139.5, 140.4)		
	10	Male	2,706	143.7 (143.2, 144.3)	143.7 (143.4, 144.1)	144.6 (144.1, 145.1)	0.077 (0.019)	7×10^{-5}
		Female	2,674	143.6 (143.1, 144.2)	144.2 (143.8, 144.5)	144.8 (144.3, 145.3)		
	11	Male	2,540	150.0 (149.4, 150.6)	149.9 (149.5, 150.3)	150.8 (150.2, 151.3)	0.061 (0.020)	0.002
		Female	2,560	151.0 (150.4, 151.6)	151.6 (151.2, 152.0)	152.0 (151.4, 152.5)		
Birth ^c								
EFSOCH	0	Male	394	50.5 (50.2, 50.9)	50.5 (50.2, 50.7)	50.7 (50.3, 51.1)	0.041 (0.046)	0.37
		Female	354	49.5 (49.2, 49.9)	49.7 (49.5, 50.0)	49.7 (49.3, 50.1)		
ALSPAC	0	Male	3,175	51.1 (51.0, 51.3)	51.1 (51.0, 51.2)	51.1 (51.0, 51.3)	0.010 (0.017)	0.56
		Female	2,904	50.2 (50.1, 50.4)	50.3 (50.2, 50.4)	50.4 (50.2, 50.5)		

^a P values were calculated under an additive model using linear regression, corrected for sex (for height in children at specific ages) or for sex and gestational age (for birth length). Effect size (regression coefficient) and standard error values are expressed in s.d. units. ^bALSPAC children are offspring of the participants included in the adult study (Table 1); thus, associations observed are not completely independent of the adult data. Data are shown at five available ages. ^cALSPAC birth data are for the same participants as those in the child study. EFSOCH data are from newborns of the parents in the adult study. Non-singleton births and individuals born at gestation <36.00 weeks were excluded from the birth length analysis.

exclusively British and/or Irish origin. All subjects used in this study gave written informed consent, and the project protocols were approved by the local research ethics committees. Anthropometric measurements were taken as described previously²³. The DGI GWA study for type 2 diabetes has been described previously¹¹ (see also URLs section below). Participants were of European ancestry from Finland and Sweden. In both studies, extensive quality control steps were taken to exclude poorly performing samples and those of non-European descent^{10,11}.

UK Type 2 Diabetes Genetics Consortium (UK2 GCC). This study has been described previously²⁴. All subjects were self-reported "white" and of European descent, living in the Tayside region of Dundee, UK. Height and weight measurements were made as for the WTCCC samples. This study was approved by the Tayside Medical Ethics Committee, and informed consent was obtained from all subjects.

ALSPAC. ALSPAC recruited pregnant women with expected delivery dates between April 1991 and December 1992 from Bristol, UK²⁵. Self-reported "non-white" individuals were excluded from all analyses. The mothers' height and weight data were self reported. Where the data set included singleton siblings born to the same mother, only the first born was included in the analyses. All multiple births and individuals born before 36 weeks' gestation were excluded from birth length analyses. For the analyses of children 7 to 11 years old, only the first born of each twin pair was included. Birth measurement protocols have been described previously²⁶. At the ages of 7 to 11 years, anthropometric measurements were taken²⁷. At age 9, 7,470 children underwent a whole-body dual-energy X-ray absorptiometry (DXA) scan²⁶.

All aspects of the study were reviewed and approved by the ALSPAC Law and Ethics Committee and by local research ethics committees. Parents gave written consent for children in this study.

Exeter Family Study of Childhood Health (EFSOCH). EFSOCH is a prospective study of parents and children from a consecutive birth cohort²⁸. Subjects were recruited from a postcode-defined region of Exeter, UK between 2000 and 2004 and were of self-reported "white" European descent. Parental height and weight were measured by the research midwife at 28 weeks' gestation. Maternal pre-pregnant weight was self reported. Ethical approval was given by the North and East Devon Local Research Ethics Committee, and informed consent was obtained from the parents of the newborns.

