

ARTICLE

Genome-wide linkage screen for stature and body mass index in 3.032 families: evidence for sex- and population-specific genetic effects

Sampo Sammalisto¹, Tero Hiekkalinna^{1,2}, Karen Schwander³, Sharon Kardia⁴, Alan B Weder⁵, Beatriz L Rodriguez⁶, Alessandro Doria⁷, Jennifer A Kelly⁸, Gail R Bruner⁸, John B Harley⁸, Susan Redline⁹, Emma K Larkin⁹, Sanjay R Patel¹⁰, Amy JH Ewan¹¹, James L Weber¹¹, Markus Perola^{1,2} and Leena Peltonen^{*,1,2,12}

¹Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland; ²Department of Medical Genetics, Faculty of Medicine, University of Helsinki, Helsinki, Finland; ³Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; ⁴Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA; ⁵Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, MI, USA; ⁶Pacific Health Research Institute, Honolulu, HI, USA; ⁷Research Division, Joslin Diabetes Center, Harvard Medical School, Boston, MA, USA; ⁸Oklahoma Medical Research Foundation, Arthritis and Immunology, Oklahoma City, OK, USA; ⁹Center for Clinical Investigation, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; ¹⁰Division of Pulmonary Medicine, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; ¹¹Marshfield Clinic Research Foundation, Marshfield, WI, USA; ¹²The Broad Institute, MIT, Boston, MA, USA

Stature (adult body height) and body mass index (BMI) have a strong genetic component explaining observed variation in human populations; however, identifying those genetic components has been extremely challenging. It seems obvious that sample size is a critical determinant for successful identification of quantitative trait loci (QTL) that underlie the genetic architecture of these polygenic traits. The inherent shared environment and known genetic relationships in family studies provide clear advantages for gene mapping over studies utilizing unrelated individuals. To these ends, we combined the genotype and phenotype data from four previously performed family-based genome-wide screens resulting in a sample of 9.371 individuals from 3.032 African-American and European-American families and performed variance-components linkage analyses for stature and BMI. To our knowledge, this study represents the single largest family-based genome-wide linkage scan published for stature and BMI to date. This large study sample allowed us to pursue population- and sex-specific analyses as well. For stature, we found evidence for linkage in previously reported loci on 11q23, 12q12, 15q25 and 18q23, as well as 15q26 and 19q13, which have not been linked to stature previously. For BMI, we found evidence for two loci: one on 7q35 and another on 11q22, both of which have been previously linked to BMI in multiple populations. Our results show both the benefit of (1) combining data to maximize the sample size and (2) minimizing heterogeneity by analyzing subgroups where within-group variation can be reduced and suggest that the latter may be a more successful approach in genetic mapping.

European Journal of Human Genetics (2009) 17, 258–266; doi:10.1038/ejhg.2008.152; published online 10 September 2008

Keywords: body height; body mass index; linkage mapping; quantitative trait loci

*Correspondence: Professor L Peltonen, Department of Medical Genetics, University of Helsinki, National Public Health Institute, Biomedicum, Haartmaninkatu 8, Helsinki, FIN 00251, Finland. Tel: +358 9 4744 8393; Fax: +358 9 4744 8480; E-mail: leena.peltonen@ktl.fi
Received 30 October 2007; revised 1 July 2008; accepted 16 July 2008; published online 10 September 2008

Introduction

Stature, that is adult body height, and body mass index (BMI) are typical complex traits that have a well-established genetic component as well as well-known environmental and lifestyle factors that act independently and perhaps in concert with genetic components in the determination of the trait phenotype. Both stature and BMI have relatively high proportion of variance explained by genetic components (heritability)^{1,2} and should thus be amenable to genetic mapping of underlying quantitative trait loci (QTL).

Genetic linkage mapping has been highly successful in the study of Mendelian traits, where a single-gene mutation (or several closely linked mutations within the same gene) explains most of the observed trait variation. Linkage mapping focuses on observed recombinations and is thus useful for narrowing down the chromosomal region that contains loci influencing the trait. Association (or linkage disequilibrium) mapping on the other hand is utilized to directly identify the genetic variant(s) influencing the trait (or variants in linkage disequilibrium with it). Recently with the advent of rapid and affordable genome-wide single-nucleotide polymorphism genotyping, genome-wide association (GWA) studies have been proposed as a more powerful approach compared to traditional linkage mapping for identifying genetic variants underlying complex traits.³ However, even though the GWA study design holds great promise for genetic discoveries, there is also a plethora of genome-wide marker data for family samples that have yet to be utilized for genetic linkage analyses of available phenotypes such as stature and BMI. As these family samples may still hold unexplored potential for genetic discoveries,⁴ they should be exploited to the fullest extent possible.

We were able to obtain genome-wide genotype and phenotype data for 9,371 individuals from 3,032 families and performed genome-wide variance-components linkage analysis for locating QTL that influence stature and BMI. To our knowledge, this study represents the single largest family-based genome-wide scan published to date that is based on combining primary data instead of meta-analysis of primary results. We corroborate multiple previously reported QTL for stature and BMI as well as provide linkage evidence for two novel loci for stature. In addition, we were

able to demonstrate population- and sex-specific genetic effects underlying these phenotypes.

Materials and methods

Family samples

The samples analyzed in the study consisted of four independent genome-wide screened family cohorts: the Joslin study of type 2 diabetes, the Family Blood Pressure Program (FBPP), systemic lupus erythematosus (SLE) families and the Cleveland Family Study. The number of families and individuals with genotype and phenotype information are listed in Table 1. Details of the patient ascertainment and genotyping can be found from the respective initial publications. For the purpose of minimizing within-family genetic heterogeneity, we discarded all families with members of multiple ethnicities in each sample before analysis.

