

# Genome-wide association study in Han Chinese identifies three novel loci for human height

Yongchen Hao · Xuehui Liu · Xiangfeng Lu · Xueli Yang · Laiyuan Wang · Shufeng Chen · Hongfan Li · Jianxin Li · Jie Cao · Jichun Chen · Ying Li · Liancheng Zhao · Yongyong Shi · Chong Shen · Weili Yan · Jiang He · Jianfeng Huang · Dongfeng Gu

Received: 7 November 2012 / Accepted: 18 February 2013 / Published online: 3 March 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** Human height is a complex genetic trait with high heritability but discovery efforts in Asian populations are limited. We carried out a meta-analysis of genome-wide association studies (GWAS) for height in 6,534 subjects with in silico replication of 1,881 subjects in Han Chinese. We identified three novel loci reaching the genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ), which mapped in or near *ZNF638* (rs12612930,  $P = 2.02 \times 10^{-10}$ ), *MAML2* (rs11021504,  $P = 7.81 \times 10^{-9}$ ), and *C18orf12* (rs11082671,  $P = 1.87 \times 10^{-8}$ ). We also confirmed two loci previously reported in European populations including *CS* (rs3816804,  $P = 2.63 \times 10^{-9}$ ) and

*CYP19A1* (rs3751599,  $P = 4.80 \times 10^{-10}$ ). In addition, we provided evidence supporting 35 SNPs identified by previous GWAS ( $P < 0.05$ ). Our study provides new insights into the genetic determination of biological regulation of human height.

## Introduction

Human height is a complex genetic trait that involves multiple loci with heritability estimates of over 80 % (Silventoinen et al. 2003). Insights into the genetic determination of height will provide a better understanding of human development and growth. With the boom of genome-wide association studies (GWAS) in the past 5 years, researchers have identified a large number of height-related genes and genetic loci (Carty et al. 2012; Estrada et al. 2009; Johansson et al. 2009; Lettre et al. 2008; Liu et al.

Y. Hao and X. Liu contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-013-1280-9) contains supplementary material, which is available to authorized users.

Y. Hao · X. Liu · X. Lu (✉) · X. Yang · L. Wang · S. Chen · H. Li · J. Li · J. Cao · J. Chen · Y. Li · L. Zhao · J. Huang · D. Gu (✉)  
State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China  
e-mail: xiangfenglu@sina.com

D. Gu  
e-mail: gudongfeng@vip.sina.com

L. Wang · H. Li  
National Human Genome Center at Beijing, Beijing, China

Y. Shi  
Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Bio-X Institutes, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China

C. Shen  
Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China

W. Yan  
Division of Clinical Epidemiology, Pediatric Institute, Children's Hospital of Fudan University, Shanghai, China

J. He  
Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

2010; Sanna et al. 2008; Tonjes et al. 2009; Weedon et al. 2007, 2008). In 2010, a large-scale meta-analysis of GWAS reported 180 height-related loci that reached genome-wide significance (Lango Allen et al. 2010). However, single-nucleotide polymorphisms (SNPs) within these loci explain only approximately 10 % of the variation of adult height. These results suggest that there might be height-related genetic variants that need to be discovered (Weedon and Frayling 2008).

Understanding the homogeneity and heterogeneity of genetic basis among populations could provide evidence for translation of genetic findings across different ethnic groups. Up to now, most GWAS of height were conducted in European populations. In Asian populations, GWAS of height were reported in Chinese (Lei et al. 2009), Japanese (Okada et al. 2010), Filipinos (Croteau-Chonka et al. 2011), and Koreans (Cho et al. 2009; Kim et al. 2009, 2010). While a few of the previously reported genetic loci have been replicated in different studies, many novel variations were identified in specific ethnic groups. That prompted us to conduct a meta-analysis of GWAS for human height in 6,534 subjects with in silico replication in 1,881 subjects to identify genetic underpinnings in Chinese. We also evaluated the cumulative effects of height-related loci in our study.

## Materials and methods

### Study population in discovery stage

This study included GWAS discovery studies of 6,534 Han Chinese subjects from Beijing Atherosclerosis Study (BAS) and China Atherosclerosis Study (CAS). The BAS consisted of 505 cases of myocardial infarction and 1,021 controls (Hou et al. 2009). The CAS consisted of 1,010 cases of coronary artery disease and 3,998 controls. Details of studies mentioned above have been published elsewhere (Lu et al. 2012). All the studies obtained approval from the institutional review boards of Fuwai Hospital, the Chinese Academy of Medical Sciences and Peking Union Medical College, and other respective medical institutions. All participants provided written informed consent.

### Measurement of height

Height was measured by trained staffs using a standard protocol (He et al. 2004). Subjects were measured without shoes. Subjects stood on a level floor with feet parallel and pointing forward. Subjects stood with the back and head straight to ensure that the Frankfurt plane was horizontal and the eyes were focused forward. The moveable headboard was gently lowered until it touched the crown of the head.

