npg

ORIGINAL ARTICLE

The regulation-of-autophagy pathway may influence Chinese stature variation: evidence from elder adults

Feng Pan[1,5], Xiao-Gang Liu[1,5], Yan-Fang Guo[1], Yuan Chen[1], Shan-Shan Dong[1,2], Chuan Qiu[1], Zhi-Xin Zhang[1], Qi Zhou[1], Tie-Lin Yang[1,2], Yan Guo[1,2], Xue-Zhen Zhu[1], Hong-Wen Deng[1,2,3,4]

Recent success of genome-wide association studies (GWASs) on human height variation emphasized the effects of individual loci or genes. In this study, we used a developed pathway-based approach to further test biological pathways for potential association with stature, by examining ~370 000 single-nucleotide polymorphisms (SNPs) across the human genome in 618 unrelated elder Han Chinese. A total of 626 biological pathways annotated by any of the three major public pathway databases (KEGG, BioCarta and Ambion GeneAssist Pathway Atlas) were tested. The regulation-of-autophagy (ROA) (nominal $P=0.012$) pathway was marginally significantly associated with human stature after our family wise error rate multiple-testing correction. We also used 1000 random recruited US whites for further replication. Interestingly, the ROA pathway presented the strongest signals in whites for height variation (nominal $P=0.002$). The results correspond to biological roles of the ROA pathway in human long bone development and growth. Our findings also implied that multiple-genetic factors may work jointly as a functional unit (pathway), and the traditional GWASs could have missed important genetic information imbedded in those less significant markers.

INTRODUCTION

Human stature has a long research history in the community, as it is the most visible traits with remarkable high heritability and can be easily and accurately measured.[1] Specifically in Han Chinese, heritability of stature is >0.6.[2] The major driver of those studies on stature is to provide clues for human growth, bone development and onset of some diseases such as cancer, type 2 diabetes, H syndrome and osteoporosis.[1,3,4]

Recent success in genome-wide association studies (GWASs) has revealed that some pathways can influence the human height.[5–10] Strong statistical and experimental evidence showed that pathways underlying bone and cartilage development have critical roles in the formation of adult height,[1] such as pathways of Hedgehog signaling,[6] bone morphogenic[9,11] and extracellular matrix.[1] Meanwhile, some pathways may have indirect impact on human height and highlight more fundamental biological processes, including the pathways of chromatin structure regulation,[6] cell-cycle regulation[1] and the C-type natriuretic peptide signaling.[12]

Our earlier GWAS performed in Han Chinese[13] pinpointed some single-nucleotide polymorphisms (SNPs) that was influencing Chinese height variation. Nevertheless, the most significant SNPs can only explain limited variance (~0.3%) in stature.[13] The pathway-based association test integrates current knowledge about functional interactions among genes and represents a promising approach to look into genome-wide data to find the potential joint effects or functional interactions of genes, which were imbedded in the remaining unreported SNPs that did not achieve the genome-wide significant level.[14]

In this study, the genome-wide pathway-based analysis was conducted to detect the potential association between 626 pathways and height variation in Han Chinese. We aimed to identify pathways that can be influencing Chinese stature and/or to provide novel hypotheses in the mechanisms of human growth. This study may provide new insights into height variation, especially for the Chinese ethnic, which was less studied and genetically differs from the whites.[1,9,13]

MATERIALS AND METHODS

Subjects

The study was approved by the Research Administration of Xi'an Jiaotong University to assure that the general ethnic rules were addressed. All study volunteers have signed the informed-consent forms before they entered the

[1]The Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China; [2]Departments of Orthopedic Surgery and Basic Medical Science, School of Medicine, University of Missouri-Kansas City, Kansas City, MO, USA; [3]College of Life Sciences and Engineering, Beijing Jiao Tong University, Beijing, PR China and [4]Center of Systematic Biomedical Research, Shanghai University of Science and Technology, Shanghai, PR China
[5]These authors contributed equally to this work.
Correspondence: Dr H-W Deng, Departments of Orthopedic Surgery and Basic Medical Sciences, University of Missouri-Kansas City, 2411 Holmes Street, Room M3-C03, Kansas City, MO 64108, USA.
E-mail: dengh@umkc.edu

study. A total of 618 elder Han adults were randomly recruited from the Xi'an City and its surrounding areas in northern China. Basic characteristics of the study sample are given in Table 1a. Stature for each subject was measured using a stadiometer wearing no shoes. The daily calibrated stadiometer was in well maintenance and the long-term precision was monitored. The coefficient of variation of stature measurements was 5.8%. Strict data quality control was conducted by excluding the outliers and database inputting errors. The whole data set was double crosschecked by at least two different researchers.

All the individuals with chronic diseases and conditions that might potentially affect hormone level, bone structure or metabolism were excluded from this study, which included but not limited to chronic disorders involving vital organs, diabetes, hyperthyroidism, hypo-/hyper-parathyroidism and other serious skeletal diseases (Paget's disease, osteogenesis imperfecta, rheumatoid arthritis, kyphosis, serious vertebral fracture), chronic use of drugs affecting bone metabolism (hormone replacement therapy, corticosteroid therapy, anticonvulsant drugs, anti-bone-resorptive or bone anabolic agents) and malnutrition conditions (such as anorexia nervosa).