From the Joslin study of type 2 diabetes, we included 45 families originally ascertained for maturity-onset diabetes of the young (MODY).⁵ As the majority of the sample consisted of European-American families (39 of 45), we excluded one Asian-American, two Hispanic and three African-American families before analysis. A total of 39 European-American families consisting of 282 with genotype and phenotype information (152 females and 130 males) were included in all subsequent analyses.

The FBPP is a multicenter collaborative effort established by the National Heart, Lung, and Blood Institute (NHLBI) to localize and identify genes underlying blood pressure and hypertension.⁶ The FBPP consists of the African-American, Mexican-American, Asian and non-Hispanic white populations but because we had access to a limited number of Asian- and Mexican-American families from the other studies, we included only African-American and European-American families in our analyses. In total, we analyzed 1496 African-American families with 4009 individuals (2663 females and 1346 males) and 1151 European-American families with 3641 individuals (2004 females and 1637 males).

The SLE families were ascertained from the Lupus Family Registry and Repository (<http://lupus.omrf.org>). All SLE cases met 4 of the 11⁷ criteria for SLE. These families included 73 African-American families with 281

Table 1 Number of individuals with genotype and phenotype information by cohort and ethnicity

Cohort	African-American		European-American		All	
	Families	Individuals (females/males)	Families	Individuals (females/males)	Families	Individuals (females/males)
Joslin study of type 2 diabetes	0	0	39	282 (152/130)	39	282 (152/130)
Family Blood Pressure Program	1496	4009 (2663/1346)	1151	3641 (2004/1637)	2647	7650 (4667/2983)
Cleveland Family Study	73	281 (231/50)	148	720 (518/202)	221	1001 (749/252)
Obstructive Sleep Apnea	59	176 (100/76)	66	262 (120/142)	125	438 (220/218)
Total	1628	4466 (2994/1472)	1404	4905 (2794/2111)	3032	9371 (5788/3583)

individuals (231 females and 50 males) and 148 European-American families with 720 individuals (518 females and 202 males).^{8,9} These families are partially included in the paper,¹⁰ which reports individuals from 56 African-American families, 92 European-American families and 12 families of other ethnic backgrounds. This paper builds on the samples described by Gray-McGuire *et al*⁸ and Moser *et al*⁹ with 34 additional families.

Participants from the Cleveland Family Study were ascertained from first or selected second-degree relatives of a proband with either laboratory diagnosed obstructive sleep apnea or neighborhood control of an affected proband. Families were selected for genotyping on the basis of genetic informativity, including multigenerational data or individuals from the extremes of the distribution of apnea phenotype.^{11,12} These families included 59 African-American families with 176 individuals (100 females and 76 males) and 66 European-American families with 262 individuals (120 females and 142 males) with genotype and phenotype information.

Statistical analysis of the phenotype data

Phenotypic distributions were analyzed within each cohort using SPSS 14.0 (SPSS, Chicago, IL, USA). We excluded the phenotype information of all individuals less than 23 years of age, as they may still be growing and may have not reached their final adult height. However, the genotypes of these individuals were retained to maximize linkage phase information and to obtain more accurate allele frequency estimates. In each cohort, we removed phenotypic outliers (defined as individuals whose phenotypic value deviated 3 standard deviations or more from their respective ethnic- and sex-specific mean) because they may have disproportionate weight in the genetic analyses.¹³ Both stature and BMI were approximately normally distributed, thus no transformations were applied before variance-component analyses.

Familial relationship and genotype checks

As all genetic analyses performed assume that the familial relationship information is correct, we utilized genome-wide genotype data to verify relationships using the GRR program.¹⁴ The genotype and phenotype data of all pairs whose reported relationship was likely to have been misspecified were removed before analysis.

Mendelian consistency of the genotype data was verified with PedCheck¹⁵ and all inconsistencies were removed before genetic analysis. We also performed the multipoint method of detecting unlikely yet Mendelian consistent genotypes implemented in Merlin¹⁶ and excluded those that were unlikely using the Pedwipe-program included in the Merlin package. We tested each marker for deviation from Hardy–Weinberg equilibrium (HWE) with PedStats and excluded all markers deviating from HWE after controlling for multiple testing using the Sidak's correction¹⁷ with $\alpha = 0.05$.

Genetic map construction

We used the published deCODE genetic map¹⁸ to assign the genetic location for each marker. For those markers that were not included in the original deCODE genetic map, we used the physical position for linear interpolation of the genetic location using the physical and genetic locations of the flanking deCODE markers. Building of the genetic map was automated using Cartographer.¹⁹

Variance-components linkage analyses

Variance-components linkage analyses were performed using Merlin. Marker allele frequencies were estimated before linkage analysis, within each cohort for European-Americans and African-American separately using the observed genotype data by maximum likelihood as implemented in Merlin. The sex and age of the individuals at the time of phenotype determination were used as covariates in the variance-components model, and to control for heterogeneity between the cohorts, the cohort identifier and ethnicity were included as covariates in the combined analyses. As it has been shown that inclusion of ungenotyped and unphenotyped individuals that are not required for specification of pedigree structure introduces noise in the variance-components analyses,²⁰ we used Merlin's trim option to automatically remove these individuals before genetic analysis. All Merlin analyses were automated using AUTOGSCAN.²¹

Results

We collected genome-wide genetic marker information for families ascertained for various adult onset phenotypes genotyped at the National Heart, Lung, and Blood Institute (NHLBI) Mammalian Genotyping Service. The total number of informative individuals with genome-wide genotype and phenotype data was 9,371 from 3,032 families (Table 1). The numbers of informative relative pairs for the variance-components analysis by sex and ethnicity are shown in Table 2, and the phenotypic characteristics are summarized in Table 3. The mean BMI and standard deviation were largest in the African-American female sample. In the European-Americans, the mean BMI was slightly smaller for females compared to males, but the female standard deviation was larger. In males, the mean BMI was larger in European-Americans, but the standard deviation was larger in African-American males. The mean and standard deviations for stature between African-Americans and European-Americans were nearly identical in both females and males.

