

Confirmation linkage study in support of the X chromosome harbouring a QTL underlying human height variation

Y-Z Liu, F-H Xu, H Shen, H Deng, Y-J Liu, L-J Zhao, V Dvornyk, T Conway, J-L Li, Q-Y Huang, K M Davies, R R Recker, H-W Deng

J Med Genet 2003;40:825–831

Human height is a complex trait determined by both genetic and environmental factors. An initial whole genome study showed several genomic regions with suggestive linkage to height in a sample of 630 subjects from 53 human pedigrees. The present study was conducted in an extended sample of 1816 subjects from 79 pedigrees in an attempt to replicate and confirm the results of the previous whole genome scan. Xq24–25 on the X chromosome was confirmed as the region suggestive of linkage to height. In the previous whole genome study, a microsatellite marker of the region DXS1001 achieved a two point LOD score of 1.91 for linkage to height. In the present study on the 79 pedigrees, another marker of the same region, DXS8067, which is only 2.7 cM away from the former marker, attained a higher two point LOD score of 2.66. Moreover, the region's significant linkage to height was sustained, with a two point LOD score of 1.00 achieved in a subset of the current sample (1026 subjects from 26 new pedigrees), which is independent of the original 630 subjects used in the whole genome study. Our results—together with identification of several syndromes with short stature, which are in linkage to Xq24–25—strongly suggest that this region may harbour a quantitative trait locus (QTL) underlying human height variation.

Human height is a typical complex trait determined by both genetic and environmental factors. Nutritional status and diseases are the most important environmental factors controlling human linear growth.^{1–4} However, genetic factors play a more dominant role in height determination. This is indicated by a significant familial aggregation of the trait, translating into a heritability of well above 50%.^{5–9}

The search for genes underlying height variation has long been an endeavour in the field of genetic studies of complex traits. Several segregation analyses on populations of different races have suggested the existence of major genes for height, though with different inheritance models.^{10–11} Association studies identified a list of candidate genes showing a significant relation to height variation. These genes include vitamin D receptor (VDR) gene,¹² D2 dopamine receptor gene,¹³ collagen I α 1 (COL1A1) gene,¹⁴ oestrogen receptor α (ER- α) gene,¹⁵ and luteinising hormone β gene.¹⁶ Stimulated by high throughput genotyping technology, genome-wide linkage scanning has recently become a major approach to the search for genes determining height variation. Whole genome studies using microsatellite markers have been undertaken by our group and others.^{11–21} Several genomic regions with significant or suggestive linkage to height were identified.

Our whole genome scan for QTLs underlying height variation was initially done on 630 subjects from 53 multiplex pedigrees of white European origin. We detected suggestive linkage to height at the regions of 5q31, 8p22, 17q25, Xp22, and Xq24–25.²⁰ The current study was undertaken to confirm those results by doing linkage analysis with 16 markers,

Key points

- An initial whole genome study in a sample of 630 subjects from 53 human pedigrees showed several genomic regions that might have linkage to height.
- The present study was conducted on an extended sample of 1816 subjects from 79 pedigrees in an attempt to replicate/confirm the results of the previous whole genome scan.
- Xq24-25 on X chromosome is confirmed as the region that may have linkage to height. Linkage to this region is suggested in both the original 53 pedigrees and the extended 79 pedigrees. Moreover, the linkage signal sustained its significance in a subset of the extended sample independent of the original 53 pedigrees.
- These results, together with identification of several syndromes with short stature which are also in linkage to Xq24-25, strongly suggest that this region may harbour a quantitative trait locus underlying human height variation.

concentrating on the above five regions in a much more extended sample of 1816 subjects from 79 pedigrees. The intermarker density was narrowed down from the previous ~10 cM to ~5 cM in the current study.

METHODS

Subjects

The study was approved by the Creighton University institutional review board. All the study subjects signed informed consent documents before entering the project. The study subjects came from an expanding database being created for studies to search for genes underlying the risk of developing osteoporosis and obesity which is under way in the Osteoporosis Research Center (ORC) of Creighton University. The sampling scheme and exclusion criteria have been detailed elsewhere.²⁰ Briefly, patients with chronic diseases and conditions which could affect bone mass were excluded from the study. These diseases/conditions include chronic disorders involving vital organs (heart, lung, liver, kidney, brain), serious metabolic diseases (diabetes, hypo- and hyperparathyroidism, hyperthyroidism, and so on), other skeletal diseases (such as Paget's disease, osteogenesis imperfecta, and rheumatoid arthritis), chronic use of drugs affecting bone metabolism (corticosteroids, anticonvulsant drugs), and conditions leading to malnutrition (chronic diarrhoea, chronic ulcerative colitis).

All the study subjects were of European origin. These subjects can be divided into three groups: 630 individuals from 53 pedigrees in our previous whole genome study were

included in the current study; 128 individuals were newly recruited but still belonged to the original 53 pedigrees; and 1058 individuals were newly recruited and made up 26 new pedigrees. The pedigrees varied in size from four to 416 individuals (mean (SD), 31.9 (48.9)). The distribution of pedigree size is summarised in table 1.

