

## A Genome-Wide Linkage Scan for Age at Menarche in Three Populations of European Descent

Carl A. Anderson,\* Gu Zhu,\* Mario Falchi, Stéphanie M. van den Berg, Susan A. Treloar, Timothy D. Spector, Nicholas G. Martin, Dorret I. Boomsma, Peter M. Visscher, and Grant W. Montgomery

Genetic Epidemiology and Molecular Epidemiology Laboratories (C.A.A., G.Z., S.A.T., N.G.M., P.M.V., G.W.M.), Queensland Institute of Medical Research, Brisbane, Queensland, Australia 4029; Twin Research and Genetic Epidemiology Unit (M.F., T.D.S.), St. Thomas' Hospital, Kings College, London, United Kingdom SE1 7EH; Department of Biological Psychology (S.M.v.d.B., D.I.B.), Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; Department of Veterinary Medicine (S.M.v.d.B.), Utrecht University, 3508 GA Utrecht, The Netherlands; Institute of Evolutionary Biology (C.A.A.), School of Biological Sciences, University of Edinburgh, Edinburgh, Scotland EH9 3JT, United Kingdom; and Bioinformatics and Statistical Genetics (C.A.A.), Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom OX3 7BN

**Context:** Age at menarche (AAM) is an important trait both biologically and socially, a clearly defined event in female pubertal development, and has been associated with many clinically significant phenotypes.

**Objective:** The objective of the study was to identify genetic loci influencing variation in AAM in large population-based samples from three countries.

**Design/Participants:** Recalled AAM data were collected from 13,697 individuals and 4,899 pseudo-independent sister-pairs from three different populations (Australia, The Netherlands, and the United Kingdom) by mailed questionnaire or interview. Genome-wide variance components linkage analysis was implemented on each sample individually and in combination.

**Results:** The mean,  $SD$ , and heritability of AAM across the three samples was 13.1 yr, 1.5 yr, and 0.69, respectively. No loci were detected that reached genome-wide significance in the combined analysis, but a suggestive locus was detected on chromosome 12 (logarithm of the odds = 2.0). Three loci of suggestive significance were seen in the U.K. sample on chromosomes 1, 4, and 18 (logarithm of the odds = 2.4, 2.2 and 3.2, respectively).

**Conclusions:** There was no evidence for common highly penetrant variants influencing AAM. Linkage and association suggest that one trait locus for AAM is located on chromosome 12, but further studies are required to replicate these results. (*J Clin Endocrinol Metab* 93: 3965–3970, 2008)

Age at menarche (AAM) is a clearly defined event in the course of female pubertal development. The onset of menses is a milestone with important biological and social implications and the age at onset has been identified as a risk factor for several traits. Early AAM is a significant risk factor for depression (1), eating disorders (2), and breast cancer (3). Late AAM has been associated with osteoporotic fractures (4) and a reduced risk of coronary heart disease (5). Therefore, the identification of

loci contributing to variation in AAM could lead to a better understanding of a wide range of phenotypes.

A secular decrease in AAM has occurred in developed countries over the past century. This decrease is believed to be associated with improved nutritional status, greater amounts of adipose tissue, and the improved general health of adolescent females. A large multinational study carried out by the World Health Organization showed the median AAM to be 14 yr (6). In populations of Euro-

0021-972X/08/\$15.00/0

Printed in U.S.A.

Copyright © 2008 by The Endocrine Society

doi: 10.1210/jc.2007-2568 Received November 20, 2007. Accepted July 11, 2008.

First Published Online July 22, 2008

\* C.A.A. and G.Z. contributed equally to this work.

Abbreviations: AAM, Age at menarche; IBD, identical by descent; LOD, logarithm of the odds; pisp, pseudo-independent sibling pair; QTL, quantitative trait locus; SNP, single-nucleotide polymorphism.

pean descent, the mean AAM is typically about 13 yr (except for Mediterranean populations, who have an early AAM) (7).

Menarche is reached after a series of complex developmental and neuroendocrine events leading to full activation of the hypothalamic-pituitary-gonadal axis including maturation of the *KISS1/KISS1R* system (8). AAM is a complex trait determined by an array of genetic and environmental factors (9–11). Genetic factors clearly play a role in AAM, with monozygotic twin correlations in the range of 0.65–0.93 and dizygotic twin correlations in the range of 0.18–0.62 (11). AAM is a highly heritable trait, with twin studies reporting heritabilities of 0.45–0.95 (10).

