

Association of *HMGA2* Gene Variation with Height in Specific Pediatric Age Categories

Struan F.A. Grant^{1,2,3}, Mingyao Li⁴, Jonathan P. Bradfield¹, Cecilia E. Kim¹, Kiran Annaiah¹, Erin Santa¹, Joseph T. Glessner¹, Tracy Casalunovo¹, Edward C. Frackelton¹, F. George Otieno¹, Julie L. Shaner¹, Ryan M. Smith¹, Andrew W. Eckert¹, Marcin Imielinski¹, Rosetta M. Chiavacci¹, Robert I. Berkowitz^{5,6} and Hakon Hakonarson^{1,2,3}

¹Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, U.S.A. ²Department of Pediatrics and Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, U.S.A. ³Department of Pediatrics University of Pennsylvania, School of Medicine, Philadelphia, Pennsylvania 19104, U.S.A. ⁴Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A. ⁵Behavioral Health Center and Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia PA 19104, U.S.A. ⁶Center for Weight and Eating Disorders, Department of Psychiatry, University of Pennsylvania, Philadelphia PA 19104, U.S.A.

Background: Recently an association was demonstrated between the single nucleotide polymorphism (SNP), rs1042725, within the *HMGA2* locus and height as a consequence of a genome wide association (GWA) study of this trait in adults; this observation was also reported in children aged 7–11 years old.

Objective: We examined in our Caucasian childhood cohort the effects of two strong surrogates for this SNP at this locus with height, rs8756 and rs7968902, with respect to the same pediatric age category but also in children grouped separately as younger and older.

Methods: Utilizing data from an ongoing GWA study in our cohort of 2,619 Caucasian children with measurements for height, we investigated the association of the previously reported variation at the *HMGA2* locus with this height treated as a quantitative trait (age and sex corrected) in childhood in the 2–6 (n = 706), 7–11 (n = 617) and 12–18 (n = 1293) years old categories.

Results: The minor alleles of rs8756 and rs7968902 respectively (strong surrogates for rs1042725 i.e. $r^2 = 0.873$ and 0.761 in the CEU HapMap respectively) were significantly associated with height in the 7–11 years old age group ($P = 3.53 \times 10^{-3}$ and 2.82×10^{-4} , respectively). However in the 2–6 and 12–18 years old age groups, no association was observed.

Conclusions: We observe a strong association with height in same age group of 7–11 years old as has been previously reported. However, in the under 7s and the over 11s, no such association was observed.

Keywords: *HMGA2*, gene, association, children, height

Height in humans has always been considered highly heritable and has also been correlated with certain disorders, such as cancer (Davey Smith, Hart et al. 2000; Gunnell, Okasha et al. 2001). Twin and family studies have suggested that may be as much as 90% of variation in human height is due to genetic factors (Preece, 1996; Silventoinen, Kaprio et al. 2000; Silventoinen, Kaprio et al. 2001; Macgregor, Cornes et al. 2006; Perola, Sammalisto et al. 2007).

Weedon et al. (Weedon, Lettre et al. 2007) reported an analysis of this phenotype in the context of genome-wide data generated using the Affymetrix GeneChip Human Mapping 500 K platform on nearly 5000 individuals of self-reported European ancestry, which included approximately 2000 U.K., individuals from the Wellcome Trust Case Control Consortium (Wellcome Trust Case Control Consortium 2007) and approximately 3,000 Scandinavian participants from the Diabetes Genetics Initiative (Saxena, Voight et al. 2007).

As a consequence, they observed association to common variation in the mobility group-A2 (*HMGA2*) oncogene. Follow-up analyses in approximately 19,000 more individuals (both adults and children aged

Correspondence: Struan F.A. Grant, Email: grants@chop.edu
Hakon Hakonarson, Email: hakonarson@chop.edu



Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: <http://creativecommons.org/licenses/by/3.0/>.

7 and 11 years old) revealed strong replication of this observation, yielding a $P = 4 \times 10^{-16}$ when all data was combined.

This gene has already been implicated to play a role in height as rare but severe mutations in *HMGA2* are known to impact on body size in both mice and humans (Palmert and Hirschhorn, 2003). They estimated that the key SNP, rs1042725, explains approximately 0.3% of population variation in height (or about 0.4cm increase in adult height per C allele).

Although these *HMGA2* findings are compelling, there are continuing concerns regarding the performance of association studies in complex traits. As such, independent replication efforts are now considered mandatory (Patterson and Cardon, 2005). With the many errors and biases that can blight any individual study, replication by others can ensure that the original findings are robust and can also provide a more accurate estimate of the likely effect size (Hirschhorn, Lohmueller et al. 2002; Page, George et al. 2003).