Recruitment of the sample was initially intended for osteoporosis study and was therefore based on bone mineral density (BMD) values in the probands. Among the original 53 pedigrees for our whole genome study, 50 were recruited through probands having BMD z scores of ≤ -1.28 at the hip or spine and three were recruited without regard to BMD. Among the 26 new pedigrees, 25 were recruited through probands having BMD z scores of $\geq +1.28$ at the hip or spine, and one was recruited without regard to BMD. Recruitment through probands having extreme BMD values—that is, z scores of ≤ -1.28 or $\geq +1.28$ —may help enhance the statistical power for a linkage study of BMD related traits.^{22, 23} In our sample, the correlations of sex and age adjusted height values, with thus adjusted spine BMD and hip BMD values, are 0.17 ($p < 0.01$) and 0.18 ($p < 0.01$), respectively. Therefore our sampling scheme based on BMD values may also have positive effects on the power of linkage study of height.

Genotyping

For each subject, blood (20 ml) was drawn into EDTA containing tubes by certified phlebotomists and stored chilled (at $\sim 4^{\circ}\text{C}$) until DNA extraction, which was normally completed within the next five calendar days. DNA was extracted employing a Puregene DNA isolation kit (Gentra Systems Inc, catalogue number D-5000, Minneapolis, Minnesota, USA) following the procedures detailed therein. DNA was genotyped using fluorescently labelled markers as before.²⁴ The 16 dinucleotide markers we genotyped are commercially available through Perkin Elmer Applied Biosystems (ABI PRISM linkage mapping sets, version 2, Norwalk, Connecticut, USA). The polymerase chain reaction (PCR) was done on PE 9700 thermocyclers (GeneAmp[®] PCR System 9700, Applied Biosystems, Foster City, California, USA). PCR cycling conditions followed those suggested in the ABI PRISM linkage mapping sets, version 2.5. Genotyping was done using Applied Biosystems automated DNA sequencing systems (model 3700; Perkin Elmer-ABI, Foster City, California, USA) running the GENESCAN[™] version 4.0 and GENOTYPER[™] version 4.0 softwares for allele identification and sizing. The current 3700 system in this study is a genotyping platform upgraded from ABI 377, which we used for the previous whole genome study.

As some of the markers for the current confirmation study were also genotyped in the previous whole genome study, we retyped those markers to avoid a systemic shift in genotyping results and consequent difficulty in binning adjustment.

A genetic database management system (GenoDB)²⁵ was employed to manage the phenotype and genotype data for linkage analyses. GenoDB was also employed for allele binning (including setting up allele binning criteria and converting allele sizes to distinct allele numbers), data quality control, and data formatting for PedCheck²⁶ and linkage analyses using SOLAR. PedCheck (available at http://watson.hgen.pitt.edu/register/-soft_doc.html) was employed for checking whether the data conformed to a Mendelian inheritance pattern at all the marker loci, and for confirming the alleged relations of family members within pedigrees. A genotyping error rate of 0.3% was achieved after three rounds of data check with PedCheck and three rounds of PCR amplification of the missing data and data which could not pass the PedCheck. All the 16 markers were successfully genotyped. These markers have an average population heterozygosity of ~ 0.79 and spaced on average ~ 5 cM.

Statistical analyses

The variance component linkage analysis^{27–29} for quantitative traits was used. The program employed was SOLAR (Sequential Oligogenic Linkage Analysis Routines),²⁹ which is available online (<http://www.sfbr.org/sfbr/public/software/solar/solar.html>).

In linkage analysis, age and sex were used as covariates to adjust for height, as they generally affect human height variation significantly in our study population.³⁰ Analyses were also done without adjusting for one or both of these covariates. Adjustment for significant covariates in genetic analyses can generally increase the signal to noise ratio in linkage detection by decreasing the proportion of the residual phenotypic variation attributable to random environmental factors.³¹ This may thus improve statistical power in our linkage analyses. The height data were tested by graphical methods³² and found not to deviate from normal distributions.

We also tested genetic heterogeneity for the markers or regions achieving a LOD score of > 2.0 in the linkage analysis of the 79 pedigrees. The program we used for this purpose is HOMO. It was developed by Dr Harald Göring and is available online at hgoring@darwin.sfbr.org. HOMO has been integrated into SOLAR version 2.0. It performs a test of heterogeneity using an admixture model³³ and is, in principle, quite similar to Jürg Ott's program HOMOG.³⁴

Table 1 Distribution of pedigree size

Number of subjects in a pedigree	Number of pedigrees of a certain size			
	Original sample (53 pedigrees)	Extended original sample (53 pedigrees)	Independent sample (26 pedigrees)	Extended sample (79 pedigrees)
~10	36	33	7	40
10~20	7	9	8	17
20~30	8	5	3	8
30~40	0	3	3	6
40~50	0	0	1	1
50~	2	3	4	7
Total number of subjects: 1816	Total number of pedigrees: 79			

The original sample (53 pedigrees) represents the 630 subjects from the original 53 pedigrees on which our initial whole genome study was conducted. The extended original sample (53 pedigrees) represents the original sample plus 128 newly recruited subjects, who still belong to the original 53 pedigrees. The extended sample (79 pedigrees) represents the 1816 subjects of 79 pedigrees; it is composed of the 630 subjects of the original sample, 128 subjects who are newly recruited but still belong to the original 53 pedigree, and 1058 subjects who are newly recruited and make up the 26 new pedigrees. The independent sample (26 pedigrees) represents the 1058 subjects from the 26 new pedigrees.