To date, two genome-wide linkage scans for genes underlying variation in AAM have been performed (12, 13). Using 98 sister pairs, three suggestive linkage peaks were found on chromosomes 16q21, 16q12, and 8p12 for weight-adjusted AAM (13). A study of 2461 sister pairs found significant evidence of linkage to 22q13 and suggestive evidence of linkage to 11q23 and 22q11 (12). No replicated linkages have been reported for AAM.

Several candidate genes have been associated with AAM, including the estrogen receptor- $\alpha$  (*ESR1*) gene (14), the estrogen-biosynthetic gene aromatase *CYP19A1* (15), and the *SHBG* gene (16). The estrogen receptor- $\beta$  (*ESR2*) gene has also been associated with AAM and has been shown to interact with the *ESR1* locus (17). These associations are yet to be replicated in independent samples.

We performed a genome-wide linkage scan in large population-based samples from three populations of European descent. Our combined sample size is the largest reported study for AAM.

## Subjects and Methods

The AAM data in the present study are drawn from population based samples from Australia, The Netherlands, and the United Kingdom.

### Australian samples

The Australian samples were drawn from three different studies.

### Adolescent twin families

Adolescent twins and their families were recruited for an ongoing study of melanoma risk factors. Twins were interviewed at ages 12 and 14 yr. Nontwin siblings were asked to attend at least one interview. As part of the clinical protocol female adolescents were asked during interview to provide their age at menarche (18). The present analysis used AAM estimates collected between May 1992 and February 2006. A second sample of adolescent twin families was recruited to an ongoing study of cognitive ability. Twins and their siblings were interviewed at age 16 yr. As part of the clinical protocol, female participants were asked by a research nurse to give their AAM (19). This current study uses AAM data collected between July 1996 and February 2006. Many of the twins taking part in the cognitive ability study had previously taken part in the melanoma risk factor study. However, 226 of the 1351 adolescents had a censored AAM (*i.e.* menarche had not occurred by the last interview) and were removed from further analysis. This conservative approach will cause a small reduction in power to detect linkage but not increase the false-positive rate. The mean age at last interview for these censored individuals was 12.4 yr, well below the mean age at menarche in the Australia sample (13.0 yr). Therefore, the censored data include individuals from across the range in AAM and not just the upper end of this distribution. Furthermore, because these individuals form only a small part of the total study population (1.68%) and statistical power is pro-

portional to the total number of sibling pairs, their removal is unlikely to have a large influence on the linkage results. Data were also discarded when the reported AAM was higher than the age at interview or when participants indicated they had not yet experienced menarche but nevertheless reported an AAM. In total, 17.6% of the adolescent data were removed (16.7% due to censoring). When AAM was ascertained from the same individual more than once, the correlation between the estimates was 0.9. Three genome-wide scans using microsatellite markers were carried out on extracted DNA from blood samples collected from these families. Details of the genotyping and genetic data cleaning procedures have been described previously (20). In addition, 166 adolescent females were genotyped with 100k single-nucleotide polymorphism (SNP) chips (Affymetrix, Santa Clara, CA). Genetic data from all four genome screens are used in the present study. The adolescent samples provided 826 individuals with both AAM and genotypic data to the present analysis.

### Adult twin families

From 1980 to 1982, AAM information was collected from both members of 1888 female twin pairs as part of a health survey mailed to twins on the Australian Twin Registry. The cohort comprised twins born between 1913 and 1964. Females were asked the question, “How old were you when you had your first menstrual period?” (years and months) via questionnaire. From November 1989 onward, the first-degree relatives of this twin cohort were also surveyed and asked the same AAM question. AAM data from this cohort of female relatives are included in the present analysis. A second cohort of adult twins, born between 1964 and 1971, was recruited in 1989 and hence were aged 18–24 yr when they were initially surveyed by mailed questionnaire. Immediate family members were also asked to participate. As part of the questionnaire, participants were asked the same question regarding menarche as the 1913–1964 cohort. This second cohort were followed up, via telephone interview, between 1996 and 2000 as part of a study into alcoholism and psychiatric morbidity (21) and were again asked to give their AAM. Four microsatellite genome-wide scans have been completed on DNA extracted from blood or buccal samples collected from the two adult cohorts. The genotyping and genetic data cleaning procedures for all four scans have been described previously (22). The adult cohorts provided 3544 individuals with both phenotypic and genotypic data to the present study.