FINRISK97. FINRISK1997 is a population-based risk factor survey carried out by the National Public Health Institute of Finland²⁹ and was approved by the Ethical Committee of the National Public Health Institute on 30 October 1996 (decision number 38/96). The sample was drawn from the national population register for five geographical areas in Finland.

European American and Polish height panels. These height case-control samples have been described previously¹². For both panels, all individuals were self-described "white" or "Caucasian." For the US panel, all subjects were born in the US, and all of their grandparents were born in either the US or Europe. All subjects in the Polish panel were born in Poland, and all grandparents were born in Europe or Russia. All subjects gave informed consent, and approval was obtained from the Institutional Review Board of Children's Hospital, Boston.

Genome-wide analyses. We used 393,453 SNPs from the Affymetrix GeneChip Human Mapping 500K platform, which was used in a recent report of type 2 diabetes association using the same WTCCC type 2 diabetes samples²⁴. For the DGI data, SNP quality control and exclusion criteria are reported in detail elsewhere¹¹ and resulted in the use of 386,731 SNPs. We report the 364,301 autosomal SNPs common across the studies.

For each GWA study, summary statistics from linear regression using Z scores were generated using PLINK³⁰. We obtained a combined result for each SNP using inverse variance meta-analysis from the summary statistic beta and standard error (s.e.m.).

Within-study analyses. Height was normally distributed in all cohorts. For WTCCC, UKT2D GCC, ALSPAC and EFSOCH, gender-specific height Z scores were generated within each study, and age was included as a covariate in

subsequent analyses. FINRISK Z scores were generated by correcting for gender, age and regions of recruitment. For all studies, individuals with heights greater than 4 s.d. from the mean were excluded. We examined the associations between genotype and quantitative traits using linear regression. BMI and DXA fat mass were \log_{10} transformed before analysis, and gestational age was included as a covariate in birth length analyses. To obtain an overall estimate for the association between height and rs1042725 genotype in the children from the ALSPAC study, we performed a linear regression using a generalized estimating equation (GEE) to account for the correlation between height measurements performed repeatedly on each subject.

Meta-analysis. Meta-analysis statistics and plots were generated using StataSE version 9 (StataCorp). We used the inverse variance method to pool continuous data (Z score units). The I^2 statistic²⁰ was used to estimate between-study heterogeneity.

Eigenstrat analyses. For both GWA studies, EIGENSTRAT¹⁴ was run on the full set of markers (~390,000 SNPs) using genotypes from unrelated individuals only. Similar results were obtained when we used the first three or ten main eigenvectors. For the WTCCC sample, an LD-pruned set of 104,766 SNPs (generated using PLINK³⁰) produced similar results.

Extreme height analyses. Statistical analysis was performed using a Cochran-Mantel-Haenszel test, as implemented in PLINK³⁰. The data set was stratified according to the country of origin of the grandparents to account for population stratification within the European American height panel¹².

Family-based analyses. Using the related individuals in the DGI sample (982 individuals in 406 sibships), we performed a family-based test of association. We used the QFAM-WITHIN method, as implemented in PLINK³⁰.

Genotyping and quality control. The genotyping of the initial genome-wide association studies is described elsewhere^{10,11}. Genotyping of the UKT2D GCC and EFSOCH samples was performed using TaqMan SNP genotyping assay (Applied Biosystems) according to the manufacturer's protocol. Genotyping of the ALSPAC cohort was performed by KBiosciences (Hoddesdon) using their own system of fluorescence-based competitive allele-specific PCR (KASPar). Genotyping of FINRISK97 and the GCI Extreme Panel was performed using the platform iPLEX Sequenom MassARRAY. SNP rs1042725 was in HWE in all studies ($P > 0.05$). The duplicate concordance rate was >99.4% for each study. The genotype success rate was >95% in all studies.