Table 2 Basic characteristics of the study subjects

Age group (years)	Sex	Mean height (cm)	SD	Sample size
20 to 29	M	180.7	6.7	111
	F	166.0	6.1	155
30 to 39	M	179.7	8.0	213
	F	166.4	6.6	310
40 to 49	M	179.0	6.7	271
	F	164.8	6.3	347
50 to 59	M	178.0	7.1	161
	F	163.8	5.5	210
60 to 69	M	176.2	7.1	124
	F	161.8	6.0	196
70+	M	173.2	6.1	128
	F	158.0	6.9	192

In total there were 2418 subjects in our sample with height data, who are all included here. Among them only 1816 subjects were genotyped and have genotype information. The remaining 602 subjects have not been genotyped. They are included in the table as they were also involved in linkage analysis for height. Their phenotype information, along with that of the genotyped subjects, was used in making polygenic models for adjustment of raw height data with covariates. In addition, for some of the 602 subjects, the genotype information can be inferred from that of their genotyped relatives and hence they were also used as the genotyped subjects in linkage analysis.
F, female; M, male.

RESULTS

The distribution of pedigree size is summarised in table 1. The basic characteristics of study subjects in 79 pedigrees are summarised in table 2. The relationships contained in the original pedigrees for the whole genome study, the current extended pedigrees, and the pedigrees independent of the original ones are listed in table 3. It can be seen that all the relationships informative for linkage analysis have drastically increased in the extended pedigrees. For example, the number of the sibling, grandparent–grandchild, and avuncular pairs has been more than tripled and that for cousin pairs has increased 10-fold. This sharp increase in relationship is attributed to the incorporation of several large multigenerational pedigrees to our extended sample. The biggest one contains 416 people.

Table 4 compares the results of the present study with our initial whole genome scan. As the SOLAR software for

Table 3 Relationships contained in pedigrees

	Original pedigrees	Independent pedigrees	Extended pedigrees
Number of pedigrees	53	26	79
Number of genotyped subjects	630	1058	1816
Relative pairs			
Sibling	1249	2498	3846
Grandparent–grandchild	1098	2486	3718
Avuncular	1993	5038	7170
Half sibling	23	16	39
First cousin	2589	8046	10 916
First cousin, once removed	1607	13 165	15 163
Second cousin	1971	17 762	20 348
Second cousin, once removed	12	20 045	20 111
Total	10 542	69 056	81 311

Original pedigrees represent those employed in our initial whole genome study. Independent pedigrees are the ones newly recruited, who make up the majority of our current sample. Extended pedigrees represent the total sample size of this study, who include both the original pedigrees and independent pedigrees as well as an additional set of 128 subjects who are newly recruited but still belong to the original pedigrees.

linkage analysis does not handle multipoint linkage analysis of the X chromosome, the two regions Xp22 and Xq24–25 have only two point linkage results available. As presented in table 4, Xq24–25 is the region consistently showing linkage to height in both the initial 630 subjects and the present extended sample of 1816 subjects. While in the previous study Xq24–25 achieved a two point LOD score of 1.91 at the marker DXS1001, a higher one of 2.66 is attained in the present extension study at the marker DXS8067, which is only 2.7cM away from DXS1001 (fig 1). In addition, another region on chromosome 5 (5q31) also achieved a marginally significant two point LOD score of 0.96 and a multipoint LOD score of 0.61 in the extended sample.

We also conducted linkage analysis in the subset of the sample extended from, but still limited to, the original 53 pedigrees. This subset of the sample is made up of the original 630 subjects for the whole genome study and 128 newly recruited subjects who come from the same pedigrees as the 630 subjects. Again, Xq24–25 was confirmed as the only region of significant linkage to height, with a two point LOD score of 1.46 (table 4).

To further confirm the significance for linkage of the two regions Xq24–25 and 5q31, we computed two point or multipoint LOD scores in the newly recruited 26 pedigrees composed of 1058 genotyped subjects. This subset of sample makes up the majority of our total sample size. More importantly, it is a purely independent sample in relation to the 630 subjects, on whom our initial whole genome linkage study was conducted. For the Xq24–25 region, a significant two point LOD score of 1.00 was achieved, equivalent to a nominal p value of 0.0197. For the 5q31 region, a marginally significant two point LOD score of 0.72 ($p = 0.048$) and a multipoint LOD score of 0.60 ($p = 0.07$) were attained. The above results were summarised in table 5.

A heterogeneity test for the region Xq24–25 in the 79 pedigrees did not detect significant genetic heterogeneity of height, with an hLOD value of 2.67. However, the result should be interpreted with caution owing to the test's intrinsic limitations.³⁵

Figure 1 plots X chromosome two point LOD scores for both the previous whole genome study and the present confirmation study. The figure includes 20 microsatellite markers genotyped in the previous whole genome study in the 53 pedigrees (with LOD scores denoted by open squares) and the present confirmation study in the extended 79 pedigrees (with LOD scores denoted by filled diamonds). For the initial whole genome study, the linkage signal peaks at the two regions, the region near the pter (Xp22) and the region Xq24–25. Five markers were retyped in or near these two regions in the present confirmation study. The marker, DXS8067, of the region Xq24–25 achieved a LOD score of 2.66, higher than the one (1.91) attained by another marker, DXS1001, of the same region in the initial whole genome study.