Similar to many other quantitative traits, stature and BMI are good examples of sexually dimorphic traits that may have sex-specific genetic backgrounds.²² Therefore, we also performed sex-limited linkage analyses attempting to localize such sex-specific QTL by ignoring the phenotype values from one sex, while retaining all available genotype data to maximize identity-by-descent information. The

Table 2 Number of informative relative pairs, where both members have genotype and phenotype information available

	African-American				European-American				All			
	Female–female	Male–male	Opposite sex pairs	All pairs	Female–female	Male–male	Opposite sex pairs	All pairs	Female–female	Male–male	Opposite sex pairs	All pairs
Sib-pairs	1316	463	1149	2928	1281	855	1861	3997	2597	1318	3010	6925
Half-sibs	384	101	294	779	36	17	53	106	420	118	347	885
Cousins	53	9	20	82	131	98	193	422	184	107	213	504
Parent–child	485	83	409	977	666	324	894	1884	1151	407	1303	2861
Grandparent–grandchild	38	2	34	74	68	25	72	165	106	27	103	239
Avuncular	221	52	210	483	353	291	586	1230	574	343	796	1713

Table 3 Phenotypic characteristics of the sample analyzed in the study

	Females								n	Males								
	Stature (cm)				BMI (kg/m ²)					Stature (cm)				BMI (kg/m ²)				
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	n	
African-American	163.6	6.3	144.8	182	32.2	7.1	16.1	53.5	3161	177	6.6	158	196	28.4	5.2	13.8	43.7	1617
European-American	162.4	6.2	144	180.3	28.7	6.2	15	47.2	2884	176.4	6.5	158	195.6	29.1	4.4	16.6	41.8	2183
All	163.1	6.3	144	182	30.5	6.9	15	53.5	6045	176.7	6.6	158	196	28.8	4.8	13.8	43.7	3800

These figures are based on all available phenotype data.

Table 4 Heritability estimates from the Merlin variance-components analysis

	African-American			European-American			All		
	Females	Males	Combined	Females	Males	Combined	Females	Males	Combined
BMI	0.60	0.66	0.53	0.47	0.65	0.49	0.62	0.57	0.51
Stature	0.97	0.94	0.87	0.92	0.81	0.83	0.95	0.87	0.85

All denotes African- and European-American combined.

covariate-adjusted heritability estimates ranged between 0.49–0.66 for BMI and 0.81–0.97 for stature in the African-American, European-American cohorts and the combined cohort (Table 4). These estimates are in good agreement with estimates from other studies.^{1,2}

The average map density of the multiallelic markers was 6.4 cM and average entropy-based information content was 0.61 in the combined sample. The genome-wide linkage results for African-American, European-American cohorts and the combined cohort are shown in Figure 1, and the chromosomal loci that produced evidence for linkage (multipoint LOD score ≥ 2) are summarized in Table 5. Multipoint LOD scores for chromosomes that showed evidence for linkage (multipoint LOD score ≥ 2) in one or more cohorts are shown in detail in Supplementary Figure 1.

Linkage results for BMI

We found evidence for two separate female-specific loci for BMI: one on 7q35 (LOD=2.93) and another on 11q22 (LOD=2.21). Interestingly, both loci also seem to be population-specific as the 7q35 locus shows linkage only in European-American females and the 11q22 locus only in African-American females. The males provide no evidence for linkage to BMI at these loci.

These LOD scores correspond to the ten-base logarithm of the ratio of the likelihoods of the observed data under the alternative hypothesis (locus is linked to a QTL) and the null hypothesis (locus is not linked to a QTL). Therefore, the interpretation for LOD score of 2.0 is that the likelihood that the locus is linked is 100 times more likely than the likelihood that the locus is not linked to trait. A LOD score of 3.0 is thus evidence that linkage at this locus is 1000 times more likely than no linkage. The LOD scores can also be converted to *P*-values – LOD score of 2.0 corresponds roughly to a *P*-value of 10^{-3} and a LOD score of 3.0 to a *P*-value of 10^{-4} . However, due to the inherent multiple testing performed in genome-wide linkage screens due to the usage of 300–400 genetic markers, usually LOD scores that are considered suggestive and significant are 2.2 and 3.3, respectively.

7q35 has previously been linked to BMI and other body composition-related phenotypes in the Quebec Family Study and in the Old Order Amish population.^{23,24} This region also contains two highly interesting candidate genes for body composition phenotypes: leptin (LEP) and plasminogen activator inhibitor-1 (PAI1). LEP, originally identified from studying the C57BL/6J ob/ob mouse strain,²⁵ plays a key role in the regulation of body weight

by inhibiting food intake and stimulating energy expenditure. In humans, congenital LEP deficiency causes morbid early-onset obesity.²⁵ The PAI1 was previously found to be associated with BMI in a female cohort in the Quebec Family Study,²⁶ which is consistent with our female-specific linkage finding.