Genotyping

Genomic DNA was extracted from whole human blood using a commercial isolation kit (Gentra systems, Minneapolis, MN, USA) following the standard protocols. For each sample, genotyping with GeneChip Human Mapping 500 K set containing 250 K Nsp array and 250 K Sty array (Affymetrix, Santa Clara, CA, USA) was performed using the standard protocols recommended by the manufacturer.

From the initial full-set of 500 568 SNPs, we discarded 18 067 SNPs with sample call rate <90%, 13 041 SNPs with allele frequencies extremely deviating from Hardy–Weinberg equilibrium ($P < 10^{-7}$) and 98 202 SNPs with minor allele frequencies <1%. The final SNP number for the analyses is 371 258. This SNP set covered 14 642 genes and yielded an average marker spacing of ~7.9 kb throughout the human genome.

Population stratification

We used STRUCTURE 2.2 program (http://pritch.bsd.uchicago.edu/software.html) to detect the potential population stratification of our sample. The program used a Markov chain Monte Carlo algorithm to cluster individuals into cryptic subpopulations on the basis of multi-loci genotype data.[15] We performed the analysis assuming the number of population strata $k=2$ and using a set of 1000 unlinked SNPs randomly selected genome wide. EIGENSTRAT[16] was used to crosscheck the stratification in our analyses.

Generation of pathway/gene sets collection

To identify the pathways potentially contributing to Chinese stature variation, we comprehensively retrieved the pathways collected by three major pathway databases, KEGG (http://www.genome.ad.jp/kegg/pathway.html), BioCarta (http://www.biocarta.com/genes/index.asp) and Ambion GeneAssist Pathway Atlas (http://www.ambion.com/tools/pathway). The number of pathways finally retrieved from BioCarta, KEGG and Ambion were 364, 262 and 201, respectively. To relieve the multiple-testing issue and avoid testing over narrow or broad functional pathways, we only included those pathways with 10–200 genes and at least 85% of which were covered by at least one SNP of our qualified genotypic data. The pathways with the following characteristics were filtered: (1) the pathway contains <10 or >200 genes in the above three database; (2) after discarding those unqualified SNPs (bad sample call rate, no Hardy–Weinberg equilibrium, small minor allele frequencies), the pathway contains <10 or >200 genes in our genotypic data; (3) after filtering those SNPs that were 500 kb away from the involved gene, the pathway contains <10 or >200 genes in our cleaned data and (4) the pathway contains genes within 10–200 gene, but <85% of them were included in our analysis due to discarded SNPs or SNPs with >500 kb distance. Overall, a total of 626 pathways/functional networks were finally analyzed.

Pathway-based GWAS analysis

Gender and age are significant fixed effectors on stature[13] and they were used to adjust the raw stature data. The GWAS analysis for individual SNPs was achieved by the Wald *t*-test implemented in Plink (version 1.03).[17]

Table 1a Basic characteristics of the 618 studied Chinese (mean ± s.d.)

Trait	Total (n=618)	Female (n=351)	Male (n=267)
Age (years)	70.6 ± 7.8	70.2 ± 8.2	71.1 ± 7.1
Stature (cm)	160.9 ± 9.0	155.5 ± 6.2	168.0 ± 6.9

Table 1b Basic characteristics of the 1000 studied American whites (mean ± s.d.)

Trait	Total (n=1000)	Female (n=501)	Male (n=499)
Age (years)	50.3 ± 18.3	50.1 ± 7.7	50.5 ± 8.9
Stature (cm)	170.8 ± 9.8	163.8 ± 6.5	177.8 ± 7.0

The used pathway-based GWAS approach was proposed and developed first by Wang *et al*.[14]

SNPs that are 500 kb away from any gene are not considered. The SNPs of our identified pathway were verified without overlapping or misrepresentation problem (Figure 2). Among all the SNPs of a given gene, the SNP achieving the smallest *P*-value (the highest Wald *t*-test statistic values) in the single-marker association tests was assigned to represent the magnitude of association evidence of the gene. If in very rare cases where one SNP is located within shared regions of two overlapping genes, we map the SNP to both genes. They will all get the same score. We ranked all genes by sorting their statistic values from the largest to smallest. Then, a weighted Kolmogorov–Smirnov-like running-sum enrichment score (ES) was calculated.[14] Briefly, ES score was used to reflect the enrichment of a given gene set at the top of the entire genome-wide ranked gene list. It measures the deviation from zero to maximum level of the highest statistic value encountered in the random walk. High ES score means this set of genes (genes of pathway) concentrated at the top of ranked gene list.

The nominal significance level of ES for each pathway was estimated by shuffling phenotypic data. The procedures of ES calculation were repeated for 1000 times. Finally, an empirical distribution of ES (ES_{null}) was generated for each pathway. The nominal significance of an observed ES for a pathway (nominal *P*-value) was estimated as the percentage of permutations whose $ES_{null} >$ observed ES.