DISCUSSION

The genome-wide linkage scan has now become a major device for deciphering the genetic basis of complex traits. Theoretically, it can obviate the need to look into complicated intermediate biochemical steps and directly indicate the genomic regions that may be relevant to the development of a quantitative trait or the pathogenesis of a genetic disease.

Proven to be effective in the localisation of genes underlying monogenic characters, the whole genome linkage approach, however, faces challenges in the genetic dissection of complex traits such as height, blood pressure, and blood glucose levels. This is largely because of the complex biological basis of these characters—they are determined not only by multiple genetic and environmental factors but

Table 4 Comparison of linkage results between the previous whole genome study and the present confirmation study

	Maximum LOD scores of two point linkage analysis			Maximum LOD scores of multipoint linkage analysis		
	Whole genome scan (original sample, 53 pedigrees)	Present confirmation study		Whole genome scan (original sample, 53 pedigrees)	Present confirmation study	
		Extended original sample (53 pedigrees)	Extended sample (79 pedigrees)		Extended original sample (53 pedigrees)	Extended sample (79 pedigrees)
5q31	1.58	0.83	0.96	2.14	0.35	0.61
8p22	1.93	0.43	0.02	1.67	0.37	0
17q25	1.67	1.12	0.78	1.50	0.68	0.58
Xp22	1.95	0.21	0.47			
Xq24-25	1.91	1.46	2.66			

Original sample (53 pedigrees) represents the 630 subjects from the original 53 pedigrees, on which our initial whole genome study was conducted. Extended original sample (53 pedigrees) represents the original sample plus newly recruited 128 subjects, who still belong to the original 53 pedigrees. Extended sample (79 pedigrees) represents the 1816 subjects of 79 pedigrees. It is composed of the 630 subjects of the original sample, 128 subjects who are newly recruited but still belong to the original 53 pedigrees, and 1058 subjects who are newly recruited and make up the 26 new pedigrees.

also by their interactions, which lead to both genetic heterogeneity and phenocopies. Consequently, significant linkage results are much easier to claim than to replicate. To ensure the relevance of the results of whole genome linkage studies, stringent criteria have been proposed.³⁶ Generally, a multipoint LOD score above 3 has been accepted as a rule of thumb for a region to be statistically significant for linkage in a whole genome scan. However, for the significant levels applied to a linkage study seeking confirmation of an initial whole genome scan, as in the case of this study, no consensus has been reached. But the strict standard accounting for multiple testing in a whole genome study should definitely not be applied to a confirmation study using only a limited number of markers. In the situation where five markers are used to replicate a region, it was suggested that a nominal p value of 0.01 should be required for confirmation at a 5% significance level.³⁶

Our primary whole genome linkage scan was undertaken on 630 subjects from 53 pedigrees. In that study, 380 microsatellite markers were used, giving us an average marker density of ~10 cM/marker. Suggestive linkage results were achieved at 5q31, 8p22, 17q25, Xq24–25, and Xp22. To confirm these results, more microsatellite markers (with a density of ~5 cM/marker) within or around these regions were chosen and tested for linkage to height in our extended sample. Xq24–25 turns out to be the only region consistently showing linkage to height (tables 4 and 5). Although further

investigation is needed to validate this result, its relevance is supported by three features.

First, the significance level of linkage achieved is remarkable for a confirmation study. As suggestive linkage is already detected at the whole genome level, the current aim of confirming the linkage signal represents a common hypothesis-driven test, where nominal rather than stringent genome-wide significance levels should be used. In the light of that, the LOD score of 2.66 achieved in our extended sample should be highly significant. It is equivalent to a nominal p value of 0.00014 and even approaching the significance levels for a genome-wide linkage scan.

Second, the characteristics of our extended sample are highly suitable for a study trying to replicate a previous linkage. The majority of our extended sample is composed of newly recruited pedigrees, which themselves almost double the original sample size, making the current sample quite independent of the original one. Moreover, the newly recruited subjects are mainly from large multigeneration pedigrees. Pedigrees with at least 85 family members make up more than 60% of the newly recruited sample, the largest pedigree having 416 individuals with genotype information. These huge pedigrees provide a large number of relationships informative for linkage analysis and hence increase the statistical power of the current study substantially as compared with the original whole genome scan.

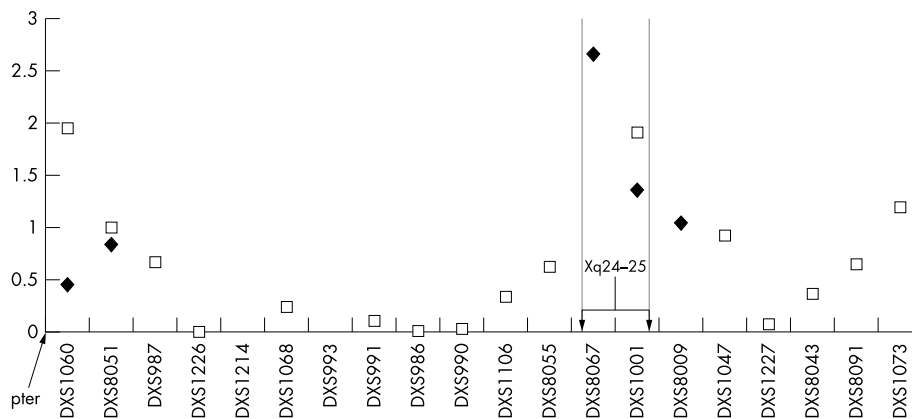


Figure 1 X chromosome two point linkage analysis results for the previous whole genome study and the present confirmation study. The X axis lists genotyped microsatellite markers arranged in the order as on the X chromosome. The Y axis is the scale for two point LOD score values. Open squares represent the LOD scores achieved in the previous whole genome study in 53 pedigrees. Filled diamonds represent the LOD scores achieved in the present confirmation study in 79 pedigrees. The grey lines enclose the region Xq24–25, where confirmative significant linkage to height was detected. Two markers, DXS8009 and DXS8067, are genotyped only in the present study. Three markers—DXS1060, DXS8051, and DXS1001—are genotyped in both the previous whole genome study and the present confirmation study.

