

Joel Wu
MPH Thesis
8/12/2005

Introduction

Human height is a highly variable trait that correlates with many social background variables and with both social and health-related outcomes. Observed associations between height and outcomes such as cardiovascular disease, diabetes and suicide may be due to direct effects of height.^{1-2, 7-14} However the relationships between height and health outcomes are often confounded by associations between height and other social and environmental factors, such as nutrition during development, socioeconomic status, and education.³⁻⁴ Social and environmental factors such as these may independently impact multiple traits including height and other health outcomes. A greater understanding of the determinants of height may help us to unravel the complicated relationships between height and health.

Short stature has been implicated as a cardiovascular risk factor by a number of studies that have observed an association between short stature and increased risk for coronary heart disease and cerebrovascular disease.⁷⁻¹⁴ One of the goals of the National Heart, Lung and Blood Institute (NHLBI) Family Heart Study (FHS) is to investigate the genetic determinants of cardiovascular risk factors.⁶ The goal of this study is to localize genes which contribute to variation in adult height.

Height is a complex trait well suited for genome-wide linkage analysis for several reasons. Compared to most other complex traits, height is easily, reliably and accurately measured. Height is highly heritable. Recent studies have reported heritability values (h^2) ranging from 0.6 to greater than 0.95.^{3, 15-23} A summary of these values is provided in Table 1. Furthermore, two recent papers by Silventoinen et al have shown that in affluent western countries variation in height is largely due to genetic variation, and not environmental influence.^{3,20} Genetic variation in height is probably influenced by a large number of genes, however it is believed that a majority of the variation in height

is due to a few major genes. Recent studies have provided some evidence for such a major gene model. Li et al found evidence suggesting that a major gene accounts for about 17.2% of the total variation in adult height in a Chinese population, and Ginsburg et al found that in 5 ethnically different populations, a major gene may affect up to 53% of the variance in body height and weight.²⁴⁻²⁵ An appropriately designed and conducted genome-wide scan should be able to identify chromosomal regions containing genes that contribute to variation in adult height.

To date, 8 papers have published results of genome scans for adult height.^{16-22,26} Appendix I is an updated reproduction of a summary of these papers published by Willemsen et al. There are several regions of interest that have produced strong evidence of linkage. A number of these chromosome regions have been identified and replicated across different study populations. A summary of these regions is presented in Table 2.

Methods

This study is based on data from the National Heart, Lung and Blood Institute Family Heart Study, a large population based study consisting of families from four different cities in the US. The NHLBI FHS is a multi-centered population-based study examining the genetic and non-genetic causes of coronary heart disease, preclinical atherosclerosis and cardiovascular risk factors¹. The objectives and design of FHS have been published in further detail in Higgins et al.

Study Population

Potential FHS participants were drawn from 3 parent studies: the Framingham Heart Study, Utah Health Family Tree Study, and the Atherosclerosis in Communities Study. Recruitment of FHS participants occurred in 2 phases. In phase I, potential probands were drawn from 4 study sites: Forsyth County, North Carolina; Minneapolis, Minnesota; Framingham, Massachusetts; Salt Lake City, Utah. At each site, approximately 2,000 individuals with a family history of heart disease, and 2,000 random individuals were selected to provide family and personal medical histories as well as the names and

contact information of their relatives. These relatives were then asked to provide their own medical histories. In Phase I, 3,156 probands provided personal and family medical histories, and 22,909 relatives of the probands provided additional personal medical histories. In phase II, 588 random families and 657 families with the highest risk of CHD were asked to participate in clinical examinations and follow-up. During these clinical examinations, anthropometric measures and blood samples were taken. In all 5,381 individuals in 1,255 families participated in the clinical examinations. In this study we have included 3,152 individuals from 535 families. Table 3 is summary of the relative pairs within the sample. Potential subjects were excluded under 2 conditions: If they had no genotype data and no children; If their self-reported race was other than White: Non-Hispanic and they were not in a family with other White: Non-Hispanic Subjects. Use of uninformative subjects introduces noise into the model. Using only one race category eliminates race categories as a possible predictor of height and heart disease. Furthermore, we also avoid dealing with the possibility of differing allele frequencies between racial groups.

Genotyping

Blood samples collected during clinical examinations were sent to the University of Minnesota for isolation of genomic DNA. Genomic DNA samples were then sent to 2 different laboratories for genotyping. There were 2,164 individuals genotyped exclusively at the NHLBI Mammalian Genotyping Service in Marshfield WI, using the Cooperative Human Linkage Center Screening Set 10. This screening set includes 402 autosomal markers that are spaced approximately 9 cM apart. There were 347 individuals genotyped exclusively at the Utah Molecular Genetics Laboratory in Salt Lake City UT, using an independently developed set of 243 autosomal markers spaced approximately 18 cM apart. A total of 846 individuals were genotyped at both locations with both marker sets. There were 1,312 individuals included in the linkage analysis that were not genotyped.