The 11q22 region has also been previously linked to BMI in Dutch, Nigerian and Old Order Amish cohorts.^{24,27,28} The linked region contains two highly relevant genes for various body composition-related phenotypes: uncoupling protein 2 (UCP2) and uncoupling protein 3 (UCP3). Both UCP2 and UCP3 have been linked with thermogenesis and energy balance^{26,29} and also possess significant sequence similarity with the uncoupling protein 1 (UCP1), an important mediator of thermogenesis-mediated energy expenditure.

Linkage results for stature

For stature loci on 11q23 (LOD=2.74), 12q12 (LOD=2.07), 15q25 (LOD=3.00), 15q26 (LOD=2.15),

18q23 (LOD=2.49) and 19q13 (LOD=2.18) showed evidence for linkage. The loci on 11q23, 12q12, 15q25 and 18q23 have also been previously reported for linkage to stature, whereas 15q26 and 19q13 have not been linked to stature previously.

The linkage observed at 11q23 is primarily driven by European-American males who clearly show the clearest evidence for linkage. This locus may not, however, be truly sex-specific, as females also show some evidence for linkage, and the total evidence for linkage is stronger when females and males are analyzed jointly. African-American families on the other hand exhibit virtually no evidence for linkage at this locus. Mukhopadhyay *et al*²⁰ also reported linkage finding consistent with our results in this region. Both African-American and European-American cohorts show evidence for linkage on chromosome 12 but peak locations are slightly different. This linkage is also mostly contributed by females; however, inclusion of males adds linkage evidence suggesting lack of

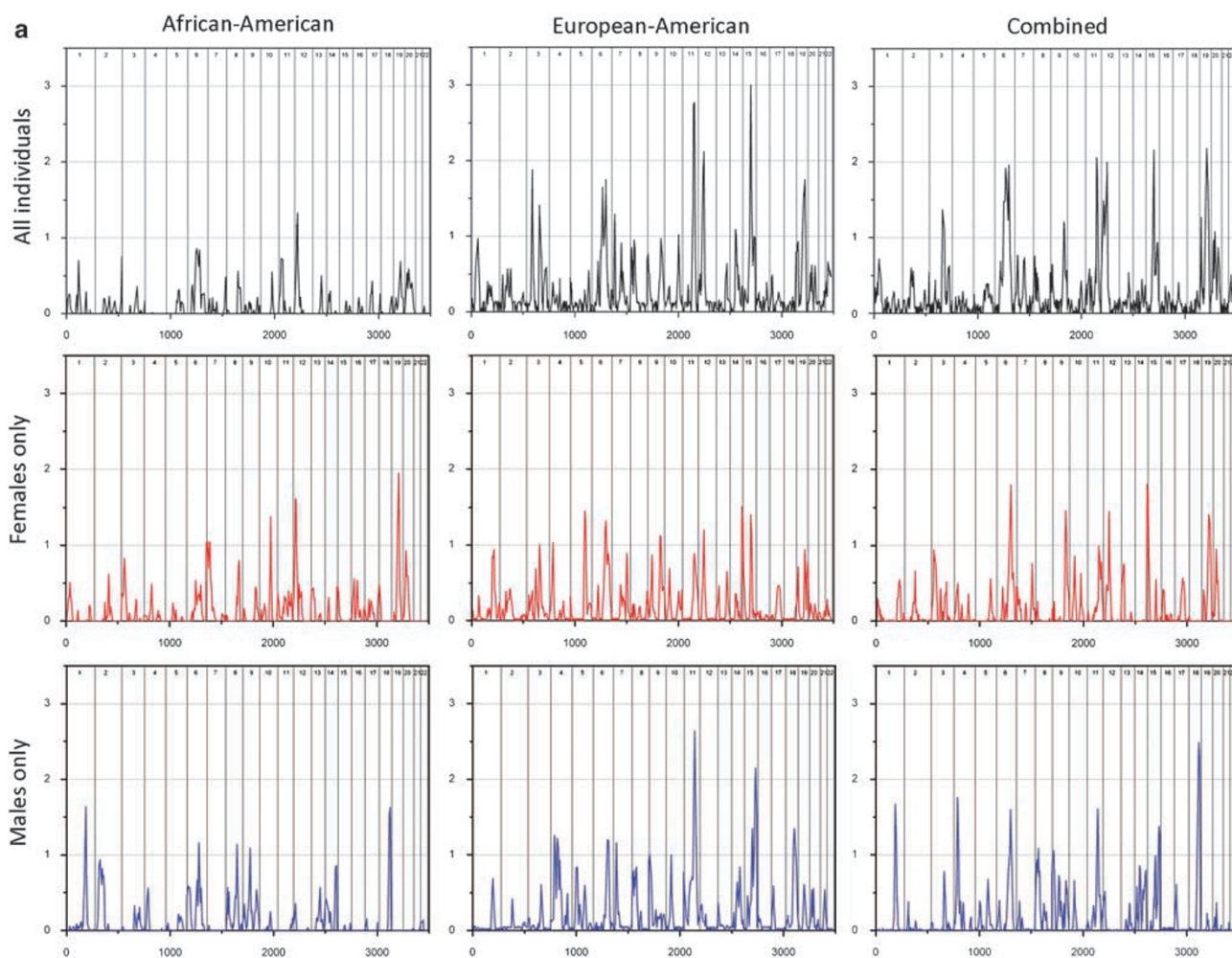


Figure 1 Continued.