URLs. HapMap: <http://www.hapmap.org>; DGI GWA study, <http://www.broad.mit.edu/diabetes/>; ALSPAC: <http://www.alspac.bris.ac.uk>. Details of the KBiosciences fluorescence-based competitive allele-specific PCR (KASPar) assay design are available at <http://www.kbioscience.co.uk>. Details of the iPLEX Sequenom MassARRAY are available at <http://www.sequenom.com/Assets/pdfs/appnotes/8876-006.pdf>

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

ACKNOWLEDGMENTS

For the UK-based studies, collection of the type 2 diabetes cases was supported by Diabetes UK, BDA Research and the UK Medical Research Council (MRC) (Biomedical Collections Strategic Grant G0000649). The UK Type 2 Diabetes Genetics Consortium collection was supported by the Wellcome Trust (Biomedical Collections Grant GR072960). The ALSPAC study was supported by The UK MRC, the Wellcome Trust and the University of Bristol. The Exeter Family Study of Childhood Health was supported by UK National Health Service Research and Development and the Wellcome Trust. We also thank the Exeter University Foundation for funding. The UK GWA genotyping was supported by the Wellcome Trust (076113), and replication genotyping was supported by the Wellcome Trust, Diabetes UK, European Commission (EURODIA LSHG-CT-2004-518153) and the Peninsula Medical School. Personal funding comes from the Wellcome Trust (A.T.H.; Research Leave Fellow; Research Career Development Fellow); UK MRC (J.R.B.P.); Diabetes UK (R.M.F.) and the Throne-Holst Foundation (C.M.L.). M.N.W. is Vandervell Foundation Research Fellow at the Peninsula Medical School. C.M.L. is a University of Oxford Nuffield Department of Medicine Scientific Leader Fellow. C.N.A.P. and A.D.M. are supported by the Scottish executive as part of the Generation Scotland Initiative. We acknowledge

the assistance of many colleagues involved in sample collection, phenotyping and DNA extraction in all the different studies. We thank K. Parnell, C. Kimber, A. Murray and K. Northstone for technical assistance. We thank S. Howell, M. Murphy and A. Wilson (Diabetes UK) for their long-term support for these studies. We also acknowledge the efforts of J. Collier, P. Robinson, S. Asquith and others at KBiosciences for their rapid and accurate large-scale genotyping. Finally, we acknowledge all participants in the various studies.

For the studies using the Scandinavian, US and Polish samples, the work was supported by a March of Dimes grant (#6-FY04-61) to J.N.H. and by a grant from The Center of Excellence in Complex Disease Genetics of the Academy of Finland (EU Projects GenomEUtwin, QLG2-CT-2002-01254) (L.P.). The FINRISK study was supported by the Sigrid Juselius Foundation. We thank members of our laboratories and of the Altshuler and Daly laboratories for helpful discussion, and we gratefully acknowledge all of the participants in the studies. We also thank J. Butler for excellent technical assistance and M. Kuokkanen for logistical help with the FINRISK97 cohort.

The whole-genome genotyping and analysis in the DGI genome scan (see **Supplementary Note** online for contributors) 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). R.S. is supported by a US National Institutes of Health (NIH) Research Service Award. G.L. is supported by a March of Dimes research grant (6-FY04-61). Members of the DGI study group acknowledge support from an NIH/National Heart, Lung, and Blood Institute grant (U01 HG004171), the Burroughs Wellcome Fund and the Doris Duke Charitable Foundation. L.C.G. and members of the Botnia Study were funded by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, the Folkhalsan Research Foundation and Clinical Research Institute HUCH. The Malmö Study was funded by a Linné grant from the Swedish Research Council. L.C.G. is supported principally by the Sigrid Juselius Foundation, the Finnish Diabetes Research Foundation, The Folkhalsan Research Foundation and Clinical Research Institute HUCH. Work in Malmö, Sweden was also funded by a Linné grant from the Swedish Research Council. We thank the Botnia and Skara research teams for clinical contributions, and colleagues at Massachusetts General Hospital, Harvard, Broad and Novartis for discussions.