Finally, two measures were used to adjust for multiple-hypothesis testing for more reliable results. We first normalized the ES to correct the influence of gene size, so that different gene sets are directly comparable to each other. The false-discovery rate (FDR) *Q* value (denoted as Q_{FDR}) and family wise error rate (FWER) *P*-value (denoted as P_{FWER}) were generated for correcting the multiple testing.[14] FDR is a procedure that is usually used to control the fraction of expected false-positive findings to stay below a certain threshold. FWER is another approach that is considered to be highly conservative seeking to ensure that the list of reported results does not include even a single false-positive gene set.

To identify the common pathway(s) that may influence height in distinct ethnics, and also as a replication effort, we compared the results of this study with those of the pathway-based GWAS analysis for height variation in our American white sample study, which was approved by the University of Missouri-Kansas City Institutional Review Board. A total of 1000 whites were randomly recruited from Midwestern of United States (Table 1b). The same genotyping and statistical approaches were applied across the Chinese and the American studies.[5]

RESULTS

Characteristics of study subjects

A total of 618 individuals were included in the analysis. Basic characteristics of the study sample are shown in Table 1a.

The STRUCTURE program results showed that all the subjects were clustered together as a single group by using 1000 unlinked SNPs, which suggested that no significant population stratification was present in our sample. The analyses by EIGENSTRAT confirmed that only one principal component was significant ($P < 0.001$), which suggested that our study sample originated from one common ancestry population. Consequently, the inflation factor λ calculated by the Genomic Control approach[18] implemented in EIGENSTRAT is 1.02, close to 1 standing for no stratification.

Pathway-based association analysis

Out of the 626 pathways, the regulation-of-autophagy (ROA) pathway achieved the strongest evidence of association with height variation (nominal $P = 0.012$, $Q_{FDR} = 0.30$ and $P_{FWER} = 0.11$). The ROA pathway was involved in 35 genes (data from KEGG database), with 28 genes covered by our SNP set (Table 2). Among the 28 genes, 5 genes (PIK3C3, ULK2, ATG5, GABARAPL1 and IFNG) got nominal $P < 0.01$, and 14 genes were ranked among the top 5000 (leading edge) genes in the single-SNP GWAS. The ROA pathway achieved a high ES score of 0.47 with nominal P-value = 0.012 (Figure 1). The individual gene with the most significant P-value among all the 28 genes is PIK3C3 (nominal $P = 0.0009$). Other ROA genes adding to the ES score (that is the genes that ranked before or at the point of maximize ES achieved, also denoted as leading-edge genes) are listed in Table 2 and Figure 1. Information of the representative SNPs for the

17 leading-edge genes that contributed positively to the ES is given in Table 3.

Comparison analysis

In the American whites, the ROA pathway was the only one that consistently presented the strongest association signals with height variation (nominal $P = 0.002$, ES score = 0.55, $Q_{FDR} = 0.042$, $P_{FWER} = 0.016$). The individual gene with the most significant nominal P-value is IFNA4 (nominal $P = 0.0026$). Seven genes of the ROA pathway (IFNA4, IFNA21, PIK3C3, ULK2, IFNA14, IFNA13 and IFNG) closely associated with height in our GWAS with nominal P-value < 0.01. The genes PIK3C3, ULK2 and IFNG presented significant nominal P-value < 0.01 in both Han Chinese and American whites. The results from United States did not indicate other novel pathways or replicate the prior known ones regulating human height (all other tested pathways had $Q_{FDR} > 0.2$ and $P_{FWER} > 0.13$).

We also compared the P-values of the same representative SNPs in Chinese with in US whites, and none of them were significant (except for gene PRKAA2) (Table 4). Thus, different sample may have different representative SNPs even for the same gene, because of the different minor allele frequency and the distinct ethnic background.

DISCUSSION

The pathway-based approach can work as a complementary strategy to traditional SNP-based analyses and could be helpful for the