Table 5 Linkage analysis results for the regions Xq25 and 5q31 in original sample, extended sample, and independent sample

	Whole genome scan (original sample, 53 pedigrees)	Present confirmation study		
		Extended original sample (53 pedigrees)	Independent sample (26 pedigrees)	Extended sample (79 pedigrees)
Xq24-25 Two point LOD score	1.91	1.46 ($p=0.0048$)	1.00 ($p=0.0197$)	2.66 ($p=0.0001432$)
5q31 Two point LOD score	1.58	0.83	0.72	0.96
Multipoint LOD score	2.14	0.35	0.60	0.61

Original sample (53 pedigrees) represents the 630 subjects from the original 53 pedigrees on which our initial whole genome study was conducted. Extended original sample (53 pedigrees) represents the original sample plus newly recruited 128 subjects, who still belong to the original 53 pedigrees. Extended sample (79 pedigrees) represents the 1816 subjects of 79 pedigrees. It is composed of the 630 subjects of the original sample, 128 subjects who are newly recruited but still belong to the original 53 pedigree, and 1058 subjects who are newly recruited and make up the 26 new pedigrees. Independent sample (26 pedigrees) represents the 1058 subjects from the 26 new pedigrees.

Third and most importantly, in the 1058 newly recruited subjects from the 26 new pedigrees, Xq24–25 sustained its significance for linkage to height, with a two point LOD score of 1.00, equivalent to a nominal significance level of 0.0197 (table 5). These 1058 subjects represent a purely independent subset of the total sample in relation to the 630 subjects used in our initial whole genome study. Confirmation of linkage in such an independent group of subjects constitutes an unequivocal replication for the region's importance to height identified in the genome-wide scan.³⁶

Recently, several whole genome studies have been done on height. The time is ripe to summarise the data generated by these studies for meta-analysis purposes. So far linkage signals for height have been reported for chromosomes 3, 5, 6, 7, 9, 12, 20, and X.^{11 17–21} Cross study comparison shows that two regions, 7q31–36 and 6q25, were both detected by different studies. Linkage to height of the former region was identified both in 614 subjects from 247 Finnish families¹⁹ and in 746 subjects from 179 Sweden families.¹⁸ Linkage to height of the latter region was detected both in 408 subjects from 58 Finnish families¹⁸ and in 1184 subjects from 200 Dutch pedigrees.¹¹ The replication achieved for these two regions implicates the relatively homogeneous genetic basis for height determination in northern Europeans. On the other hand, the importance of these two regions for height may be limited to the northern European population—an isolated population with a history marked by population bottle necks and random genetic drifts.

In contrast, in a relatively less homogeneous population—such as the white US subjects in our study—identification of genetic factors underlying height variation should be much more elusive and challenging, as genetic heterogeneity is to be expected. However, the answers to this search have more universal implications for human biology. First, in order to ensure statistical power, the gene mapping in less homogeneous—as compared to isolated—populations requires a much larger sample size, which is the prime advantage of this study over others in the field. Among all the linkage studies on height, this study has not only the largest number of individuals but also the most complex pedigrees and the most abundant informative relationships, which give us the highest statistical power over other studies. This is particularly meaningful for a replication study in two areas. First, only with high statistical power can it be possible, as well as theoretically justified, for a study to replicate linkage findings caused by QTLs with moderate effects,³⁷ which are expected to be the genetic basis of height. Second, only with high statistical power can one be confident enough to reject

previous linkage findings that could not be replicated in a confirmation study.³⁶

The X chromosome plays an important role in human height. This is evidenced by the large number of X linked diseases involving abnormal linear growth. In particular, for the genomic region Xq24–25, linkage to several syndromes with short stature has recently been discovered. These syndromes include the MRSS syndrome (X linked mental retardation with short stature),³⁸ the MRGH syndrome (X linked mental retardation with isolated growth hormone deficiency),^{39 40} panhypopituitarism,⁴¹ and a syndrome reported by Cabezas *et al* (X linked mental retardation with short stature, small testes, muscle wasting, and tremor).⁴² These findings, together with the present study confirming the linkage of Xq24–25 to normal human height variation, shed light on the region for its potential involvement in human linear growth and support the existence of a QTL in this region for height determination. Given such a critical role of the X chromosome in human height development, surprisingly little effort has been made to map the X chromosome in the whole genome study of human height. To date, our initial whole genome study²⁰ and the present confirmation study are the only ones exploring the X chromosome in a genome-wide scale for its importance in height determination.

Some genomic regions showing suggestive linkage to height in our last whole genome scan could not be replicated in this study. A significant decrease in LOD scores is seen for 8p22, 17q25, and Xp22. Confirmation for the region 5q31 is tenuous, as only marginally significant results were achieved.