AUTHOR CONTRIBUTIONS

Members of the WTCCC and DGI are listed in the **Supplementary Note**. M.N.W., G.L., R.M.F. and C.M.L. performed the analyses and wrote the first draft of the paper. J.R.B.P., K.S.E., B.S., E.Z., H.L., N.J.T. and N.W.R. were responsible for production, quality control and cleaning of the GWA data from the WTCCC samples. B.F.V., C.G., V.L., N.P.B. and R.S. were responsible for the production, quality control checking and cleaning of the GWA data from the DGI samples. B.F.V. also performed association testing and EIGENSTRAT analyses for height on the DGI data set. S.M.R. performed the data linking and quality control analysis for the ALSPAC study. R.H. and C.G. genotyped rs1042725 in the FINRISK97 cohort. K.A. is principal investigator for the near-extreme height panel. L.P. and V.S. are principal investigators for FINRISK97. L.C.G. is principal investigator for the Botnia study. G.D.S. is principal investigator of the ALSPAC study. J.H.T. and A.R.N. generated DXA measures in the ALSPAC children. C.N.A.P. and A.D.M. are principal investigators of the Tayside UKT2D-GCC study. A.T.H., M.I.M., J.N.H. and T.M.F. designed and led the study. All authors read and approved the final manuscript.

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

1. Fisher, R.A. The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinburgh*, 399–433 (1918).

2. Macgregor, S., Cornes, B.K., Martin, N.G. & Visscher, P.M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.* **120**, 571–580 (2006).
3. Preece, M.A. The genetic contribution to stature. *Horm. Res.* **45**, 56–58 (1996).
4. Silventoinen, K., Kaprio, J., Lahelma, E. & Koskenvuo, M. Relative effect of genetic and environmental factors on body height: Differences across birth cohorts among Finnish men and women. *Am. J. Public Health* **90**, 627–630 (2000).
5. Silventoinen, K. *et al.* Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).
6. 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)(doi:10.1371/journal.pgen.0030097).
7. Palmert, M.R. & Hirschhorn, J.N. Genetic approaches to stature, pubertal timing, and other complex traits. *Mol. Genet. Metab.* **80**, 1–10 (2003).
8. Davey Smith, G. *et al.* Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *J. Epidemiol. Community Health* **54**, 97–103 (2000).
9. Gunnell, D. *et al.* Height, leg length, and cancer risk: A systematic review. *Epidemiol. Rev.* **23**, 313–342 (2001).
10. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
11. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
12. Campbell, C.D. *et al.* Demonstrating stratification in a European American population. *Nat. Genet.* **37**, 868–872 (2005).
13. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246 (2005).
14. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
15. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
16. Zhou, X., Benson, K.F., Ashar, H.R. & Chada, K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* **376**, 771–774 (1995).
17. Battista, S. *et al.* The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis. *Cancer Res.* **59**, 4793–4797 (1999).
18. Ruyter-Spira, C.P. *et al.* The HMGI-C gene is a likely candidate for the autosomal dwarf locus in the chicken. *J. Hered.* **89**, 295–300 (1998).
19. Ligon, A.H. *et al.* Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas. *Am. J. Hum. Genet.* **76**, 340–348 (2005).
20. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
21. Hock, R., Furusawa, T., Ueda, T. & Bustin, M. HMG chromosomal proteins in development and disease. *Trends Cell Biol.* **17**, 72–79 (2007).
22. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33**, 177–182 (2003).
23. Mills, G.W. *et al.* Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to type 2 diabetes. *Diabetologia* **47**, 732–738 (2004).
24. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
25. Golding, J., Pembrey, M. & Jones, R. ALSPAC—the Avon Longitudinal Study of Parents and Children. I: study methodology. *Paediatr. Perinat. Epidemiol.* **15**, 74–87 (2001).
26. Rogers, I.S. *et al.* Associations of size at birth and dual-energy X-ray absorptiometry measures of lean and fat mass at 9 to 10 y of age. *Am. J. Clin. Nutr.* **84**, 739–747 (2006).
27. Leary, S.D. *et al.* Smoking during pregnancy and offspring fat and lean mass in childhood. *Obesity (Silver Spring)* **14**, 2284–2293 (2006).
28. Knight, B., Shields, B.M. & Hattersley, A.T. The Exeter Family Study of Childhood Health (EFSOCH): study protocol and methodology. *Paediatr. Perinat. Epidemiol.* **20**, 172–179 (2006).
29. Vartiainen, E. *et al.* Cardiovascular risk factor changes in Finland, 1972–1997. *Int. J. Epidemiol.* **29**, 49–56 (2000).
30. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).