### Endometriosis families

AAM information was obtained from women diagnosed with endometriosis ascertained from the Australian component of the International Endogene Study (23). From 1995 to 2002, the Australian group recruited 931 families, each with at least two affected members (mostly affected sister pairs) with surgically diagnosed endometriosis for a genome-wide linkage scan (24). Included in the sample were case-parent trios and some cases without parents (23). All affected female family members were asked the question, “How old were you when your periods began” (whole years) via mailed questionnaire. Details of the genotyping and genetic data cleaning procedures for the endometriosis cohort have been detailed elsewhere (24). A single 400-marker genome-wide scan was carried out on the Australian families. The final batch of 79 families was genotyped using only 113 markers across chromosomes 9, 10, 11, 19, 20, 21, 22, and X.

For all study cohorts, where two or more AAM estimates were available for an individual, the estimate provided at the first data collection after menarche was used. It was assumed that the recall closest in time to menarche would be the most accurate. In total, the Australian sample comprised 6892 individuals with AAM and genotype data.

### Dutch samples

In total, the Dutch sample comprises 1549 individuals with AAM and genotype data.

Phenotype data were collected from twins and their family members who are registered with The Netherlands Twin Registry (25). In 1991, the The Netherlands Twin Registry started a longitudinal survey study of health, personality, and lifestyle. Surveys were sent out in 1991, 1993,

1995, 1997, 2000, and 2002 to adolescent and adult twins and their family members. Twin pairs could participate in all waves; siblings were included from 1995; parents of twins participated in 1991, 1993, 1995, and 2002; and spouses from 2000. Families with adolescent and young adult twins were recruited through city councils in 1990–1993. After 1993 an effort was also made to recruit adult and older twins using several approaches. Further details on response rates and demographic characteristics of the sample can be found elsewhere (26).

In 1991, 1993, and 1995, female participants were asked to indicate the AAM in years and months, whereas in 2000 and 2002, participants were asked to indicate age in years. In 1993, 1995, and 2002, participants were first asked whether they had had their first menstruation, and if so, they were asked to indicate the age at which it had occurred. All data concerning AAM were rescaled to number of months, in which 6 months were added to the data from 2000 and 2002. This way we avoided bias due to the fact that we only had data on the age in years (27). Data were discarded when the reported AAM was higher than the age at time of the questionnaire and when an AAM report differed by more than 12 months from others reported for the same individual. Data points were also excluded when participants indicated they had not yet had their first menses but nevertheless reported an age. Total discarded data were less than 1%.

DNA was extracted from either whole-blood or buccal swabs using standard protocols (28). Samples were genotyped by the Mammalian Genotyping Service in Marshfield and the Molecular Epidemiology Section, Leiden University Medical Centre. Genotype data from these screens were combined. Pedigree relations were checked with graphic representation of relationships. Errors of Mendelian inheritance were detected with Pedstats. Markers and samples were removed if their total error rate was more than 1%; in all other cases the specific erroneous genotypes were set as unknown. Unlikely recombinants were detected using Merlin (29) and erroneous genotypes were removed with Pedwipe. The mean number of markers was 343 and 717 sibling pairs that had at least 200 autosomal markers genotyped for each individual were selected. In the Dutch families, 352 parents were genotyped.

### United Kingdom samples

Twins identified from the St. Thomas' United Kingdom adult twin registry (TwinsUK) were invited to participate in a study of common diseases and traits (30). Study subjects were assessed for an extensive range of clinical phenotypes. General medical and lifestyle data were also obtained by questionnaire. Genome scans were performed using DNA from leukocytes. Gemini Genomics conducted the genotyping, with 365 core microsatellites and 372 markers providing gap fills, as described previously (31), and 1494 SNP markers from the HuSNP GeneChip linkage mapping set (Affymetrix), providing approximate intermarker spacing of less than 10 cM. Twin zygosity and family relationships were rigorously investigated and discrepant pairs discarded from further analyses. The estimated genotyping error rate was less than 1%.