Given the relatively high statistical power of this study, it is unlikely that the disappearance of these regions' linkage signal for height reflects random fluctuations. However, there are still certain factors that could lead to a drastic difference in linkage results between this study and the earlier whole genome scan. First, the allele frequencies may have changed in the present study for the markers genotyped in both the current and the previous whole genome study. As IBD (identity by descent) inference is based on allele frequencies, the changes of the latter will inevitably to some extent affect the computation results of IBD and other downstream parameters, including the multipoint and two point LOD scores. Second, the genotyping errors may have substantial effects on linkage results. Despite the similar genotyping error rate (0.3%) for both the current study and the initial whole genome study, the distribution of the errors may vary between the two studies. Such errors can seriously deflate power and inflate recombination fraction estimates in two

point linkage analysis.^{43–45} In multipoint linkage analysis, the effects of the errors can be further magnified as marker density increases,⁴⁶ leading to increased potential for false exclusion of the true disease loci. Third, even if those unreplicated regions do harbour QTLs underlying height variation, it is still possible that those regions have gone undetected in a replication study owing to the expected minor effects of each QTL on height. Such a situation can be explained by the fact that a study should have a much greater power—given a certain sample size—to detect than to repeat the findings of any one of multiple genes underlying a complex trait.⁴⁷

A limitation of our study is the absence of multipoint linkage results for the X chromosome. This is because of the inability of the SOLAR software to handle multipoint analysis for the X chromosome. Other softwares, such as GENEHUNTER,⁴⁸ which has that function, unfortunately cannot handle the large multiplex pedigrees that make up most of our study sample. Breaking down these large pedigrees into smaller ones is an option, but this procedure may result in a considerable loss of statistical power—especially for a sample like the present study with huge multigeneration pedigrees as the main component. Although multipoint linkage results are more accurate than two point linkage, our two point results on the X chromosome still provide a clear and consistent pattern for linkage of Xq24–25 to height. The selection of this region for confirmation study is based on our linkage scan of the X chromosome using 18 microsatellite markers,²⁰ giving us a marker density of ~10 cM/marker. A clear peak of linkage signal was formed in the Xq24–25 region, with the marker DXS1001 having the highest LOD score (fig 1). Three markers, including DXS1001 and two flanking ones, are thus chosen to saturate this region in the present study. The marker DXS8067, which is just 2.7 cM away from DXS1001, achieved a much higher LOD score of 2.66 (fig 1). More notably, in the purely independent subset of the sample—that is, the 1058 subjects from 26 pedigrees—the significant linkage of the marker DXS8067 to height is maintained. Given such a remarkable increase in LOD score in the extended sample, and a consistently significant linkage in the independent subset of the sample, the chance of a type I error for the linkage of Xq24–25 is remote. Further investigation of this region's importance is justified.

This study represents the first follow up study in the field of genome-wide scan for height. It is also the first one bringing to light the X chromosome as the potential genetic basis of normal human height variation. Based on the result, subsequent fine mapping studies will be pursued. Candidate genes inside the Xq24–25 region will be subject to association studies and transmission disequilibrium test studies⁴⁹ in pursuit of functional mutations. Our efforts will contribute to the understanding of genetics of human growth and the genetic architecture of complex traits as a whole.

ACKNOWLEDGEMENTS

The investigators were partially supported by grants from Health Future Foundation, NIH grants (K01 AR02170–01, R01 AR45349–01, R01 GM60402–01A1, P01 DC01813–07), grants from State of Nebraska cancer and smoking related disease research program and the State of Nebraska tobacco settlement fund, US Department of Energy grant DE-FG03–00ER63000/A00, grants from Creighton University, grants from National Science Foundation of China, a grant from Huo Ying Dong Education Foundation, and grants from HuNan Normal University and the Ministry of Education of China.

Authors' affiliations

Y-Z Liu, F-H Xu, H Shen, H Deng, Y-J Liu, L-J Zhao, V Dvornyk, T Conway, Q-Y Huang, K M Davies, R R Recker, H-W Deng, Osteoporosis Research Center, Creighton University, Omaha, Nebraska, USA

J-L Li, Center for Medical Informatics, School of Medicine, Yale University, New Haven, Connecticut, USA

Correspondence to: Dr Hong-Wen Deng, Osteoporosis Research Center, Creighton University, 601 N 30th St, Suite 6787, Omaha, Nebraska 68131, USA; deng@creighton.edu