Statistical analysis

Separate histograms and quantile-quantile plots (Appendix II) of the male and female study

participants revealed adult stature is approximately normally distributed in the study sample. Linear regression of the combined study sample revealed that age, age², and gender are predictors of height (Appendix II); therefore subsequent analyses are adjusted for age, age² and gender. In addition to a simple linear regression of the sample data, a second regression using Generalized Estimating Equations (GEE) was performed prior to heritability estimation in order to more accurately represent the clustering of values within families.³¹ The histograms, qqplots and linear regressions were performed in the statistical package “R”. Heritability (h²) was estimated by maximum likelihood in a polygenic model in SOLAR.²⁷ The estimation was performed with age, age² and gender as covariates.

Unlikely genotypes were checked and removed prior to linkage analysis by using the –error and pedwipe commands in MERLIN.²⁸ Quantitative trait loci associated with variation in adult height were identified by performing multipoint linkage analysis in SOLAR.²⁷ The variance component estimation methods implemented in SOLAR require identity-by-descent probabilities (IBD). These probabilities were estimated using 2 different methods. The IBD probabilities of 1077 small families were estimated using a version of the Lander-Green algorithm implemented in MERLIN.²⁸ The IBD probabilities of 7 remaining large families were estimated using Markov chain Monte Carlo methods as implemented in LOKI.²⁹⁻³⁰ IBD matrices calculated in MERLIN and LOKI were converted into a single set of multipoint IBD files in SOLAR format using MER2SOL (<http://taxa.epi.umn.edu/mer2sol>) a program developed by Michael Miller at the University of Minnesota.

Results

The basic characteristics of the study sample are summarized in Tables 4 and 5. In all, there were 1832 women, and 1620 men in the study sample. The final GEE model estimated 60% of the variance in height within the sample is due to three significant covariates: age, age², and sex. Heritability analysis was adjusted according to this model in order to minimize confounding. The residual heritability of height in the study sample is 0.87 (SE=0.023). The complete result of this

analysis is presented in Appendix III.

Appendix IV contains plots of the multipoint linkage results. Strong evidence for linkage was detected on chromosomes 5 and 14. On chromosome 5 a lod score of 3.09 was observed near GX134B03 at 31 cM. On chromosome 14 a lod score of 2.38 was observed near D14S556 at 49 cM. A summary of other points of interest are summarized in Table 6.

Discussion

The locus detected on chromosomes 5 did not replicate the results of any previous linkage studies. The locus detected on chromosome 14 replicates findings of Perola et al, who reported a LOD score of 1.67 at 45.12-47.51 cM. A smaller signal (LOD 1.158) on chromosome 4 at 112 cM replicates findings of Hirschorn et al, who reported a LOD score of 2.28 at 112.63 cM. It is encouraging to find our results replicate the results of genome scans of other populations. This suggests that these loci may play an important role in human height.

However, replication of results between studies and between populations is still rare. Location estimates of quantitative trait loci for complex traits may vary about the true locus for several reasons. Simulations have shown that chance variation introduced in statistical modeling of complex polygenic traits causes variation in locus estimates.³² Furthermore, studies utilizing different statistical methods implemented in different software packages may result in varying estimates. Mukhopadhyay et al observed differing results after analysis of the same dataset with two different, but well established methods for linkage analysis.²² Finally, though stature is a highly heritable trait, it is still a complex trait determined by the effects and interactions of many genetic and environmental components. The high variability in genome scan results may suggest that height is influenced by a large number of genes of small effect. There is still a great deal of inconsistency in the results of published linkage studies, nevertheless each additional study conducted will continue to contribute to a greater understanding of the genetics of stature.