### Genotyping and quality control

The detailed genotyping information of our discovery studies has been described elsewhere (Lu et al. 2012). In short, BAS subjects were genotyped with the Affymetrix GeneChip® Human Mapping 500 K Array Set, and CAS subjects were genotyped with the Axiom™ Genome-Wide CHB 1 Array. We excluded SNPs with minor allele frequency (MAF) <0.01, genotype call rates below 95 %, or deviations from Hardy–Weinberg equilibrium ( $P$  value <10<sup>-4</sup>). We also excluded samples of gender discordance, high genotype missing rate (>3.0 %), cryptic relatedness (IBD >0.1875), or population outliers. After quality control, 1,526 samples and 367,129 autosomal SNPs remained for BAS, and 5,008 samples and 613,724 autosomal SNPs were retained for CAS for genotype imputation.

### Genotype imputation

In the discovery stage, imputation of ungenotyped SNPs considering for uncertainty was carried out using MACH (Li et al. 2009, 2010) based on the HapMap Phase 2 (JPT + CHB) data. After excluding imputed SNPs with imputation quality scores below a threshold ( $R^2 < 0.30$ ), call rate <0.90, MAF <0.01, and Hardy–Weinberg equilibrium  $P < 1 \times 10^{-5}$ , a total of 1,532,051 SNPs from the BAS and 2,042,781 SNPs from the CAS were retained for subsequent analyses.

### Population stratification

Population structure was evaluated by principal component (PC) analysis in the software package EIGENSTRAT (Price et al. 2006). The first 10 PCs were calculated for each subject. To minimize the effect of population stratification, the first two PCs that could explain most of ancestral variation were used as covariates during the association analysis.

### Association analysis

Height was adjusted for sex, age, and the first two PCs, and then normalized into Z-scores. Association analysis was performed using linear regression in case and control groups separately. Association analyses were performed using PLINK (Purcell et al. 2007). A fixed-effects inverse variance-weighted meta-analysis as implemented by METAL (Willer et al. 2010) was used to combine the results from BAS and CAS study. A quantile–quantile (QQ) plot generated using R (<http://www.r-project.org/>) was used to evaluate the overall significance of the GWAS results and the potential impact of population stratification. The genomic inflation factor ( $\lambda$ ) (Devlin and Roeder 1999)

was estimated from the median of the  $\chi^2$  statistic divided by 0.456.

### Replication and meta-analysis

Top SNPs in loci that achieved genome-wide significance ( $P < 5.0 \times 10^{-8}$ ) or suggestive evidence ( $5.0 \times 10^{-8} < P < 1.0 \times 10^{-5}$ ) were replicated in the GenSalt study (GenSalt Collaborative Research Group 2007). The GenSalt study is a family-based study including 1,881 individuals from 637 families. All GenSalt study participants were Han Chinese recruited from six sites in rural areas of northern China. The institutional review board at all participating institutes approved the GenSalt study. Informed consent was obtained from each participant. Genome-wide SNPs were genotyped using Affymetrix® Genome-Wide Human Array 6.0 at the Affymetrix genotyping facility. To estimate the association between height and SNPs, a mixed linear model was used for the association test and the family structure was taken into account using the GWAF package in R (Chen and Yang 2010). Results of discovery stage and replication stage were combined using fixed-effects inverse variance-weighted meta-analysis as implemented by METAL (Willer et al. 2010).

### Cumulative effects of height-related loci

We assessed the cumulative effects of five height loci that showed genome-wide significance of associations (Table 1). The doses of effect alleles were weighted by its effect size and summed for each individual to calculate weighted height scores. We then calculated average height according to the quintiles of weighted height scores.

## Results

### Genome-wide meta-analysis

The discovery meta-analysis included 6,534 subjects. The main characteristics of study subjects were listed in Supplementary Table 1. The average height of the four study groups ranged from 167.6 to 171.2 cm in males and from 155.0 to 159.2 cm in females. The overall inflation factor of our meta-analysis was 1.06. The QQ plot for the distribution of  $P$  values showed that the observed  $P$  values departed from the expected  $P$  values only at the right end tail of the distribution (Supplementary Fig. 1). This result suggested that the signals discovered in our study were more likely the result of true variants rather than potential population stratification or genotyping error.

Through meta-analysis, we identified two loci that reached the genome-wide significance level of  $P < 5 \times 10^{-8}$ . These included one locus at *CYP19A1* reported in populations of European descent (Lango Allen et al. 2010) and one previously unreported locus at *ZNF638*. The SNP most strongly associated with height at the *CYP19A1* locus (15q21.2) was rs3751599 with a  $P$  value of  $1.86 \times 10^{-9}$  (Table 1). We observed that rs3751599 was in weak linkage disequilibrium (LD) ( $r^2 = 0.27$  in HapMap CHB) with rs16964211 identified in European populations (Lango Allen et al. 2010). The LD pattern of this region on 15q21.2 in populations of Chinese was slightly different from that of the European populations from HapMap data (Supplementary Fig. 2). A cluster of 20 SNPs at the *ZNF638* locus, which spanned a genome region of  $\sim 110$  kb on 2p13.2, were in strong LD with each other. Among these SNPs, the top one was rs12612930 located in *ZNF638* with a  $P$  value of  $1.07 \times 10^{-8}$  (Fig. 1). In addition, we found 15 loci that showed  $P$  value  $< 1 \times 10^{-5}$ , but did not achieve the genome-wide significance threshold (Table 1).