REFERENCES

- Allen LH. Malnutrition and human function: a comparison of conclusions from the INCAP and nutrition CRSP studies. *J Nutr* 1995;**125**:1119–26S.
- Fewtrell MS, Cole TJ, Bishop NJ, Lucas A. Neonatal factors predicting childhood height in preterm infants: evidence for a persisting effect of early metabolic bone disease? *J Pediatr* 2000;**137**:668–73.
- Newell ML, Borja MC, Peckham C. Height, weight, and growth in children born to mothers with HIV-1 infection in Europe. *Pediatrics* 2003;**111**:e52–60.
- Torres AM, Peterson KE, de-Souza AC, Orav EJ, Hughes M, Chen LC. Association of diarrhea and upper respiratory infections with weight and height gains in Bangladeshi children aged 5 to 11 years. *Bull WHO* 2000;**78**:1316–23.
- Pearson K, Lee A. On the laws of inheritance in man. *Biometrika* 1903;**2**:356–462.
- Phillips K, Matheny AP. Quantitative genetic analysis of longitudinal trends in height: preliminary results from the Louisville Twin Study. *Acta Genet Med Gemellol* 1990;**39**:143–63.
- Carmichael CM, McGue M. A cross-sectional examination of height, weight, and body mass index in adult twins. *J Gerontol A Biol Sci Med Sci* 1995;**50**:B237.
- Preece MA. The genetic contribution to stature. *Horm Res* 1996;**45**:56–8.
- 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 Pub Health* 2000;**90**:627–30.
- Ginsburg E, Livshits G, Yakovenko K, Kobylansky E. Major gene control of human body height, weight and BMI in five ethnically different populations. *Ann Hum Genet* 1998;**62**:307–22.
- Xu JF, Bleecker ER, Jongepier H, Howard TD, Koppelman GH, Postma DS, Meyers DA. Major recessive gene(s) with considerable residual polygenic effect regulating adult height: confirmation of genomewide scan results for chromosomes 6, 9, and 12. *Am J Hum Genet* 2002;**71**:646–50.
- Minamitani K, Takahashi Y, Minagawa M, Yasuda T, Niimi H. Difference in height associated with a translation start site polymorphism in the Vitamin D receptor gene. *Pediatric Res* 1998;**44**:628–32.
- Miyake H, Nagashima K, Onigata K, Nagashima T, Takano Y, Morikawa A. Allelic variations of the D2 dopamine receptor gene in children with idiopathic short stature. *J Hum Genet* 1999;**44**:26–9.
- Garnero P, Borel O, Grant SFA, Ralston SH, Delmas PD. Collagen 1 α 1 Sp1 polymorphism, bone mass, and bone turnover in healthy French premenopausal women: the OFELY study. *J Bone Miner Res* 1998;**13**:813–17.
- Lorentzon M, Lorentzon R, Bäckström T, Nordström P. Estrogen Receptor gene polymorphism, but not estradiol levels, is related to bone density in healthy adolescent boys: a cross-sectional and longitudinal study. *J Clin Endocrinol Metab* 1999;**84**:4597–601.
- Raivio T, Huhtaniemi I, Anttila R, Siimes MA, Hagenäs L, Nilsson C, Pettersson K, Dunkel L. The role of luteinizing hormone- β gene polymorphism in the onset of progression of puberty in healthy boys. *J Clin Endocrinol Metab* 1996;**81**:3278–82.
- Thompson DB, Ossowski V, Janssen RC, Knowler WC, Bogardus C. Linkage between stature and a region on chromosome 20 and analysis of a candidate gene, bone morphogenetic protein 2. *Am J Med Genet* 1995;**59**:495–500.
- Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burtt NP, Altshuler D, Parker A, Rioux JD, Platko J, Gaudet D, Hudson TJ, Groop LC, Lander ES. Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. *Am J Hum Genet* 2001;**69**:106–16.
- Perola M, Öhman M, Hiekkalinna T, Leppävuori J, Pajukanta P, Wessman M, Koskenvuo M, Palotie A, Lange K, Kaprio J, Peltonen L. Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *Am J Hum Genet* 2001;**69**:117–23.
- Deng HW, Xu FH, Liu YZ, Shen H, Deng HY, Huang QY, Liu YJ, Conway T, Li JL, Davis KM, Recker RR. A whole-genome linkage scan suggests several genomic regions potentially containing QTLs underlying the variation of stature. *Am J Med Genet* 2002;**113**:29–39.
- Wiltshire S, Frayling TM, Hattersley AT, Hitman GA, Walker M, Levy JC, O'Rahilly S, Groves CJ, Menzel S, Cardon LR, McCarthy MI. Evidence for linkage of stature to chromosome 3p36 in a large UK family data set ascertained for type 2 diabetes. *Am J Hum Genet* 2002;**70**:543–6.
- Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 1995;**268**:1584–9.
- Deng HW, Li J. The effect of selective sampling on transmission disequilibrium test of a QTL. *Genet Res* 2002;**79**:161–74.
- Deng HW, Zhou Y, Johnson M, Li J, Recker RR. Product difference of multiplex and singleplex PCR and its practical implications for human genetic studies. *BioTechniques* 2000;**29**:298–308.
- Li JL, Deng HY, Lai DB, Recker RR, Deng HW. Towards high-throughput genotyping: A dynamic and automatic software for manipulating large-scale genotype data using fluorescently labeled dinucleotide markers. *Genome Res* 2001;**11**:1304–14.
- O'Connell JR, Weeks DE. PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.