References

- 1 Han TS, Feskens EJM, Lean MEJ, Seidell JC. Associations of body composition with Type 2 diabetes mellitus. *Diabetic Medicine* 1998 15(2):129-135.
- 2 Magnusson PKE, Gunnell D, Tynelius P, Smith GD, Rasmussen F. Strong inverse association between height and suicide in a large cohort of Swedish men: Evidence of early life origins of suicidal behavior? *Am J Psychiatry* 2005 162:1373-1375.
- 3 Silventoinen K, Krueger RF, Bouchard TJ, Kaprio J, McGue M. Heritability of body height and educational attainment in an international context: Comparison of adult twins in Minnesota and Finland. *Am J Human Biology* 16:544-555.
- 4 Sargent JD, Blanchflower DG. Obesity and stature in adolescence and earnings in young adulthood. Analysis of a British birth cohort. 1994 *Arch Pediatrics Adolescent Med* 148(7):681-687.
- 5 Goldberg MJ, Yassir W, Sadeghi-Nejad A. Clinical Analysis of Short Stature. *J Pediatric Orthopaedics* 2002 22:690-696.
- 6 Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, Rao DC, Sprafka JM, Williams R. NHLBI Family Heart Study: Objectives and Design. *Am J Epidemiology* 1996 143(12):1219-1228.
- 7 Lundberg M, Diderichsen F, Hallqvist J. *Scand J Public Health* 2002 30:249-258.
- 8 Silventoinen K, Kaprio J, Koskenvuo M, Lahelma E. The association between body height and coronary heart disease among Finnish twins and singletons. *Int J Epidemiology* 2003 32:78-82.
- 9 Parker DR, Lapane KL, Lasater TM, Carleton RA. Short Stature and cardiovascular disease among men and women from two southeastern New England communities. *Int J Epidemiology* 1998 27:970-975.
- 10 Forsen T, Eriksson J, Qiao Q, Tervahauta M, Nissinen A, Tuomilehto J. Short stature and coronary heart disease: a 35-year follow-up of the Finnish cohorts of The Seven Countries Study. *J Int Med* 2000 248:326-332.
- 11 Smith GD, Hart C, Upton M, Hole D, Gillis C, Watt G, Hawthorne V. Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *J Epidemiology Community Health*. 2000 54:97-103
- 12 Wannamethee SG, Shaper AG, Whincup PH, Walker M. Adult height, stroke and coronary heart disease. *Am J Epidemiology* 1998 148(11):1069-1076.
- 13 Goldbourt U, Tanne D. Body height is associated with decreased long-term stroke but not coronary heart disease mortality? *Stroke* 2002 33:743-748.
- 14 Njolstad I, Arnesen E, Lund-Larsen PG. Body height, cardiovascular risk factors, and risk of stroke in middle-aged men and women. A 14 year follow-up of the Finnmark Study. *Circulation* 1996 94:2877-2882.
- 15 Silventoinen K, Kaprio J, Lahelma E, Koskenvuo M. Relative effect of genetic and environmental factors on body height: Differences across birth cohorts among Finnish men and women. *Am J Public Health* 2000 90:627-630.
- 16 Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burt 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 Human Genetics* 2001 69:106-116.
- 17 Perola M, Ohman M, Hiekkalinna T, Leppavuori 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 Finish study groups. *Am J Human Genetics* 2001 69:117-123.
- 18 Deng HW, Xu FH, Liu YZ, Shen H, Deng H, Huang QY, Liu YJ, Conway T, Li JL, Davies KM, Recker RR. A whole-genome linkage scan suggests several genomic regions potentially containing QTL's underlying the variation in stature. *Am J Medical Genetics*. 2002 113:29-39.
- 19 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 3p26 in a large UK family data set ascertained for Type-2 Diabetes. *Am J Human Genetics* 2002 70:543-546.
- 20 Xu J, Bleeker 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 chromosome 6, 9, and 12. *Am J Human Genetics* 2002 71:646-650.
- 21 Wu X, Cooper RS, Boerwinkle E, Turner ST, Hung S, Myers R, Olshen RA, Curb D, Zhu X, Kan D, Luke A. Combined analysis of genomewide scans for adult height: results from the NHLBI Family Blood Pressure Program. *Eur J Human Genetics* 2003 11:271-274.
- 22 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. *Human Heredity* 2002 55:191-201.

- 23 Silventoinen K, Sammalisto S, Perola M, Boomsma D, Cornes B, Davis C, Dunkel L, de Lange M, Harris JR, Hjelmborg JVB, Luciano M, Martin NG, Mortensen J, Nistico L, Pedersen NL, Skytthe A, Spector TD, Stazi MA, Willemssen G, Kaprio J. Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research* 2003 6(5):399-408.
- 24 Li MX, Liu PY, Li YM, Qin YJ, Liu YZ, Deng HW. A major gene model of adult height is suggested in Chinese. *J Hum Genetics* 2004 49:148-153.
- 25 Ginsburg E, Livshits G, Yakovenko K, Kobylankys E. Major gene control of human body height, weight and BMI in five ethnically different populations. *Ann Hum Gen* 1998 62:307-322.
- 26 Willemsen G, Boomsma D, Beem L, Vink JM, Slagboom PE, Posthuma D. QTL's for height: Results of a full genome scan in Dutch sibling pairs. *Eur J Human Genetics* 2004 12:820-828.
- 27 Almasy L, Blangero J. Multipoint Quantitative-Trait Linkage Analysis in general pedigrees. *Am J Human Genetics* 1998 62:1198-1211.
- 28 Abecasis GR, Cherny SS, Cookston WO, Cardon LR. Merlin- rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* 2002 30:97-101.
- 29 Heath SC. Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Human Genetics* 1997 61:748-760.
- 30 Heath SC, Snow GL, Thompson EA, Tseng C, Wijsman EM. MCMC segregation and linkage analysis. *1997 Genetic Epidemiology* 14:1011-1015.
- 31 Hardin JW, Hilbe JM. *Generalized Estimating Equations*. Chapman & Hall/CRC; 1st edition (July 30, 2002).
- 32 Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: An examination of variation in location estimates. *Am J Hum Gen* 1999 65:876-884.