### In silico replication and combined analysis

We selected 17 top SNPs in 17 loci ( $P < 1 \times 10^{-5}$ ) from the discovery study for silico replication in the GenSalt study (GenSalt Collaborative Research Group 2007). Six SNPs showed nominal significance ( $P < 0.05$ ) with  $P$  values ranging from  $4.53 \times 10^{-5}$  to  $4.26 \times 10^{-2}$  in the GenSalt study. Combined analysis of discovery and replication studies strengthened the original associations of the *CYP19A1* and *ZNF638* loci with height (rs3751599,  $P = 4.80 \times 10^{-10}$ ; rs12612930,  $P = 2.02 \times 10^{-10}$ ). We also observed another three loci reached genome-wide significance in combined analysis. Two SNPs rs11021504 on 11q21 (*MAML2*,  $P = 7.81 \times 10^{-9}$ ) and rs11082671 on 18q21.1 (*C18orf12*,  $P = 1.87 \times 10^{-8}$ ) (Fig. 1) were newly identified. Another locus on 12q13.3 (rs3816804,  $P = 2.63 \times 10^{-9}$ ) had been reported previously (Lango Allen et al. 2010).

We evaluated whether the height-related variants identified in our samples from Chinese population were associated with height in Europeans, using the results from the GIANT (Lango Allen et al. 2010). Of the three SNPs, rs12612930 was monomorphic in European populations, and rs2670744, a SNP in perfect LD with rs12612930 in Chinese ( $r^2 = 1$  in HapMap CHB) showed nominal association ( $P = 0.017$ ) with height in the population of European ancestry. The SNP rs4852777, which was approximately 8 kb away from rs12612930, showed most strong signal in GIANT ( $P = 3.30 \times 10^{-4}$ ) at this locus. Associations for the other two SNPs were not detected in data from the GIANT study.

**Table 1** Association results for height loci identified in the Chinese GWAS

SNP <sup>a</sup>	CHR	BP <sup>a</sup>	Nearest Gene	Effect/ other Allele	EAF <sup>b</sup>	Discovery stage			Replication stage			Combined			Reported
						N	Effect size (SE) <sup>c</sup>	P	N	Effect size (SE) <sup>c</sup>	P	N	Effect size (SE) <sup>c</sup>	P	
<b>Significant SNPs</b>															
rs12612930	2	71379293	ZNF638	C/T	0.60	6169	0.103(0.018)	1.07E-08	1817	0.092(0.033)	4.75E-03	7986	0.100(0.016)	2.02E-10	7.69E-01
rs11021504	11	95635096	MAML2	A/T	0.51	6259	0.071(0.016)	8.86E-06	1831	0.135(0.033)	4.53E-05	8090	0.083(0.015)	7.81E-09	2.56E-01
rs3816804	12	54967012	CS	C/T	0.77	4910	0.113(0.024)	2.81E-06	1880	0.150(0.040)	1.72E-04	6790	0.123(0.021)	2.63E-09	4.71E-02
rs3751599	15	49360825	CYP19A1	G/A	0.93	6398	0.208(0.035)	1.86E-09	1794	0.123(0.061)	4.26E-02	8192	0.187(0.030)	4.80E-10	7.69E-01
rs11082671	18	44119595	C18orf12	G/A	0.79	5929	0.095(0.021)	8.97E-06	1775	0.151(0.041)	2.54E-04	7704	0.107(0.019)	1.87E-08	7.53E-01
<b>Suggestive SNPs</b>															
rs56864685	2	168926539	CERS6	G/A	0.87	4729	0.139(0.031)	5.76E-06				4729			6.88E-02
rs17382723	2	241702219	PASK	C/G	0.96	6444	0.214(0.046)	3.28E-06	1879	0.182(0.082)	2.72E-02	8323	0.206(0.040)	2.76E-07	2.73E-01
rs11923600	3	177410145	NAALADL2	G/A	0.71	4753	0.099(0.022)	9.76E-06	1871	0.048(0.036)	1.83E-01	6624	0.084(0.019)	8.22E-06	4.47E-01
rs255646	4	175674188	HPGD	A/C	0.33	4709	0.102(0.022)	3.06E-06				4709			3.31E-01
rs16895057	6	29628856	OR211P	T/C	0.95	6439	0.176(0.038)	3.51E-06	1835	0.087(0.065)	1.81E-01	8274	0.153(0.033)	2.91E-06	4.33E-01
rs13258397	8	112720825	EEF1A1P37	G/C	0.38	6413	0.083(0.018)	5.40E-06	1876	0.025(0.034)	4.64E-01	8289	0.070(0.016)	1.38E-05	6.43E-01
rs7006353	8	123533855	ZHX2	G/T	0.45	6134	0.085(0.018)	3.60E-06	1819	-0.016(0.034)	6.35E-01	7953	0.062(0.016)	1.13E-04	3.72E-02
rs10509091	10	59880796	TFAM	T/G	0.35	6474	0.086(0.018)	3.37E-06	1880	0.033(0.034)	3.32E-01	8354	0.074(0.016)	5.55E-06	6.86E-01
rs4246215	11	61320875	FEN1	G/T	0.69	6468	0.087(0.019)	4.11E-06	1880	0.025(0.036)	4.90E-01	8348	0.074(0.017)	1.06E-05	3.13E-01
rs10773568	12	127752143	TMEM132C	A/G	0.42	6285	0.089(0.018)	6.46E-07	1880	0.019(0.034)	5.87E-01	8165	0.074(0.016)	3.11E-06	3.51E-01
rs5676819	13	108788141	MYO16	A/G	0.85	4531	0.129(0.029)	8.07E-06				4531			3.39E-01
rs154663	16	88253536	C16orf55	T/C	0.29	6387	0.093(0.020)	1.96E-06	1856	0.055(0.037)	1.32E-01	8243	0.084(0.017)	9.20E-07	2.25E-01