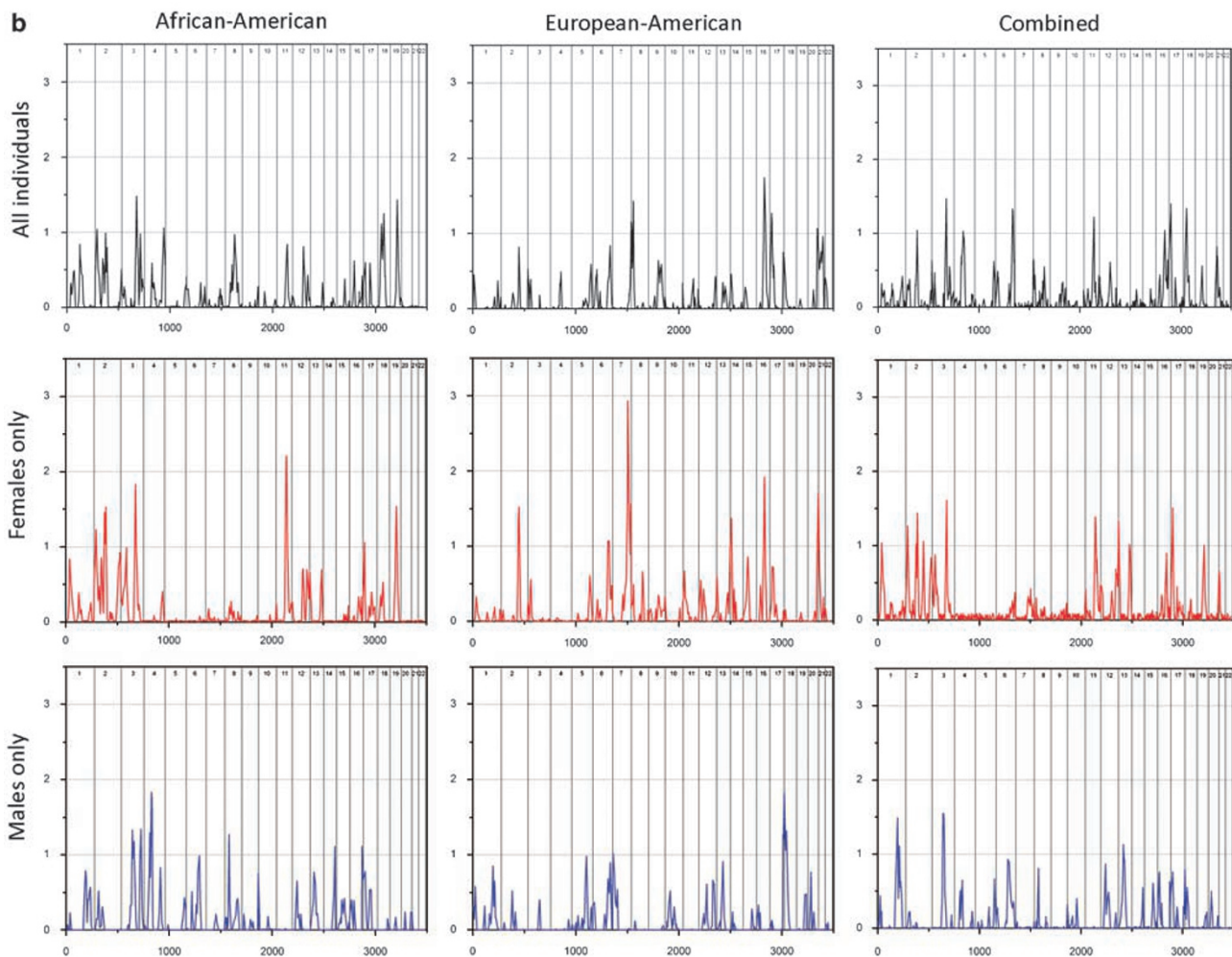


Figure 1 (a) Genome-wide multipoint linkage results for stature with LOD scores on the y axis and cumulative cM on x axis. The left panel displays results for African-American families, the middle panel for European-American and the right panel results for all families combined. Black lines show results for females and males, red lines for females and blue lines for males. (b) Genome-wide multipoint linkage results for body mass index with LOD scores on the y axis and cumulative cM on x axis. The left panel displays results for African-American families, the middle panel for European-American and the right panel results for all families combined. Black lines show results for females and males, red lines for females and blue lines for males.

Table 5 Multipoint LOD scores for loci where evidence for linkage (LOD >2.0) was observed in one or more cohorts

Phenotype	Chr	Peak marker	cM	African-American			European-American			All		
				Females	Males	Combined	Females	Males	Combined	Females	Males	Combined
BMI	7	D7S2195	152	0.02	0.00	0.16	2.93	0.00	0.04	0.42	0.00	0.00
BMI	11	D11S2000	106	2.21	0.00	0.84	0.01	0.00	0.40	1.34	0.00	1.02
Stature	11	D11S1986	110	0.41	0.00	0.09	0.71	2.21	2.74	1.00	1.35	2.06
Stature	12	D12S1301	58	0.49	0.00	0.22	1.19	0.22	2.07	1.45	0.02	2.00
Stature	15	D15S655	92	0.00	0.00	0.00	1.40	1.35	3.00	0.55	0.99	2.16
Stature	15	D15S1515	122	0.00	0.00	0.05	0.12	2.15	1.00	0.00	1.20	0.92
Stature	18	D18S844	113	0.00	1.54	0.17	0.03	1.01	0.05	0.00	2.49	0.30
Stature	19	D19S246	83	1.95	0.00	0.69	0.08	0.37	1.56	1.41	0.14	2.18