Table 2 Genes in the ROA pathway

Gene symbol	Gene ID	Cytoband	Full name
ATG12[a]	9140	5q21–q22	ATG12 autophagy related 12 homolog (*S. cerevisiae*)
ATG3	64422	3q13.2	ATG3 autophagy related 3 homolog (*S. cerevisiae*)
ATG4B	23192	2q37.3	ATG4 autophagy related 4 homolog B (*S. cerevisiae*)
ATG4C[a]	84938	1p31.3	ATG4 autophagy related 4 homolog C (*S. cerevisiae*)
ATG4D	84971	19p13.2	ATG4 autophagy related 4 homolog D (*S. cerevisiae*)
ATG5[a]	9474	6q21	ATG5 autophagy related 5 homolog (*S. cerevisiae*)
ATG7[a]	10533	3p25.3–p25.2	ATG7 autophagy related 7 homolog (*S. cerevisiae*)
BECN1	8678	17q21	Beclin 1, autophagy related. (also know as ATG6)
GABARAPL1[a]	23710	12p13.2	GABA(A) receptor-associated protein-like 1
GABARAPL2	11345	16q22.3–q24	GABA(A) receptor-associated protein-like 2
IFNA1	3439	9p22	Interferon, α1
IFNA10[a]	3446	9p22	Interferon, α10
IFNA13[a]	3447	9p22	Interferon, α13
IFNA14[a]	3448	9p22	Interferon, α14
IFNA16[a]	3449	9p22	Interferon, α16
IFNA17	3451	9p22	Interferon, α17
IFNA2	3440	9p22	Interferon, α2
IFNA21[a]	3452	9p22	Interferon, α21
IFNA4[a]	3441	9P22	Interferon, α4
IFNA5[a]	3442	9p22	Interferon, α5
IFNA7[a]	3444	9p22	Interferon, α7
IFNA8	3445	9p22	Interferon, α8
IFNG[a]	3458	12q14	Interferon, γ
PIK3C3[a]	5289	18q12.3	Phosphoinositide-3-kinase, class3 (Vps34)
PIK3R4	30849	3q22.1	Phosphoinositide-3-kinase, regulatory subunit 4
PRKAA1	5562	5p12	AMP-activated protein kinase α1 catalytic subunit
PRKAA2[a]	5563	1p31	AMP-activated protein kinase α2 catalytic subunit
ULK2[a]	9706	17p11.2	unc-51-like kinase

Abbreviations: ES, enrichment score; ROA, regulation-of-autophagy.
[a]Genes contributing positively to the ES score (leading-edge genes).
The first 10 genes encode critical proteins actively involved in the autophagy process.
The IFNA family and IFNG encode pleiotropic cytokines (type I and type II interferons), which regulate the autophagy process.
The Gene ID was retrieved from NCBI GenBank (http://www.ncbi.nlm.nih.gov/Genbank/).

detection of novel mechanisms underlying complex diseases/phenotypes. In this study, this approach was used to analyze our genomic genotyping data for height variation in an elder Han Chinese population.[13] The ROA pathway was suggested at the genome-wide level pathway-based approach after the multiple-testing correction. The result implied that autophagy-related biological processes may
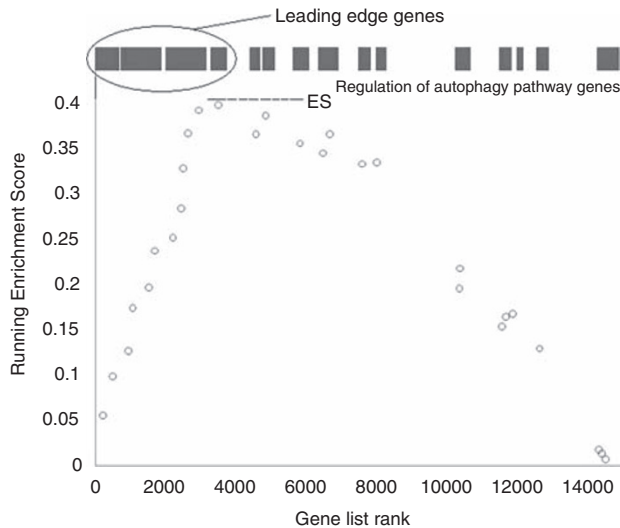

Figure 1 Running-sum plot for the ROA pathway, including the location of the maximum enrichment score (ES) and the leading-edge subset. The *x* axis is the rank of the 28 genes in ROA pathway in the whole gene list generated. These 28 genes are, from left to right, PIK3C3, ULK2, ATG5, GABARAPL1, IFNG, ATG7, PRKAA2, IFNA10, IFNA7, IFNA14, IFNA21, ATG4C, IFNA4, IFNA5, ATG12, IFNA13, IFNA16, ATG3, IFNA2, ATG4D, IFNA8, ATG4B, PIK3R4, IFNA1, IFNA17, PRKAA1, GABARAPL2 and BECN1. The *y* axis represents the running-enrichment score. The bar plot at the top of the figure also shows the distribution of the 28 genes from the ROA pathway at genome-wide level.

have important roles in human stature variation, and thus may participate in the development and growth disorders of bones.

The ROA pathway (Pathway ID hsa04140 in KEGG database) is the unique pathway related to autophagy among all the pathways tested. Autophagy is a conservative house-keeping catabolic process that involves degradation of a cell's own components by formation and separation of a membrane around the targeted region of the cell. The resultant vesicle then fuses with a lysosome and then the inner contents degrade. The autophagy process was generally induced and enhanced by nutrient starvation, defense of infection or cell programmed death.