In total, the U.K. sample comprises 5181 females in dizygotic twin families with records on AAM and genotype data.

### Genetic marker maps

To analyze the Australian, Dutch, and U.K. data jointly, a combined marker map was produced from the marker maps supplied by the three research groups. The Dutch and U.K. markers were placed onto the Australian map using a LOESS regression method implemented through R (32). Where the same marker appeared in multiple scans, the multiple versions were included in the analysis by offsetting the markers by 0.001cM on the marker map. This method negates the need for the rebinning of alleles across multiple genomic screens. Samples from Australia and The Netherlands show a very low level of population differentiation ( $F_{ST} = 0.003$ ), indicating that our combined analysis is appropriate (33). Furthermore, when we contrasted the allele frequencies of 317,000 SNPs between a sample of 462 female Australians and the HapMap sample of 60 individuals of northwestern European descent, we found no evidence for stratification (results not shown).

### Data cleaning

To reduce the influence of univariate outliers, all estimates which differed from the mean by more than 3 SD were regressed back to this threshold. Phenotypic bivariate outliers were identified and removed to ensure that linkage signals were not overly influenced by extremely discordant sibling pairs (see Appendix 1). A bivariate outlier is defined as a pseudo-independent sibling pair (pisp) with an extreme difference in AAM. Any pisp with a squared difference in AAM of more than 2878 months<sup>2</sup> was identified as a bivariate outlier. This is equivalent to an absolute difference of 4.47 yr between the AAM of pisp. For each pisp as a bivariate outlier, the AAM that differed most from the sample mean was removed from further analysis. In total, 63 individuals were removed from the study because they created a bivariate outlier. The AAM data were then standardized to a normal distribution with a mean of 0 and a SD of 1. After the removal of bivariate outliers, the sample consisted of 13,618 individuals and 4,899 pisp (2893 from Australia, 1289 from the United Kingdom, and 717 from The Netherlands) with AAM and genome-wide genetic marker.

### Statistical analysis

Genome-wide variance-component linkage analysis was carried out using MERLIN (29). MERLIN estimates the probability of the number of alleles shared (0, 1, or 2) identical by descent (IBD) among relatives in a pedigree. The resulting IBD coefficient that is used in the analysis is the estimate of the proportion of alleles shared IBD. Identity by descent coefficients were calculated at 5-cM intervals. Variance-component models are based on the correlation between the genetic similarity of relatives at a given locus (IBD coefficients) and the relatives' similarity with respect to the phenotype (AAM). Phenotypic data are assumed to follow a multivariate-normal distribution. Linkage analysis was carried out on each population individually and in combination by combining within study standardized Z scores from each of the three samples. In accordance with previously proposed significance levels (33) for genome-wide linkage analysis, a significance threshold of logarithm of the odds (LOD) = 3.3 was adopted. Suggestive evidence of linkage was reported when a LOD score of 1.9 or greater was observed.

### Results

A summary of the AAM data from the three study populations is provided in Table 1. After combining phenotypic data from all three population samples, the mean and SD of AAM was 13.1 and 1.5 yr, respectively. The means and variances reported here are similar to those reported in previous studies (10). Variance-component analysis showed the heritability of AAM, estimated from the combination of the Australian, Dutch, and U.K. data, to be 72% (where all familial resemblance is assumed to be additive genetic).

Results of the genome-wide variance-component linkage analysis are shown for each of the three population samples and the combined analysis in Fig. 1, together with the locations of potential candidates and genes involved in the reproductive axis. The combined scan identified a region of suggestive linkage on chromosome 12 (LOD = 2.0), with signals coming from both the Australian and Dutch samples. The peak LOD score was located at 130 cM (110.1 Mbp), but the peak covered a broad region on chromosome 12q. The U.K. scan identified three loci of suggestive linkage to AAM on chromosomes 1, 4, and 18 with LOD scores of 2.4, 2.0, and 3.2, respectively. These peaks were not supported by scans in the other populations (Fig. 1). No loci of suggestive significance were found in either the Australian or Dutch scans individually. Results for the combined scan were