The peak LOD score for each locus is shown in bold.

sex-specific genetic effects. Interestingly, this peak overlaps with the best association signal from the first published GWA study for stature³⁰ that reported convincing associa-

tion of the high-mobility group at-hook 2 gene (HMGA2) to stature in the general population. Also, this locus has also been previously reported to contain a QTL for stature

in a Finnish cohort³¹ and contains several other interesting candidate genes for stature, such as SRY-box 5 (SOX5), vitamin D receptor (VDR), collagen type 2 α -1 gene (COL2A1). 15q25 shows linkage only in European-American families and is contributed by females and males. This locus has also been previously reported with linkage to stature 13 cM upstream in an Australian sample (LOD score of 3.43 at 79 cM) by Perola *et al.*³² However, a seemingly distinct male-specific linkage on 15q26 was also observed in the European-American cohort. Combined, these linkage regions contain several relevant genes for stature with aggrecan 1 (AGC1) and insulin-like growth factor I receptor (IGF1R) being the most noteworthy. There is some evidence that mutations in the IGF1R gene, resulting in IGF1 resistance, may underlie some cases of prenatal and postnatal growth failure.³³

The male-specific linkage observed on 18q23 is provided roughly equally by both African-American and European-American cohorts, although the exact peak location is slightly different between the cohorts. Interestingly, this finding overlaps well with the same male-specific linkage observed in our previous joint analysis of four Finnish cohorts¹⁹ and a male-driven linkage reported in the Framingham Heart Study.³⁴ Chromosome 19 has not been previously linked to stature in other genome-wide screens. The linked region on 19q is very large containing hundreds of known genes; however, the most prominent candidates are transforming growth factor- β 1 (TGF β 1), a cluster genes belonging to the insulin-like growth factor family of signaling molecules (IGFL1, IGFL2, IGFL3 and IGFL4).

Discussion

The success of the 'positional cloning' paradigm utilizing family-based linkage designs used in gene identification for Mendelian traits has yet to be paralleled for complex traits, where solid evidence for genetic variants has been extremely difficult to establish. Likely, reasons for these shortcomings are (1) insufficient sample sizes and (2) heterogeneity from population to population with respect to either environmental or genetic factors. In this study, we aim to locate QTL underlying the observed variation in stature and BMI and overcome these hurdles by (1) maximizing the sample size by combining the primary genome-wide genotype data of four independent family-based studies and (2) dissecting genetic and environmental heterogeneity by also performing sex- and population-specific linkage analyses. Using this approach, we have performed, to our knowledge, the largest family-based genome-wide screen for stature and BMI, which is not based on meta-analysis, and report two loci linked to variation in BMI and six loci linked to variation in stature most of which add evidence to previously reported loci for these traits.

The genetic background of the continuous variation observed for stature and BMI may be either oligo- or

polygenic in nature. There is some supporting evidence for major genes underlying stature and BMI in humans,³⁵ although solid proof is still elusive perhaps barring some exceptions such as the HMGA2 gene for stature³⁰ as well as a variant upstream of the INSIG2^{36,37} and FTO gene for BMI. These recent GWA results that have demonstrated QTL explain no more than 1% of the trait variance.^{30,38} Mapping QTL of such small effects by linkage analysis is likely to require unrealistic sample sizes, while GWA studies may succeed. Therefore, it is not surprising that in our linkage study, we do not detect the HMGA2 region associated to stature or the FTO region associated to BMI.

The existence of major genes is critical for successful genetic mapping at least in out bred populations.³⁹ We believe that our results do support Fisher's infinitesimal model, where in the human population the genetic background of stature and BMI is controlled by a large number of genes each having a minute effect on the phenotype. In our opinion, this would at least to some extent explain the relatively modest statistical evidence for linkage observed in this study as well as the lack of consistent findings in other studies as well (see http://www.genomeutwin.org/stature_gene_map.htm and <http://obesitygene.pbr.edu> for overview). Another factor that may have resulted in false-negative findings in this study and previous comparable studies is that many traditional linkage-based genome-wide screens contain a relatively low proportion of inheritance information due to relatively sparse genetic maps (>5 cM) and missing founder genotypes. Regentyping these family samples utilizing high-density single-nucleotide polymorphism map to increase inheritance information has been shown to be a successful strategy in simulation⁴⁰ and empirical studies.⁴¹

Family-based linkage studies are based on examining patterns of allele sharing within individual families and then summing up these results across the study sample. Therefore, they are less liable to allelic heterogeneity than association studies and might identify genes for a trait where differing variants may contribute to the trait variance in different populations. In such circumstances even a dense GWA study might miss the signal, especially if the direction of effect is different for allelic variants. Thus, although there are obvious technological advantages in GWA studies, there is still room for traditional linkage studies⁴ in identification of biologically important genes for traits such as stature and BMI, where allelic heterogeneity across populations is quite probable. Another important issue in study design arises from the fact that family-based linkage mapping and association mapping in unrelated individuals are optimal under very different genetic models, and therefore it is unwise to invest solely on one or the other, as we do not know a priori the genetic architecture of the trait we are interested in. Simplistically speaking linkage studies are geared for relatively rare alleles with large effects within families (that may be of little

effect in the population), whereas association studies are designed to detect common genetic effects that have smaller effects. In the case of rare monogenic disease, multiple rare variants at linked loci (allelic heterogeneity) seem to be the rule not the exception.⁴² For common polygenic disease and quantitative traits this question is still unanswered – there are examples for both common⁴³ and rare alleles,⁴⁴ and theoretical and empirical studies suggest a role for both rare and common variants.⁴⁵ One must also remember that GWA can be performed in family samples, although utilizing unrelated samples is more straight-forward and powerful. However, due to the other beneficial qualities inherent to family samples, investigators already in possession of family samples should use them for GWA studies, as loss of statistical power is relatively small.⁴⁶ Considering these and other examples, it is clear that genome-wide linkage mapping in families and GWA mapping in population cohorts should rather be considered as complimentary, not alternative, strategies in mapping polygenic traits.