- 27 **Amos CI.** Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;**54**:535–43.
- 28 **Amos CI, Zhu DK, Boerwinkle E.** Assessing genetic linkage and association with robust components of variance approaches. *Ann Hum Genet* 1996;**60**:143–60.
- 29 **Almasy L, Blangero J.** Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;**62**:1198–211.
- 30 **Deng HW, Li J, Li JL, Johnson M, Dowd R, Gong G, Deng HY, Recker RR.** Association of Estrogen Receptor- α (ER) Genotypes with Body Mass Index in Normal Healthy Postmenopausal Caucasian Women. *J Clin Endocrinol Metab* 2000;**85**:2748–51.
- 31 **Deng HW, Li J, Li JL, Johnson M, Recker RR.** Association of VDR and ER genotypes with bone mass in postmenopausal women: different conclusions with different analyses. *Osteoporosis Int* 1999;**9**:499–507.
- 32 **Sokal RR, Rohlf FJ.** *Biometry*, 3rd ed. New York: WH Freeman and Co, 1995.
- 33 **Smith CAB.** Testing for heterogeneity of recombination fraction values in human genetics. *Ann Hum Genet* 1963;**27**:175–82.
- 34 **Ott J.** Linkage analysis and family classification under heterogeneity. *Ann Hum Genet* 1983;**47**:311–20.
- 35 **Whittemore AS, Halpern J.** Problems in definition, interpretation, and evaluation of genetic heterogeneity. *Am J Hum Genet* 2001;**68**:457–65.
- 36 **Lander E, Kruglyak L.** Genetic dissection of complex traits: guidelines for interpreting the reporting results. *Nature Genetics* 1995;**11**:241–7.
- 37 **Liu YZ, Liu YJ, Recker RR, Deng HW.** Molecular studies of gene identification for osteoporosis: the 2002 update. *J Endocrinol* 2003;**177**:147–96.
- 38 **Vitale E, Specchia C, Devoto M, Angius A, Rong S, Rocchi M, Schwalb M, Demelas L, Paglietti D, Manca S, Mastropaolo C, Serra G.** Novel X-linked mental retardation syndrome with short stature maps to Xq24. *Am J Med Genet* 2001;**103**:1–8.
- 39 **Hamel BCJ, Smits APT, Otten BJ, van den Helm B, Ropers HH, Mariman ECM.** Familial X-linked mental retardation and isolated growth hormone deficiency: clinical and molecular findings. *Am J Med Genet* 1996;**64**:35–41.
- 40 **Raynaud M, Ronce N, Ayrault AD, Francannet C, Malpuech G, Moraine C.** X-linked mental retardation with isolated hormone deficiency is mapped to Xq22-Xq27.2 in one family. *Am J Med Genet* 1998;**76**:255–61.
- 41 **Lagerstrom-Femer M, Sundvall M, Johnsen E, Warne GL, Forrest SM, Zajac JD, Rickards A, Ravine D, Landegren U, Pettersson U.** X-linked recessive panhypopituitarism associated with a regional duplication in Xq25-26. *Am J Hum Genet* 1997;**60**:910–16.
- 42 **Cabezas DA, Slaugh R, Abidi F, Arena JF, Stevenson RE, Schwartz CE, Lubs HA.** A new X-linked mental retardation (XLMR) syndrome with short stature, small testes, muscle wasting, and tremor localizes to Xq24-25. *J Med Genet* 2000;**37**:663–8.
- 43 **Smith HF.** Test of significance for Mendelian ratios when classification is uncertain. *Ann Eugen* 1937;**8**:94–5.
- 44 **Terwilliger JD, Weeks DE, Ott J.** Laboratory errors in the reading of marker alleles cause massive reductions in LOD score and lead to gross overestimation of the recombination fraction [abstract]. *Am J Hum Genet* 1990;(suppl 47):A201.
- 45 **Buetow KH.** Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* 1991;**49**:985–94.
- 46 **Shields DC, Collins A, Buetow KH, Morton NE.** Error filtration, interference, and the human linkage map. *Proc Natl Acad Sci USA* 1991;**88**:6501–5.
- 47 **Recker RR, Deng HW.** Role of genetics in osteoporosis. *Endocrine* 2002;**17**:55–66.
- 48 **Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES.** Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996;**58**:1347–63.
- 49 **Spielman RS, McGinnis RE, Ewens WJ.** Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;**52**:506–16.

ECHO

***IBD5* risk haplotype and *CARD15* variants add up to a bigger chance of Crohn's disease**



Please visit the Journal of Medical Genetics website [www.jmedgenet.com] for a link to the full text of this article.

A genetic study has extended our understanding of susceptibility to Crohn's disease, showing that the 5q31 (*IBD5*) cytokine cluster in British Caucasians does not interact with *CARD15* variants and is rare in Japanese people. Nevertheless, potential genetic interactions still remain possible if variants in other susceptibility loci are identified, the researchers say.

IBD5 haplotype showed strong linkage disequilibrium with CD phenotype in a family association based study in British Caucasians. No evidence of genetic interaction between *IBD5* haplotype, *CARD15* variants, and *TNF*₋₈₅₇ was identified in stratified transmission disequilibrium tests or log linear analysis. *IBD5* haplotype was rare in the Japanese cohort: none was homozygous for the haplotype, preventing further tests.

The researchers performed family based association studies on 282 CD trios identified from the Oxford Gastroenterology Unit, UK, who were screened for *IBD5* haplotype, *CARD15*, and *TNF*₋₈₅₇ variants. The Japanese cohort contained 178 patients with CD from a hospital in Sendai, Japan, and 156 volunteers acting as controls, who were screened for *IBD5* haplotype.

Among Caucasians, variants at three loci are associated with CD: chromosome 5q31 (*IBD5*), *TNF*₋₈₅₇, and *CARD15/NOD2*. Multiple variants are suspected in this disease, with the potential for genetic—epistatic—interaction among them, not just an additive effect. Evidence in a mouse model of IBD has shown that epistatic interactions can occur among loci for susceptibility to colitis.

▲ *Gut* 2003;**52**:541–546.