We have an ongoing genome wide association (GWA) study on our pediatric cohort from Philadelphia, which includes heights measured for all participants. However, we are not using the Affymetrix arrays employed by Weedon et al. (Weedon, Lettre et al. 2007) but rather the competing Illumina Infinium™ II HumanHap550 BeadChip technology. rs1042725 is not present on this platform but two strong surrogates for the same variant are included in the design, namely rs8756 and rs7968902 (r^2 to rs1042725 = 0.873 and 0.761 respectively in the CEU HapMap). In this study we demonstrate that these surrogates associate strongly with height, when treated as a quantitative trait, in our pediatric cohort, but only in the same age category of 7–11 years old (as previously reported); however this association is absent in the 2–6 and 12–18 years old age categories.

Materials and Methods

Study subjects

All subjects were consecutively recruited from the Greater Philadelphia area from 2006 to 2007 at the Children's Hospital of Philadelphia. Our study consisted of 2,619 Caucasian children of European descent. All subjects were biologically unrelated and were aged between 2 and 18 years old. The number of individuals in the 2–6, 7–11 and

12–18 years old age groups were 709, 617 and 1,293 respectively, with average heights (in centimeters) of 105.07 ± 12.27 , 135.07 ± 12.72 and 162.21 ± 12.09 . All subjects were between -3 and $+3$ standard deviations with respect CDC corrected BMI i.e. outliers were excluded to avoid the consequences of potential measurement error. This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia.

Genotyping

We performed high throughput genome-wide SNP genotyping using the Illumina Infinium™ II HumanHap550 BeadChip technology (Gunderson, Steemers et al. 2005; Steemers, Chang et al. 2006) in the same manner as our center has reported previously (Hakonarson, Grant et al. 2007). The resources available for this project included the Illumina technology platform itself plus nine Tecan pipetting robotic systems, eight scanners, a laboratory information management system (LIMS) and automated allele-calling software. The workflow was robotic-based for automatic sample processing and included algorithms for quality control of genotypes. The facility infrastructure had sufficient computational power and servers for data processing and storing, including a series of computers that were integrated (warehouse setting) to perform continuous datamining of all gathered and generated datasets.

Analysis

SNP rs1042725 was not included on the Illumina 550 K BeadChip. However, we searched the HapMap database, and found that two SNPs, which are present on our 550 K BeadChip, rs8756 and rs7968902 respectively were strong surrogates in the CEU HapMap ($r^2 = 0.873$ and 0.761 respectively). Therefore, we queried the data with a test for these two SNPs to investigate if they were associated with height (age and sex corrected) in our pediatric cohort in the pre-specified age categories.

By treating height as a quantitative trait, association analysis for each SNP was carried out using linear regression with the SNP included as an independent variable (coded as 0, 1, and 2). Both SNPs were in Hardy-Weinberg equilibrium.

All statistical analyses were carried out using the software package *plink* (<http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>) (Purcell, Neale et al. 2007).

Results

Using quantitative trait analysis, we observed significant association between both rs8756 and rs7968902 (strong surrogates for rs1042725 i.e. $r^2=0.873$ and 0.761 in the CEU HapMap respectively) and height, but only in the previously age category of 7–11 years old (Table 1). The minor alleles of rs8756 and rs7968902 respectively were significantly associated with height in the 7–11 years old age group ($P=3.53 \times 10^{-3}$ and 2.82×10^{-4} , respectively). However in the 2–6 and 12–18 years old age groups, no association was observed (Table 1).

Discussion

From an interim analysis of our ongoing GWA study of height in children, we have investigated, with respect to specific age categories, variation in the *HMGA2* locus previously reported to be associated with height in both children and adults (Weedon, Lettre et al. 2007). Consequently, we have replicated the association of this gene with height by further demonstrating its effect in the childhood form of the phenotype, but only for the previously reported 7–11 years old age category. More specifically, the common non-coding variants, rs8756 and rs7968902 respectively (strong surrogates for the previously reported rs1042725), were shown to associate with height in the 7–11 year olds but not for the age groupings below and above this category.

As the association we observe is indeed of a very similar magnitude to that of the original report (Weedon, Lettre et al. 2007), this independent replication confirms *HMGA2* as a genuine childhood height gene and further refines where the association is most strongest with respect to age. The age categories were selected based on a mixture of what had been previously described and what represented sufficiently powered cohorts. As Weedon et al. (Weedon, Lettre et al. 2007) had focused on the 7–11 age group, we defined our grouping around this previous definition, such that we grouped all children in this age category and then separately grouped those children that were either older or younger than this category; as such, we had similar numbers in each of the age categories. Note that we did not analyze each distinct age in turn as we were insufficiently powered to do so.

The association may be clouded by extreme growth during early development and then later in puberty and may explain why we do not observe association in the 2–6 and 12–18 years old age categories. Another possible explanation is that age and sex correction is insufficient and that other, yet to be determined, co-factors would need to be incorporated in to the model.

Our results lend further support for the role of the *HMGA2* gene in height determination. The variants that we observe association to may directly dictate splicing or some other regulatory

Table 1. Quantitative analysis of height results for the minor alleles of the *HMGA2* markers strongly associated with Weedon et al. (Weedon, Lettre et al. 2007) in Caucasians, by pre-specified age group.