Body mass index is a derived variable and thus dependent on both height and weight and used mainly as a means for classifying people as underweight, normal or obese. Although BMI does not provide specific information on physiological intermediate phenotypes, such as basal metabolic rate, its widespread use clinically, ease of measurement, high levels of reproducibility and significant heritability across populations make it an important phenotype for genetic analyses. Interest in this phenotype is exemplified by the Obesity Gene Map (<http://obesitygene.pbrc.edu>), which is currently reporting 169 linkages and 183 associations to BMI from various studies, across all human chromosomes. Although there are probably several true signals among those reported, it is plausible to assume that all findings are not due to the obvious polygenic background of BMI, but many are just reflections of analyzing a variable that signals different biological backgrounds in different ascertainment schemes, populations and cohorts. The modest LOD scores we observed for BMI despite the large study sample likely reflects the heterogeneity of our study populations, and may suggest that there are relatively few common loci with strong effects for BMI across these populations. From the results of our analysis, we conclude that there is marked locus heterogeneity between males and females as well as African-American and European-American cohorts. This is evident from the fact that some loci are linked only in sex-specific (BMI 7q35 and stature loci on 12q and 18q) and/or population-specific (BMI loci on 7q and 11q and stature loci on 11q and 15q25) analyses. The relatively large sample size of even subgroups reduced, but cannot eliminate, the probability that the observed linkages in these stratified analyses were spurious.

For some loci the benefit of combining a large number of families resulted in increased statistical significance; for

example, the African-American and European-American families provide consistent evidence for linkage at the stature loci on 18q and 19q13.

It is well established that in the setting of genome-wide screening for trait loci combining the primary data from independent genome-wide screens is superior to meta-analytic approaches in terms of power to detect loci and reducing sources of variation.⁴⁷ However, combining data across multiple cohorts may also increase genetic and environmental heterogeneity and thus hamper successful locus identification.⁴⁸

Our results show both the benefit of (1) combining data to maximize the sample size as well as (2) minimizing heterogeneity by analyzing subgroups where within-group variation can be reduced. Our results however suggest that the latter may be a more fruitful approach in genetic mapping. This approach is analogous to utilizing special populations such as population isolates to reduce genetic and environmental heterogeneity in gene mapping efforts.

Acknowledgements

This study was supported by the European Commission under the program 'Quality of Life and Management of the Living Resources' of 5th Framework Program no. QLG2-CT-2002-01254 (SS, TH, MP and LP) and by the Center of Excellence in Disease Genetics of the Finnish Academy, Biocentrum Helsinki (LP). Joslin Diabetes study (AD) is supported by NIH Grant DK55523. FBPP is supported by the National Heart, Lung, and Blood Institute (NHLBI). This manuscript has been reviewed by FBPP investigators for scientific content and consistency of data interpretation with previous FBPP publications, and significant comments have been incorporated before submission for publication. The SLE study (JAK, GRB and JBH) is supported by NIH Grants AR062277, AR042460 and AI024717. We thank the study participants of the Cleveland Family Study for generously donating their time. The Cleveland Family Study authors are indebted also to our staff members: Joan Aylor, Kathryn Clark, Heather Rogers, Rawan Nawabit and Jennifer Frame for their dedication to this research. We declare no competing interests. All genotypes analyzed in the study were produced by the NHLBI Mammalian Genotyping Service, Marshfield, WI, USA.

References

- 1 Maes HH, Neale MC, Eaves LJ: Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 1997; 27: 325–351.
- 2 Silventoinen K, Sammalisto S, Perola M *et al*: Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res* 2003; 6: 399–408.
- 3 The Wellcome Trust Case Control Consortium: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; 447: 661–678.
- 4 Clerget-Darpoux F, Elston RC: Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum Hered* 2007; 64: 91–96.
- 5 Kim SH, Ma X, Weremowicz S *et al*: Identification of a locus for maturity-onset diabetes of the young on chromosome 8p23. *Diabetes* 2004; 53: 1375–1384.