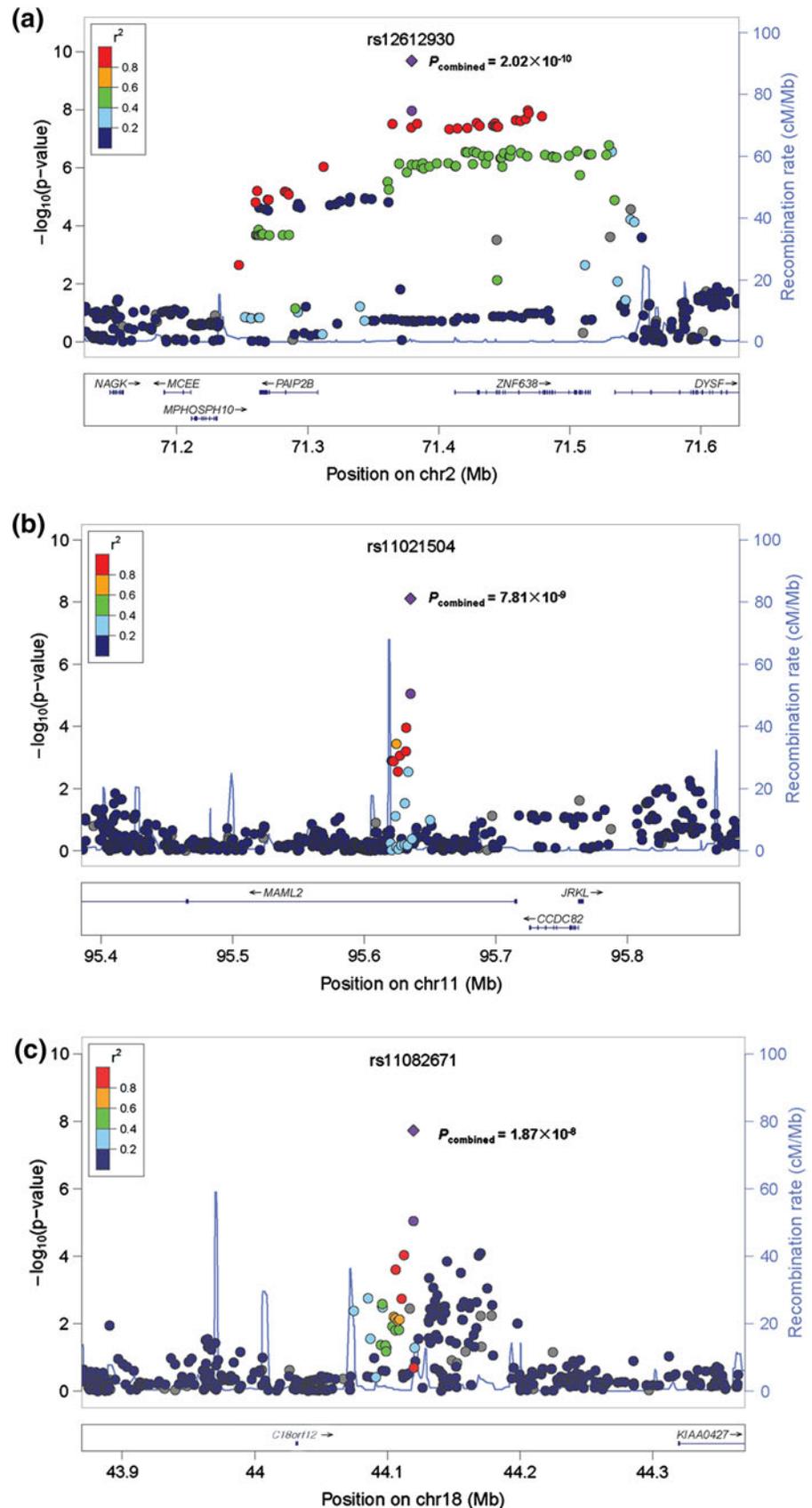
<sup>a</sup> Information for SNP ID and chromosomal position was based on NCBI genome build 36. The most significant SNP in each locus were listed

<sup>b</sup> Effect allele frequency (EAF) for each SNP was calculated based on discovery samples

<sup>c</sup> Effect size represented the difference in height Z-score associated with each additional allele and was estimated for effect allele

<sup>d</sup> P<sub>het</sub> represented the P value for between study heterogeneity

**Fig. 1** Regional plots of newly identified loci for height. The three newly identified loci were at 2p13.2 (a), 11q21 (b), and 18q21.1(c). Genotyped and imputed SNPs passing quality control measures in the discovery stage are plotted with their  $P$  values of the discovery meta-analysis. In the *top panel* of each, the  $-\log_{10}P$  values of the SNPs are presented according to their genomic positions, with the estimated recombination rates from HapMap Phase 2 CHB + JPT samples. The lead SNP is shown with a *purple circle* for the discovery meta-analysis and by a *purple diamond* for the combined analysis labeled by rsID and  $P_{\text{combined}}$  value. The  $r^2$  values of linkage disequilibrium between the lead SNP and the other SNPs are indicated by *different colors*. In the *bottom panels* the genes within the interested region are annotated and are shown as *arrows*. All chromosome positions are based on NCBI build 36