Involvement of the ROA pathway in human bone and cartilage growth has been demonstrated in earlier studies. Long bone growth and consequent increase in height take place through the activities of chondrocytes embedded in the epiphyseal growth plate. The embedded chondrocytes of long bones undergo autophagy and express autophagic proteins before the induction of osteogenesis.[19] This is because autophagy enhances the survival of chondrocytes in the hypoxic microenvironments (the cartilage is avascular, thus contains little nutrition and oxygen) and induces the synthesis of the calcified extracellular matrix.[20–22] In addition, the crosstalk of bone growth factors such as interleukin-3 with the ROA pathway regulates the utilization of exogenous nutrients to maintain cell viability.[23] As a result, autophagy is an important intermediate stage during the terminal differentiation of chondrocytes and long bone development. Notably, the interferon-α family genes in the ROA pathway have some other effects on bone remodeling (Table 2). They can broadly suppress the proliferation of osteoprogenitor cells and inhibit the differentiation of monocytes to osteoclasts.[24,25]

Our study for the first time raises the importance of the ROA pathway to Chinese stature variation. The pathway was also highly suggested by our pathway-based genome-wide analysis that could influence the Caucasian height variation. In addition to highlighting its potential contribution to human height variation, the ROA pathway was demonstrated that associated with long bone mineral density at the ultra distal skeletal site (our unpublished data). Our findings in the bone mineral

Table 3 The representative SNPs mapped to the leading edge genes of the ROA pathway in Chinese and US whites

Gene	SNPs in Chinese	Allele	Location	MAF	Nominal P-value (Chinese)	SNPs in US whites	Allele	Location	MAF	Nominal P-value (US Whites)
PIK3C3	rs2587613	C/T	5′-Upstream	0.12	0.0009	rs16975296	A/G	5′-Upstream	0.07	0.0041
ULK2	rs205096	C/T	Intronic	0.30	0.0024	rs17794370	A/G	Intronic	0.27	0.0046
ATG5	rs3827644	C/G	Intronic	0.05	0.0053	rs553307	A/G	5′-Upstream	0.08	0.0400
GABARAPL1	rs11829144	C/T	Intronic	0.06	0.0063	rs10734824	A/G	3′-Downstream	0.02	0.3300
IFNG	rs2906852	C/T	5′-Upstream	0.16	0.0097	rs1904559	C/T	3′-Downstream	0.12	0.0087
ATG7	rs6442260	A/G	Intronic	0.29	0.0112	rs346081	C/G	Intronic	0.17	0.0200
PRKAA2	rs2796512	A/G	Intronic	0.46	0.0162	rs10889008	A/G	Intronic	0.47	0.0400
IFNA10	rs7849862	A/G	5′-Upstream	0.40	0.0186	rs7849862	A/G	3′-Downstream	0.17	0.0100
IFNA7	rs7856345	A/T	3′-Downstream	0.41	0.0194	rs7849623	A/T	3′-Downstream	0.18	0.0100
IFNA14	rs10811497	A/T	3′-Downstream	0.41	0.0209	rs7037147	A/G	Intronic	0.16	0.0061
IFNA21	rs7047811	A/T	5′-Upstream	0.40	0.0246	rs10964882	A/G	5′-Upstream	0.17	0.0034
ATG4C	rs1981067	C/T	Intronic	0.30	0.0324	rs11811517	A/T	Intronic	0.23	0.0100
IFNA4	rs7852323	C/T	3′-Downstream	0.41	0.0512	rs7852323	A/T	3′-Downstream	0.17	0.0026
IFNA5	rs10964938	G/T	3′-Downstream	0.34	0.0563	rs10964938	A/C	3′-Downstream	0.18	0.0200
ATG12	rs6883910	G/T	5′-Upstream	0.32	0.0800	rs6883910	A/C	5′-Upstream	0.39	0.3100
IFNA13	rs637949	C/G	5′-Upstream	0.27	0.0985	rs637949	C/G	5′-Upstream	0.17	0.0064
IFNA16	rs10811525	A/G	5′-Upstream	0.22	0.1043	rs10811525	A/C	3′-Downstream	0.41	0.7200

Abbreviations: MAF, minor allele frequency; ROA, regulation-of-autophagy; SNP, single-nucleotide polymorphism.

Table 4 The comparison of *P*-values for same representative SNPs in Chinese and US whites

Gene name	SNPs in Chinese	Allele	Physical position	Location	MAF (Chinese)	MAF (whites)	Nominal P-value (Chinese)	Nominal P-value (whites)
PIK3C3	rs2587613	C/T	chr18: 39164485	5′-Upstream	0.12	0.32	0.0009	0.1052
ULK2	rs205096	C/T	chr17: 19682327	Intronic	0.30	0.05	0.0024	0.5146
ATG5	rs3827644	C/G	chr6: 106685195	Intronic	0.05	0.23	0.0053	0.1797
GABARAPL1	rs11829144	C/T	chr12: 10350033	Intronic	0.08	0.06	0.0063	0.9644
IFNG	rs2906852	C/T	chr12:68379481	5′-Upstream	0.16	0.24	0.0097	0.3415
ATG7	rs6442260	A/G	chr3: 11590751	Intronic	0.29	0.35	0.0112	0.4654
PRKAA2	rs2796512	A/G	chr1: 57149497	Intronic	0.47	0.44	0.0162	0.0220
IFNA10	rs7849862	A/G	**chr9: 21191684**	5′-Upstream	0.40	0.16	0.0186	0.2923
IFNA7	rs7856345	A/T	**chr9: 21196490**	3′-Downstream	0.41	0.18	0.0194	0.3437
IFNA14	rs10811497	A/T	**chr9: 21193432**	3′-Downstream	0.41	0.18	0.0209	0.4041
IFNA21	rs7047811	A/T	**chr9: 21181454**	5′-Upstream	0.40	0.18	0.0246	0.3507
ATG4C	rs1981067	C/T	chr1: 63322743	Intronic	0.30	0.46	0.0324	0.1463
IFNA4	rs7852323	C/T	**chr9: 21174643**	3′-Downstream	0.41	0.17	0.0512	0.1746
IFNA5	rs10964938	G/T	**chr9: 21258474**	3′-Downstream	0.34	0.17	0.0563	0.2559
ATG12	rs6883910	G/T	chr5:115296082	5′-Upstream	0.32	0.36	0.0800	0.3616
IFNA13	rs637949	C/G	**chr9: 21369315**	5′-Upstream	0.27	0.15	0.0985	0.3268
IFNA16	rs10811525	A/G	**chr9: 21288176**	5′-Upstream	0.22	0.43	0.1043	0.2450