CHR	SNP	B35 location	NMISS	BETA	SE	R2	T	P
2–6 yrs								
12	rs8756	64646019	612	−0.2362	0.566	0.000285	−0.4173	0.677
12	rs7968902	64649337	617	−0.198	0.5746	0.000193	−0.3446	0.731
7–11 yrs								
12	rs8756	64646019	705	1.623	0.5545	0.01205	2.928	3.53×10^{-3}
12	rs7968902	64649337	709	2.01	0.5508	0.01849	3.649	2.82×10^{-4}
12–18 yrs								
12	rs8756	64646019	1282	0.4977	0.4147	0.001124	1.2	0.230
12	rs7968902	64649337	1293	0.3843	0.4248	0.000634	0.9047	0.366

NMISS, number of individuals tested; BETA, regression coefficient for the test SNP; SE, standard error of the regression coefficient; R2, r^2 value in linear regression; T, test statistic; P, P-value (additive model).

mechanism, but more likely are in linkage disequilibrium with the causative variant(s).

Once our GWA study is complete, we will have the opportunity to look for other variants in the genome that are associated with height in children, as a consequence of our use of a high resolution BeadChip. In addition, we will explore the *HMGA2* gene further to elucidate other potential variants that may contribute to height determination in our cohort and to investigate the influence by each specific age rather than the categories employed in this current study.

Acknowledgements

We would like to thank Adrienne Alexander, Chioma Onyiah, Elvira Dabaghyan, Kenya Fain, Maria Garris, Wendy Glaberson, Kisha Harden, Andrew Hill, Crystal Johnson-Honesty, Lynn McCleery, Robert Skraban, Kelly Thomas and Alexandria Thomas for their expert assistance with genotyping or data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnsson of Raförnninn ehf for their extensive software design and contribution. This research was financially supported by the Children's Hospital of Philadelphia and a Developmental Research Award from the Cotswold Foundation.

Abbreviations Used in this Paper

SNP, single nucleotide polymorphism; *HMGA2*, high mobility group-A2 gene.

Author Contributions

S.F.A.G. and H.H. designed the study and supervised the data analysis and interpretation. S.F.A.G., M.L. and J.P.B. conducted the statistical analyses. C.E.K., T.C., E.C.F., F.G.O., J.L.S., R.M.S. and A.W.E. directed lab procedures. J.P.B., K.A., E.S., J.T.G. M.I. and A.W.E. provided bioinformatics support. S.F.A.G., R.M.C., R.I.B. and H.H.

coordinated cohort recruitment. S.F.A.G., M.L. and H.H. drafted the manuscript.

Disclosure

The authors report no conflicts of interest.

References

- Davey Smith, G., Hart, C. et al. 2000. Height and risk of death among men and women: aetiological implications of associations with cardiorespiratory disease and cancer mortality. *J. Epidemiol. Community Health*, 54(2):97–103.
- Gunderson, K.L., Steemers, F.J. et al. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.*, 37(5):549–54.
- Gunnell, D., Okasha, M. et al. 2001. Height, leg length, and cancer risk: a systematic review. *Epidemiol. Rev.*, 23(2):313–42.
- Hakonarson, H., Grant, S.F.A. et al. 2007. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*, 448(7153):591–4.
- Hirschhorn, J.N., Lohmueller, K. et al. 2002. A comprehensive review of genetic association studies. *Genet. Med.*, 4(2):45–61.
- Macgregor, S., Cornes, B.K. et al. 2006. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.*, 120(4):571–80.
- Page, G.P., George, V. et al. 2003. Are we there yet?: Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits. *Am. J. Hum. Genet.*, 73(4):711–9.
- Palmert, M.R. and Hirschhorn, J.N. 2003. Genetic approaches to stature, pubertal timing, and other complex traits. *Mol. Genet. Metab.*, 80(1–2):1–10.
- Patterson, M. and Cardon, L. 2005. Replication publication. *PLoS Biol.*, 3(9):e327.
- Perola, M., Sammalisto, S. et al. 2007. Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet.*, 3(6):e97.
- Preece, M.A. 1996. The genetic contribution to stature. *Horm. Res.*, 45(Suppl 2):56–8.
- Purcell, S., Neale, B. et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81(3):559–75.
- Saxena, R., Voight, B.F. et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, 316(5829):1331–6.
- Silventoinen, K., Kaprio, J. et al. 2000. Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *Am. J. Public Health*, 90(4):627–30.
- Silventoinen, K., Kaprio, J. et al. 2001. Sex differences in genetic and environmental factors contributing to body-height. *Twin Res.*, 4(1):25–9.
- Steemers, F.J., Chang, W. et al. 2006. Whole-genome genotyping with the single-base extension assay. *Nat. Methods*, 3(1):31–3.
- Weedon, M.N., Lettre, G. et al. 2007. A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat. Genet.*, 39(10):1245–50.
- Wellcome Trust Case Control Consortium, . 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–78.