- 6 The FBPP Investigators: Multi-center genetic study of hypertension: the Family Blood Pressure Program (FBPP). *Hypertension* 2002; **39**: 3–9.
- 7 Hochberg MC: Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**: 1725.
- 8 Gray-McGuire C, Moser KL, Gaffney PM *et al*: Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. *Am J Hum Genet* 2000; **67**: 1460–1469.
- 9 Moser KL, Neas BR, Salmon JE *et al*: Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc Natl Acad Sci USA* 1998; **95**: 14869–14874.
- 10 Olson JM, Song Y, Dudek DM *et al*: A genome screen of systemic lupus erythematosus using affected-relative-pair linkage analysis with covariates demonstrates genetic heterogeneity. *Genes Immun* 2002; **3** (Suppl 1): S5–S12.
- 11 Palmer LJ, Buxbaum SG, Larkin E *et al*: A whole-genome scan for obstructive sleep apnea and obesity. *Am J Hum Genet* 2003; **72**: 340–350.
- 12 Palmer LJ, Buxbaum SG, Larkin EK *et al*: Whole genome scan for obstructive sleep apnea and obesity in African-American families. *Am J Respir Crit Care Med* 2004; **169**: 1314–1321.
- 13 de Andrade M, Fridley B, Boerwinkle E, Turner S: Diagnostic tools in linkage analysis for quantitative traits. *Genet Epidemiol* 2003; **24**: 302–308.
- 14 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: GRR: graphical representation of relationship errors. *Bioinformatics (Oxford, England)* 2001; **17**: 742–743.
- 15 O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- 16 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 17 Šidák Z: Rectangular confidence regions for the means of multivariate normal distributions. *JASA* 1967; **62**: 626–633.
- 18 Kong A, Gudbjartsson DF, Sainz J *et al*: A high-resolution recombination map of the human genome. *Nat Genet* 2002; **31**: 241–247.
- 19 Sammalisto S, Hiekkalinna T, Suviolahti E *et al*: A male-specific quantitative trait locus on 1p21 controlling human stature. *J Med Genet* 2005; **42**: 932–939.
- 20 Mukhopadhyay N, Finegold DN, Larson MG, Cupples LA, Myers RH, Weeks DE: A genome-wide scan for loci affecting normal adult height in the Framingham Heart Study. *Hum Hered* 2003; **55**: 191–201.
- 21 Hiekkalinna T, Terwilliger JD, Sammalisto S, Peltonen L, Perola M: AUTOGSCAN: powerful tools for automated genome-wide linkage and linkage disequilibrium analysis. *Twin Res Hum Genet* 2005; **8**: 16–21.
- 22 Weiss LA, Pan L, Abney M, Ober C: The sex-specific genetic architecture of quantitative traits in humans. *Nat Genet* 2006; **38**: 218–222.
- 23 Borecki IB, Rice T, Perusse L, Bouchard C, Rao DC: An exploratory investigation of genetic linkage with body composition and fatness phenotypes: the Quebec Family Study. *Obes Res* 1994; **2**: 213–219.
- 24 Platte P, Papanicolaou GJ, Johnston J *et al*: A study of linkage and association of body mass index in the Old Order Amish. *A J Med Genet* 2003; **121**: 71–80.
- 25 Ingalls AM, Dickie MM, Snell GD: Obese, a new mutation in the house mouse. *J Hered* 1950; **41**: 317–318.
- 26 Argyropoulos G, Brown AM, Willi SM *et al*: Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes. *J Clin Invest* 1998; **102**: 1345–1351.
- 27 Adeyemo A, Luke A, Cooper R *et al*: A genome-wide scan for body mass index among Nigerian families. *Obes Res* 2003; **11**: 266–273.
- 28 van Tilburg JH, Sandkuijl LA, Strengman E, Pearson PL, van Haeften TW, Wijmenga C: Variance-component analysis of obesity in type 2 diabetes confirms loci on chromosomes 1q and 11q. *Obes Res* 2003; **11**: 1290–1294.
- 29 Fleury C, Neverova M, Collins S *et al*: Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia. *Nat Genet* 1997; **15**: 269–272.
- 30 Weedon MN, Lettre G, Freathy RM *et al*: A common variant of HMG2 is associated with adult and childhood height in the general population. *Nat Genet* 2007; **39**: 1245–1250.
- 31 Hirschhorn JN, Lindgren CM, Daly MJ *et al*: Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. *Am J Hum Genet* 2001; **69**: 106–116.
- 32 Perola M, Sammalisto S, Hiekkalinna T *et al*: Combined genome scans for body stature in 6602 European twins: evidence for common Caucasian Loci. *PLoS Genet* 2007; **3**: e97.
- 33 Abuzzahab MJ, Schneider A, Goddard A *et al*: IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *New Engl J Med* 2003; **349**: 2211–2222.
- 34 Mukhopadhyay N, Weeks DE: Linkage analysis of adult height with parent-of-origin effects in the Framingham Heart Study. *BMC Genet* 2003; **4** (Suppl 1): S76.
- 35 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**: 307–322.
- 36 Herbert A, Gerry NP, McQueen MB *et al*: A common genetic variant is associated with adult and childhood obesity. *Science (New York, NY)* 2006; **312**: 279–283.
- 37 Lyon HN, Emilsson V, Hinney A *et al*: The association of a SNP upstream of INSIG2 with body mass index is reproduced in several but not all cohorts. *PLoS Genet* 2007; **3**: e61.
- 38 Frayling TM, Timpson NJ, Weedon MN *et al*: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, NY)* 2007; **316**: 889–894.
- 39 Tanksley SD: Mapping polygenes. *Ann Rev Genet* 1993; **27**: 205–233.
- 40 Evans DM, Cardon LR: Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. *Am J Hum Genet* 2004; **75**: 687–692.
- 41 Sawcer SJ, Maranian M, Singlehurst S *et al*: Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Hum Mol Genet* 2004; **13**: 1943–1949.
- 42 Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* 2002; **11**: 2417–2423.
- 43 Maller J, George S, Purcell S *et al*: Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006; **38**: 1055–1059.
- 44 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science (New York, NY)* 2004; **305**: 869–872.
- 45 Di Rienzo A: Population genetics models of common diseases. *Curr Opin Genet Dev* 2006; **16**: 630–636.
- 46 Visscher PM, Andrew T, Nyholt DR: Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. *Eur J Hum Genet* 2008; **16**: 387–390.
- 47 Guerra R, Etzel CJ, Goldstein DR, Sain SR: Meta-analysis by combining *P*-values: simulated linkage studies. *Genet Epidemiol* 1999; **17** (Suppl 1): S605–S609.
- 48 Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; **69**: 936–950.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)