Abbreviations: MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
The bold gene names indicated these eight genes cluster in a small region of chromosome 9 (also see Figure 2).

density study indicated that the ROA pathway may influence the height variation through different rate of long bone development and remodeling. Further functional researches are warranted.

A major strength of this study may be embodied by raised importance to height of the ROA pathway, which are largely neglected by previous GWASs.[1] This may partially attribute to the pathway-based approach applied here that allows to spontaneously examine the genetic information imbedded in multiple SNPs and genes. This nature of the pathway-based method alleviates the punishment for the multiple-testing problem, and allows examining the combined effects other than individual genes whose effects are too modest to be detected through single-SNP tests. Given the complexity of biological systems, such circumstance should not be rare. Therefore, we argue that the pathway-based tests can serve as a useful complement of traditional GWAS. In addition, we emphasized that as the pathway analysis aims to jointly consider multiple variants and measures the enrichment of gene set at the top of the entire ranked gene list, only a few dramatically significant SNPs in GWAS study (for example SNPs in IHH gene of Hedgehog Signaling) may not necessarily pull the entire pathway to the top. This phenomenon cannot be defined as the 'non-replication' for earlier studies.

However, several concerns need to be clarified in our study. First, the average age is over 70 years for Chinese and >50 years for Whites. To use stature data of aged group cannot represent a typical adult height variation. As the bone remodeling is active in the elder people, we may actually detect a pathway that enhanced the bone absorbing process. Some famous validated pathways (for example Hedgehog Signaling) in whites were not found by our study, one of the potential confounding factors could be the sample age bias. The ethnic difference was also another factor for non-replication because those pathways were initially identified at the Caucasian ethnic context.[1]

Second, our sample size is small. Large sample size can dramatically improve the statistical power for detecting the potential pathways. The association signal of ROA in our study (618 individuals) was weak (nominal *P*=0.012) compared with our US replication study (nominal

P=0.002, 1000 individuals). Therefore, we reported the ROA pathway based on the statistical evidence with a consistently increased trend.

Third, although pathway-based analysis is a promising strategy to data mining of GWAS data sets, it is currently unclear what the best strategy is. Several methods are developed for pathway hunting recently.[26–30] All of these methods used relatively complex model or simulation to improve the accuracy and statistical power. In fact, how to properly account for the gene size, gene number and linkage disequilibrium (LD) and how to adopt the optimal SNP weighting scheme (for example the representative SNPs located in the coding sequence could be more important than others) are worth further investigating. We used this method because it was generally robust to gene size, LD, gene number and, thus, accepted by the community.[14,31,32] In addition, it was the real 'hypothesis-free' approach for pathway investigation, which means it can comprehensively test the association across different up-to-date pathway databases without prior hypotheses. Some studies first identified promising genes in one hypothesized pathway, then assessed the association of those genes with specific traits such as Parkinson disease.[33,34] However, similar approaches were limited for quantitative complex traits (height), because the unknown fundamental biological processes may influence the phenotypes, such as the EphrinA–EphR pathway for bone geometry,[31] the C-type natriuretic peptide signaling for human height.[12]