## Comparison with previous results in Asians

Table 2 summarized association results in our discovery study from previously reported height-related loci of Asians. We evaluated the associations of eight SNPs that achieved genome-wide significance in GWAS of other Asian populations. All these SNPs were in the same effect directions as previous studies, and four showed nominal significance ( $P < 0.05$ ) including rs3791675 (*EFEMP1*,  $P = 4.32 \times 10^{-4}$ ), rs7571816 (*DIS3L2*,  $P = 9.50 \times 10^{-4}$ ), rs7678436 (*NCAPG-LCORL*,  $P = 5.52 \times 10^{-4}$ ), and rs12338076 (*LHX3-QSOX2*,  $P = 3.42 \times 10^{-2}$ ) (Table 2). We also investigated whether the height-associated SNPs identified by non-Asian GWAS were associated in our sample. We found that 35 SNPs showed directionally consistent and nominally significant associations in the discovery study ( $P < 0.05$ ) (Supplementary Table 2).

## Cumulative effects of height-related loci

Weighted height scores, incorporating the five loci achieving genome-wide significance in our study, were calculated to examine the aggregate effect of associated loci on height in the discovery study. Average height increased linearly with increasing quintiles of weighted height scores (Fig. 2). Compared with those in bottom quintiles, the mean heights of individuals in top quintiles were 2.89 and 2.03 cm higher for males and females, respectively. In total, these five SNPs could explain 0.89 % of the variance for height in Han Chinese. When we included the previously reported SNPs that were associated with height (Supplementary Table 2) in our analysis, the proportion of variance explained was 3.67 %.

## Discussion

We carried out a meta-analysis of GWAS in 6,534 subjects with in silico replication in 1,881 subjects. We identified three novel loci reaching the genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ) including *ZNF638*, *MAML2*, and *C18orf12*. We also confirmed two loci previously reported in European populations including *CS* and *CYP19A1* (Lango Allen et al. 2010).

Of the novel discoveries, rs12612930 located in *ZNF638* showed the most significant association with height in our study. ZNF proteins, as one large class of transcription factors, have been shown to participate in a variety of processes in skeletal development (Ganss and Jheon 2004). A recent study demonstrated that *ZNF638* played a role in regulating adipocyte differentiation (Meruvu et al. 2011). The MAF of rs12612930 was quite different across

**Table 2** Association results of the discovery study in the Chinese for the previously reported SNPs associated with height in Asians

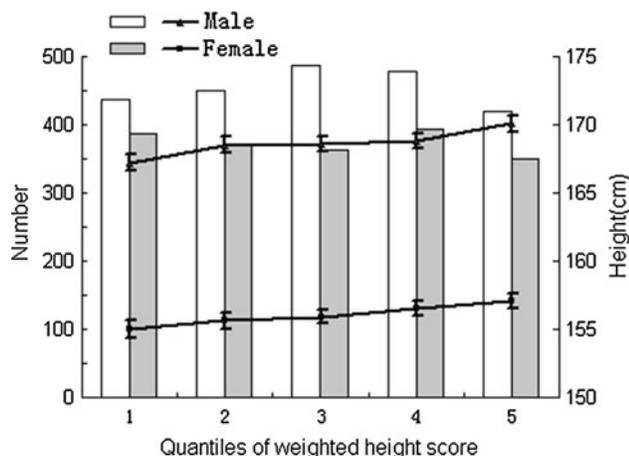
Locus	SNP <sup>a</sup>	CHR	BP <sup>a</sup>	Previous Asian GWAS			Our study					
				Effect/other allele	EAF	Effect size (SE)	P	Reported	Effect/other allele	EAF <sup>b</sup>	Effect size (SE) <sup>c</sup>	P
<i>EFEMP1</i>	rs3791675	2	55964813	G/C	0.22	0.445(0.096) <sup>d</sup>	1.7E–09	Cho et al. (2009), Okada et al. (2010)	G/C	0.22	0.074(0.021)	4.32E–04
<i>DIS3L2</i>	rs7571816	2	232785308	A/G	0.45	0.062(0.010) <sup>c</sup>	8.8E–09	Okada et al. (2010)	A/G	0.48	0.058(0.018)	9.50E–04
<i>ZBTB38</i>	rs10513137	3	142626120	A/G	0.26	0.492(0.091) <sup>d</sup>	5.6E–12	Cho et al. (2009), Okada et al. (2010)	A/G	0.23	0.030(0.021)	1.44E–01
<i>NCAPG-LCORL</i>	rs7678436	4	17407064	G/A	0.74	0.093(0.011) <sup>c</sup>	7.9E–15	Okada et al. (2010)	G/A	0.71	0.077(0.022)	5.52E–04
<i>HMGAI</i>	rs7742369	6	34273699	G/A	0.14	0.113(0.014) <sup>c</sup>	1.1E–13	Cho et al. (2009), Okada et al. (2010)	G/A	0.10	0.066(0.035)	6.32E–02
<i>PLAG1</i>	rs7833986	8	57262703	G/A	0.92	0.124(0.019) <sup>c</sup>	7.7E–10	Cho et al. (2009), Okada et al. (2010)	G/A	0.94	0.064(0.037)	8.00E–02
<i>LHX3-QSOX2</i>	rs12338076	9	138261561	C/A	0.34	0.063(0.011) <sup>c</sup>	2.2E–08	Okada et al. (2010)	C/A	0.32	0.047(0.022)	3.42E–02
<i>IGF1</i>	rs5742692	12	101323728	A/G	0.73	0.066(0.011) <sup>c</sup>	3.9E–08	Okada et al. (2010)	A/G	0.76	0.033(0.021)	1.06E–01