Finally, there are eight genes (eight representative SNPs) of ROA pathway cluster in a small region as shown in Figure 2 and Table 4. In addition, we constructed the LD block map for them (Figure 3), and six of them were located in one strong LD block. Although the permutation approaches may relieve this issue,[14,32] the cluster of SNPs could still influence the result, and we lack a reliable method to accurately evaluate the extent of the influence. In our recent study, we performed the pathway-based study through shuffling the genotypic data (not phenotypic as the original).[35] This procedure can completely eliminate the influence of strong LD, because it breaks the structure of original LD. However, the simulated results showed that the false-positive rate was not reduced; rather, it was slightly larger than the one by shuffling phenotypic data. Therefore, the influence of strong LD
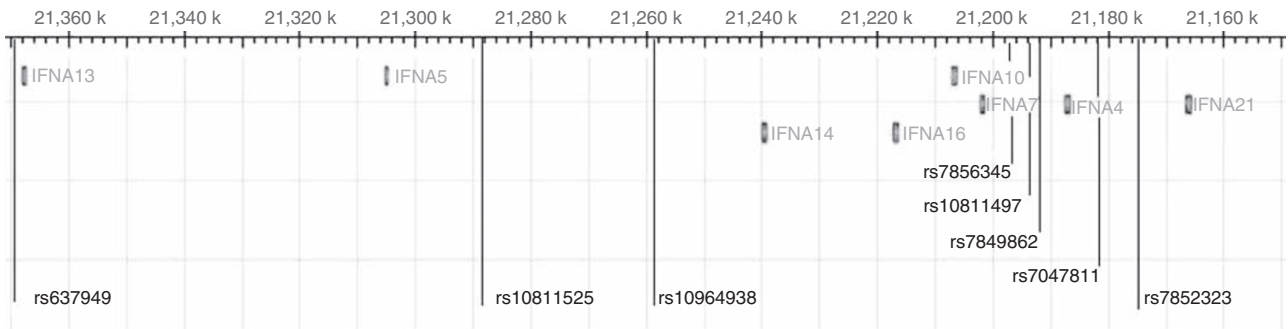
Figure 2 Physical position of the eight IFNA family members, including their corresponding representative SNPs located on the chromosome 9.
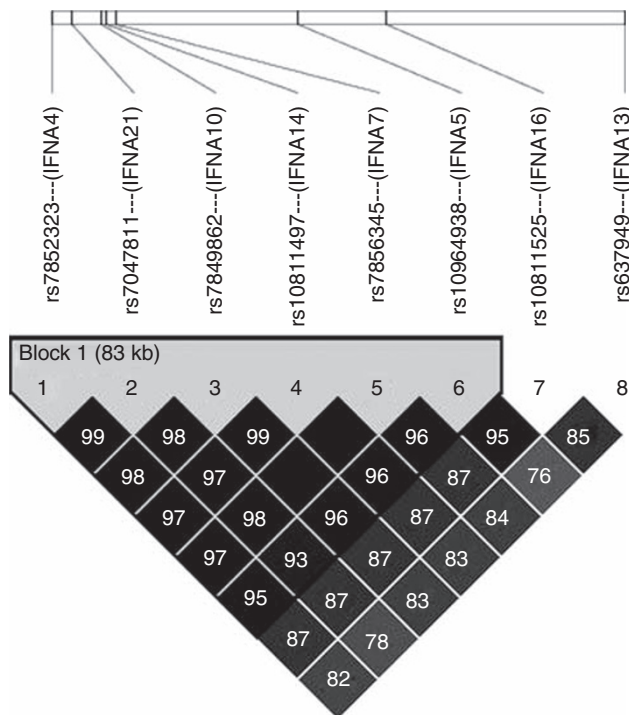

Figure 3 The haplotype block map for the eight representative SNPs of clustered IFNA family members. It was constructed using our own genotype data showing pair-wise linkage disequilibrium in r^2.

actually is far more less than one can expect. To permutate, the phenotypic data were a more conservative pathway-based association approach. To report this result, we mainly based on the statistical evidence and the independent replication study. Therefore, our study may only be served as a tentative investigation using this kind of test strategy. The future studies with larger sample size are necessary to validate our result.

In summary, the ROA pathway was suggestively associated with the height variation in Han Chinese in our pathway-based GWAS analyses. The genetic findings were supported by their biological implications in long bone development and replication study in the whites. Further functional researches are still warranted to fill the biological mechanism details.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Weedon, M. N. & Frayling, T. M. Reaching new heights: insights into the genetics of human stature. *Trends Genet.* **24,** 595–603 (2008).
2 Li, M. X., Liu, P. Y., Li, Y. M., Qin, Y. J., Liu, Y. Z. & Deng, H. W. A major gene model of adult height is suggested in Chinese. *J. Hum. Genet.* **49,** 148–153 (2004).
3 Molho-Pessach, V., Lerer, I., Abeliovich, D., Agha, Z., Abu Libdeh, A., Broshtilova, V. *et al.* The H syndrome is caused by mutations in the nucleoside transporter hENT3. *Am. J. Hum. Genet.* **83,** 529–534 (2008).
4 Ofluoglu, D., Unlu, F. & Akyuz, G. Relationship between arm span and height in postmenopausal osteoporotic women. *Rheumatol. Int.* **28,** 737–741 (2008).
5 Lei, S. F., Tan, L. J., Liu, X. G., Wang, L., Yan, H., Guo, Y. F. *et al.* Genome-wide association study identifies two novel loci containing FLNB and SBF2 genes underlying stature variation. *Hum. Mol. Genet.* **18,** 1661–1669 (2009).
6 Lettre, G., Jackson, A. U., Gieger, C., Schumacher, F. R., Berndt, S. I., Sanna, S. *et al.* Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* **40,** 584–591 (2008).
7 Johansson, A., Marroni, F., Hayward, C., Franklin, C. S., Kirichenko, A. V., Jonasson, I. *et al.* Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.* **18,** 373–380 (2009).
8 Weedon, M. N., Lettre, G., Freathy, R. M., Lindgren, C. M., Voight, B. F., Perry, J. R. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat. Genet.* **39,** 1245–1250 (2007).
9 Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* **40,** 575–583 (2008).
10 Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V., Zusmanovich, P. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40,** 609–615 (2008).
11 Krane, S. M. & Inada, M. Matrix metalloproteinases and bone. *Bone* **43,** 7–18 (2008).
12 Estrada, K., Krawczak, M., Schreiber, S., van Duijn, K., Stolk, L., van Meurs, J. B. *et al.* 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 (2009).
13 Lei, S. F., Yang, T. L., Tan, L. J., Chen, X. D., Guo, Y., Guo, Y. F. *et al.* Genome-wide association scan for stature in Chinese: evidence for ethnic specific loci. *Hum. Genet.* **125,** 1–9 (2009).
14 Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* **81,** 1278–1283 (2007).
15 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
16 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
17 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
18 Zheng, G., Freidlin, B., Li, Z. & Gastwirth, J. L. Genomic control for association studies under various genetic models. *Biometrics* **61,** 186–192 (2005).