<sup>a</sup> Information for SNP ID and chromosomal position is based on NCBI genome build 36

<sup>b</sup> Effect allele frequency (EAF) for each SNP was calculated based on discovery samples

<sup>c</sup> Effect size represented the difference in height Z-score associated with each additional allele and was estimated for effect allele

<sup>d</sup> Effect size represented the difference in height (cm) associated with each additional allele and was estimated for effect allele



**Fig. 2** Quintiles of weighted height score comprised of five validated SNPs and height in Chinese males and females. Weighted height scores were calculated to examine the aggregate effect of associated loci on height in the discovery study. For each quintile, the average height was plotted separately for males and females with corresponding 95 % confidence interval (CI). The bars present the proportion of sample in each group

Chinese and European populations. The SNP rs12612930 had a MAF of 40 % in our study, but was monomorphic in European populations (HapMap Phase 2 CEU data). This difference in MAF may explain why this SNP has not been identified for height in European populations.

The SNP rs11021504 is located in the first intron of the *MAML2* gene. As a transcriptional coactivator for NOTCH proteins (Lin et al. 2002), *MAML2* strengthens NOTCH signals which control cell fate decisions in a variety of developmental processes including skeletal development (Conlon et al. 1995; Hilton et al. 2008; Mead and Yutzey 2009). *NOTCH* mutant mice showed complete embryonic lethality and abnormal somite development, as well as decreased embryo size (Conlon et al. 1995). Several other *in vivo* studies also demonstrated that NOTCH signaling was crucial for chondrocyte progenitor proliferation and for the normal osteogenesis of hypertrophic chondrocyte during skeletal development (Kohn et al. 2012; Mead and Yutzey 2009). As a positive transcriptional regulator of *NOTCH*, *MAML2* can be a potential candidate gene influencing human height and merits further exploration. The top SNP rs11082671 at the *C18orf12* locus is located 400 kb upstream of the *SMAD2* gene. *SMAD2* mediates TGF- $\beta$  signal and thus regulates cell proliferation, differentiation, and apoptosis, including osteoblast differentiation and bone formation (Chen et al. 2012).

We also confirmed two loci associated with height reported in European populations (Lango Allen et al. 2010). The SNP rs3816804 is located in the first intron of the *CS* gene and around 55 kb downstream of the *STAT2* gene. The most significant SNP in *CYP19A1* locus,

rs3751599, lies in the first intron of the *CYP19A1* gene. *CYP19A1* catalyzes the last steps of estrogen biosynthesis, and its polymorphisms have been shown to be associated with adult male height (Ellis et al. 2001). Comparison of our results with previous findings in other Asian populations (Cho et al. 2009; Okada et al. 2010) revealed that many loci associated with adult height were overlapped among populations. All eight SNPs were in the same directions of the effects as previous reports and four of them showed nominal significance. These results suggested that the Chinese population shared similar genetic backgrounds with other Asian populations for height.

We evaluated the cumulative effects of associated loci on height in our study, and these loci could explain 0.89 % of the total variance of height. The proportion of variance explained was elevated to 3.67 % when we additionally included previous reports SNPs that were associated with height in the in our study. We might have overestimated the proportion of variance explained due to the “winner’s curse effect”, as genetic effect size was usually overestimated in the discovery studies (Xiao and Boehnke 2009). As a common variant-base approach, GWAS could identify more loci when sample size increases (Weedon and Frayling 2008). However, even the GIANT study with more than 130,000 subjects, which identified 180 loci, could only explain  $\sim 10$  % of the variance in height (Lango Allen et al. 2010). It remains low when considering the high heritability of height ( $\sim 80$  %). This suggests that there are still more height-related genetic variants that need to be discovered. Common variants of low effect sizes, lower frequency variants, and genomic structural variations may also help to account for more of the missing heritability (Manolio et al. 2009). Fitting all SNPs could explain much more proportion of variance for human height (Yang et al. 2011). This method helps to find the missing heritability due to discarding SNPs with small effect size which could not pass stringent significance test.

Several strengths of this study should be noted. First, our study subjects were restricted to Han Chinese, and homogeneity of ethnicity lowered the chance of population stratification. Second, inflation factor in our study was moderate, and PC analysis was employed to control the effect of potential population stratification. Limitation of our study should also be carefully considered. The sample size of our study was moderate and had limited power to detect variants of low effect size. Though there was little evidence of heterogeneity for loci, random model may help to shrink the effect size and make the results more conservative.

In conclusion, we identified three new loci associated with height (*ZNF638*, *MAML2*, and *C18orf12*) and replicated two previously reported loci (*CS* and *CYP19A1*) in Chinese. Our data suggest that both shared and unique genetic backgrounds of human height are present in

different ancestry groups. Further studies are required to confirm the associations of these loci identified in this study across different populations and elucidate the underlying molecular mechanisms of growth in the future.

**Acknowledgments** This work is supported by the National Basic Research Program of China (973 Plan) (2011CB503901) from the Ministry of Science and Technology of China, the National Science Foundation of China (30930047, 81000040) and the High-Tech Research and Development Program of China (863 Plan) (2009AA022703, 2012AA02A516). The Genetic Epidemiology Network of Salt Sensitivity (GenSalt) is supported by a cooperative agreement project grant (HL072507, HL087263 and HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA. We acknowledge James E Hixson from University of Texas and Dabeeru C. Rao from Washington University for their contribution in GenSalt study. We thank the GIANT Consortium for providing online height association results.

**Conflict of interest** None declared.

## References

- Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, Kooperberg C (2012) Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). *Hum Mol Genet* 21:711–720
- Chen MH, Yang Q (2010) GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 26:580–581
- Chen G, Deng C, Li YP (2012) TGF-beta and BMP signaling in osteoblast differentiation and bone formation. *Int J Biol Sci* 8:272–288
- Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban HJ, Yoon D, Lee MH, Kim DJ, Park M, Cha SH, Kim JW, Han BG, Min H, Ahn Y, Park MS, Han HR, Jang HY, Cho EY, Lee JE, Cho NH, Shin C, Park T, Park JW, Lee JK, Cardon L, Clarke G, McCarthy MI, Lee JY, Oh B, Kim HL (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 41:527–534
- Conlon RA, Reaume AG, Rossant J (1995) Notch1 is required for the coordinate segmentation of somites. *Development* 121:1533–1545
- Croteau-Chonka DC, Marvelle AF, Lange EM, Lee NR, Adair LS, Lange LA, Mohlke KL (2011) Genome-wide association study of anthropometric traits and evidence of interactions with age and study year in Filipino women. *Obesity (Silver Spring)* 19:1019–1027
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997–1004
- Ellis JA, Stebbing M, Harrap SB (2001) Significant population variation in adult male height associated with the Y chromosome and the aromatase gene. *J Clin Endocrinol Metab* 86:4147–4150
- Estrada K, Krawczak M, Schreiber S, van Duijn K, Stolk L, van Meurs JB, Liu F, Penninx BW, Smit JH, Vogelzangs N, Hottenga JJ, Willemsen G, de Geus EJ, Lorentzon M, von Eller-Eberstein H, Lips P, Schoor N, Pop V, de Keijzer J, Hofman A, Aulchenko YS, Oostra BA, Ohlsson C, Boomsma DI, Uitterlinden AG, van Duijn CM, Rivadeneira F, Kayser M (2009) A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum Mol Genet* 18:3516–3524
- Ganss B, Jheon A (2004) Zinc finger transcription factors in skeletal development. *Crit Rev Oral Biol Med* 15:282–297
- GenSalt Collaborative Research Group (2007) GenSalt: rationale, design, methods and baseline characteristics of study participants. *J Hum Hypertens* 21:639–646
- He J, Neal B, Gu D, Suriyawongpaisal P, Xin X, Reynolds R, MacMahon S, Whelton PK (2004) International collaborative study of cardiovascular disease in Asia: design, rationale, and preliminary results. *Ethn Dis* 14:260–268
- Hilton MJ, Tu X, Wu X, Bai S, Zhao H, Kobayashi T, Kronenberg HM, Teitelbaum SL, Ross FP, Kopan R, Long F (2008) Notch signaling maintains bone marrow mesenchymal progenitors by suppressing osteoblast differentiation. *Nat Med* 14:306–314
- Hou L, Chen S, Yu H, Lu X, Chen J, Wang L, Huang J, Fan Z, Gu D (2009) Associations of PLA2G7 gene polymorphisms with plasma lipoprotein-associated phospholipase A2 activity and coronary heart disease in a Chinese Han population: the Beijing atherosclerosis study. *Hum Genet* 125:11–20
- Johansson A, Marroni F, Hayward C, Franklin CS, Kirichenko AV, Jonasson I, Hicks AA, Vitart V, Isaacs A, Axenovich T, Campbell S, Dunlop MG, Floyd J, Hastie N, Hofman A, Knott S, Kolcic I, Pichler I, Polasek O, Rivadeneira F, Tenesa A, Uitterlinden AG, Wild SH, Zorkoltseva IV, Meitinger T, Wilson JF, Rudan I, Campbell H, Pattaro C, Pramstaller P, Oostra BA, Wright AF, van Duijn CM, Aulchenko YS, Gyllenstein U (2009) Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum Mol Genet* 18:373–380
- Kim J-J, Lee H-I, Park T, Kim K, Lee J-E, Cho NH, Shin C, Cho YS, Lee J-Y, Han B-G, Yoo H-W, Lee J-K (2009) Identification of 15 loci influencing height in a Korean population. *J Hum Genet* 55:27–31
- Kim HN, Lee EJ, Jung SC, Lee JY, Chung HW, Kim HL (2010) Genetic variants that affect length/height in infancy/early childhood in Vietnamese-Korean families. *J Hum Genet* 55:681–690
- Kohn A, Dong Y, Mirando AJ, Jesse AM, Honjo T, Zuscik MJ, O'Keefe RJ, Hilton MJ (2012) Cartilage-specific RBPjkappa-dependent and -independent Notch signals regulate cartilage and bone development. *Development* 139:1198–1212
- Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E et al (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832–838
- Lei SF, Yang TL, Tan LJ, Chen XD, Guo Y, Guo YF, Zhang L, Liu XG, Yan H, Pan F, Zhang ZX, Peng YM, Zhou Q, He LN, Zhu XZ, Cheng J, Liu YZ, Papasian CJ, Deng HW (2009) Genome-