19 Srinivas, V. & Shapiro, I. M. Chondrocytes embedded in the epiphyseal growth plates of long bones undergo autophagy prior to the induction of osteogenesis. *Autophagy* **2**, 215–216 (2006).

20 Srinivas, V., Bohensky, J., Zahm, A. M. & Shapiro, I. M. Autophagy in mineralizing tissues: microenvironmental perspectives. *Cell Cycle* **8**, 391–393 (2009).

21 Srinivas, V., Bohensky, J. & Shapiro, I. M. Autophagy: a new phase in the maturation of growth plate chondrocytes is regulated by HIF, mTOR and AMP kinase. *Cells Tissues Organs* **189**, 88–92 (2009).

22 Roach, H. I., Aigner, T. & Kouri, J. B. Chondroptosis: a variant of apoptotic cell death in chondrocytes? *Apoptosis* **9**, 265–277 (2004).

23 Lum, J. J., Bauer, D. E., Kong, M., Harris, M. H., Li, C., Lindsten, T. *et al.* Growth factor regulation of autophagy and cell survival in the absence of apoptosis. *Cell* **120**, 237–248 (2005).

24 Avnet, S., Cenni, E., Perut, F., Granchi, D., Brandi, M. L., Giunti, A. *et al.* Interferon-alpha inhibits *in vitro* osteoclast differentiation and renal cell carcinoma-induced angiogenesis. *Int. J. Oncol.* **30**, 469–476 (2007).

25 Oreffo, R. O., Romberg, S., Virdi, A. S., Joyner, C. J., Berven, S. & Triffitt, J. T. Effects of interferon alpha on human osteoprogenitor cell growth and differentiation *in vitro*. *J. Cell Biochem.* **74**, 372–385 (1999).

26 Herold, C., Steffens, M., Brockschmidt, F. F., Baur, M. P. & Becker, T. INTERSNP: genome-wide interaction analysis guided by *a priori* information. *Bioinformatics* **25**, 3275–3281 (2009).

27 O'Dushlaine, C., Kenny, E., Heron, E. A., Segurado, R., Gill, M., Morris, D. W. *et al.* The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* **25**, 2762–2763 (2009).

28 Miller, D. J., Zhang, Y., Yu, G., Liu, Y., Chen, L., Langefeld, C. D. *et al.* An algorithm for learning maximum entropy probability models of disease risk that efficiently searches and sparingly encodes multilocus genomic interactions. *Bioinformatics* **25**, 2478–2485 (2009).

29 Torkamani, A. & Schork, N. J. Pathway and network analysis with high-density allelic association data. *Methods Mol. Biol.* **563**, 289–301 (2009).

30 Peng, G., Luo, L., Siu, H., Zhu, Y., Hu, P., Hong, S. *et al.* Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.* **18**, 111–117 (2010).

31 Chen, Y., Xiong, D. H., Guo, Y. F., Pan, F., Zhou, Q., Zhang, F. *et al.* Pathway-based genome-wide association analysis identified the importance of EphrinA-EphR pathway for femoral neck bone geometry. *Bone* **46**, 129–136 (2010).

32 Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J. P., Russell, R. K. *et al.* Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am. J. Hum. Genet.* **84**, 399–405 (2009).

33 Lesnick, T. G., Papapetropoulos, S., Mash, D. C., Ffrench-Mullen, J., Shehadeh, L., de Andrade, M. *et al.* A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet.* **3**, e98 (2007).

34 Lesnick, T. G., Sorenson, E. J., Ahlskog, J. E., Henley, J. R., Shehadeh, L., Papapetropoulos, S. *et al.* Beyond Parkinson disease: amyotrophic lateral sclerosis and the axon guidance pathway. *PLoS One* **3**, e1449 (2008).

35 Guo, Y. F., Li, J., Chen, Y., Zhang, L. S. & Deng, H. W. A new permutation strategy of pathway-based approach for genome-wide association study. *BMC Bioinformatics* **10**, 429 (2009).