- wide association scan for stature in Chinese: evidence for ethnic specific loci. *Hum Genet* 125:1–9
- Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C, Illig T, Hackett R, Heid IM, Jacobs KB, Lyssenko V, Uda M, Boehnke M, Chanock SJ, Groop LC, Hu FB, Isomaa B, Kraft P, Peltonen L, Salomaa V, Schlessinger D, Hunter DJ, Hayes RB, Abecasis GR, Wichmann HE, Mohlke KL, Hirschhorn JN (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 40:584–591
- Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. *Annu Rev Genomics Hum Genet* 10:387–406
- Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 34:816–834
- Lin SE, Oyama T, Nagase T, Harigaya K, Kitagawa M (2002) Identification of new human mastermind proteins defines a family that consists of positive regulators for notch signaling. *J Biol Chem* 277:50612–50620
- Liu JZ, Medland SE, Wright MJ, Henders AK, Heath AC, Madden PA, Duncan A, Montgomery GW, Martin NG, McRae AF (2010) Genome-wide association study of height and body mass index in Australian twin families. *Twin Res Hum Genet* 13:179–193
- Lu X, Wang L, Chen S, He L, Yang X, Shi Y, Cheng J, Zhang L, Gu CC, Huang J, Wu T, Ma Y, Li J, Cao J, Chen J, Ge D, Fan Z, Li Y, Zhao L, Li H, Zhou X, Chen L, Liu D, Duan X, Hao Y, Lu F, Liu Z, Yao C, Shen C, Pu X, Yu L, Fang X, Xu L, Mu J, Wu X, Zheng R, Wu N, Zhao Q, Liu X, Wang M, Yu D, Hu D, Ji X, Guo D, Sun D, Wang Q, Yang Y, Liu F, Mao Q, Liang X, Ji J, Chen P, Mo X, Li D, Chai G, Tang Y, Li X, Du Z, Dou C, Yang Z, Meng Q, Wang D, Wang R, Yang J, Schunkert H, Samani NJ, Kathiresan S, Reilly MP, Erdmann J, Peng X, Chen R, Qiang B, Gu D (2012) Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet* 44:890–894
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttman AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Mead TJ, Yutzev KE (2009) Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development. *Proc Natl Acad Sci USA* 106:14420–14425
- Meruvu S, Hugendubler L, Mueller E (2011) Regulation of adipocyte differentiation by the zinc finger protein ZNF638. *J Biol Chem* 286:26516–26523
- Okada Y, Kamatani Y, Takahashi A, Matsuda K, Hosono N, Ohmiya H, Daigo Y, Yamamoto K, Kubo M, Nakamura Y, Kamatani N (2010) A genome-wide association study in 19 633 Japanese subjects identified LHX3-QSOX2 and IGF1 as adult height loci. *Hum Mol Genet* 19:2303–2312
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, Shen H, Timpson N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L, Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN, Watanabe RM, Uda M, Tuomilehto J, Coresh J, Hirschhorn JN, Shuldiner AR, Schlessinger D, Collins FS, Davey Smith G, Boerwinkle E, Cao A, Boehnke M, Abecasis GR, Mohlke KL (2008) Common variants in the GDF5-UQCC region are associated with variation in human height. *Nat Genet* 40:198–203
- Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, Davis C, Dunkel L, De Lange M, Harris JR, Hjelmberg JV, Luciano M, Martin NG, Mortensen J, Nistico L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemssen G, Kaprio J (2003) Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 6:399–408
- Tonjes A, Koriath M, Schleinitz D, Dietrich K, Bottcher Y, Rayner NW, Almgren P, Enigk B, Richter O, Rohm S, Fischer-Rosinsky A, Pfeiffer A, Hoffmann K, Krohn K, Aust G, Spranger J, Groop L, Blüher M, Kovacs P, Stumvoll M (2009) Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Hum Mol Genet* 18:4662–4668
- Weedon MN, Frayling TM (2008) Reaching new heights: insights into the genetics of human stature. *Trends Genet* 24:595–603
- Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR, Elliott KS, Hackett R, Guiducci C, Shields B, Zeggini E, Lango H, Lyssenko V, Timpson NJ, Burt NP, Rayner NW, Saxena R, Ardlie K, Tobias JH, Ness AR, Ring SM, Palmer CN, Morris AD, Peltonen L, Salomaa V, Davey Smith G, Groop LC, Hattersley AT, McCarthy MI, Hirschhorn JN, Frayling TM (2007) A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* 39:1245–1250
- Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM (2008) Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40:575–583
- Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191
- Xiao R, Boehnke M (2009) Quantifying and correcting for the winner's curse in genetic association studies. *Genet Epidemiol* 33:453–462
- Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43:519–525