**TABLE 1.** Summary of the AAM sample data from the three study populations<sup>a</sup>

Study	Australia	The Netherlands	United Kingdom	Combined sample
Number of individuals	6,892	1,545	5,181	13,618
Mean (months)	156.2	160.7	156.1	156.7
Range (months)	105–208	120–209	108–209	105–209
SD (months)	16.8	16.9	18.5	17.6
Kurtosis	0.062	-0.184	-0.210	-0.013
Skewness	0.245	0.156	0.162	0.208

<sup>a</sup> After the winsorization of univariate outliers and after the removal of one individual from each pisp bivariate outlier.

similar when the endometriosis sample was excluded from the analysis or after AAM was adjusted for body mass index (supplementary Figs. 1 and 2, published as supplemental data on The Endocrine Society's Journals Online Web site at <http://jcem.endojournals.org>).

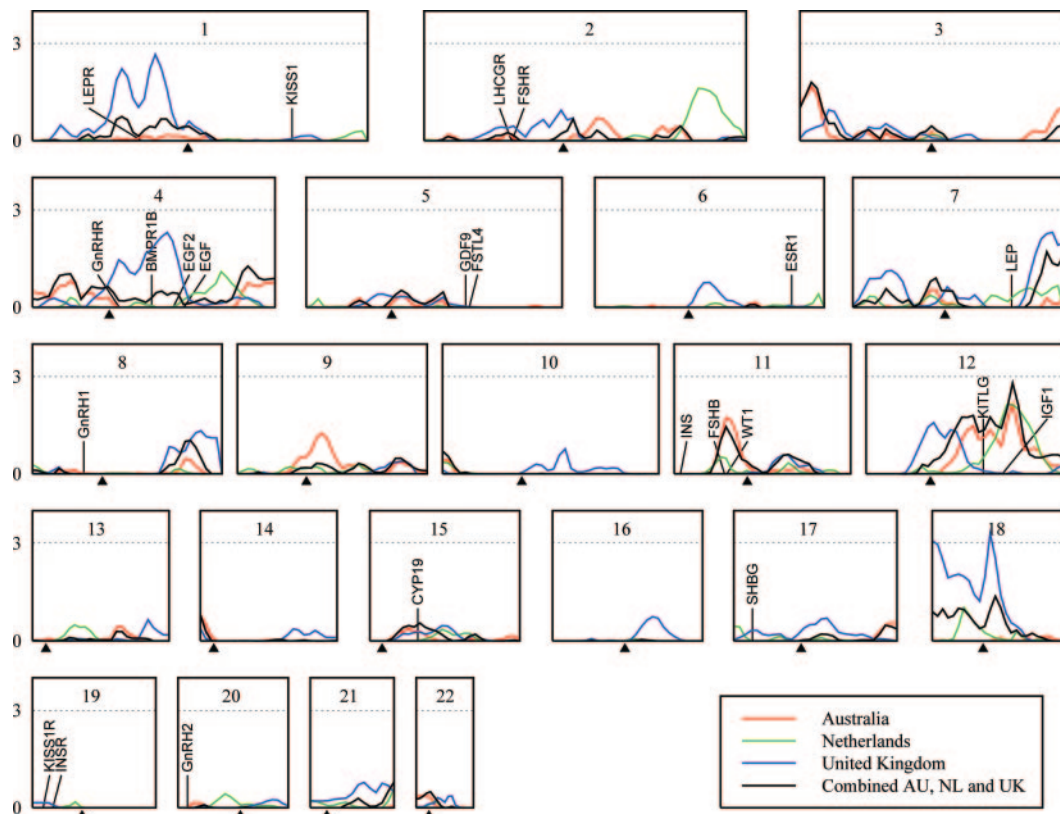
## Discussion

We performed a linkage scan on 4899 pisp from three populations, representing the largest linkage scan to date for AAM. Characteristics of AAM were generally similar in the three populations, with a slightly higher mean AAM in the Dutch sample.

No peaks reached the threshold for significant linkage in the combined scan or in individual study populations. We found one peak of suggestive linkage on chromosome 12q supported by evidence for linkage to this region in both the Australian and Dutch populations. Peaks of suggestive linkage identified in the U.K. families (chromosomes 1, 4, and 18) were not supported by results in the other two populations. The peak of suggestive linkage on chromosome 12q did not overlap with reported regions of linkage in previous studies (12, 13). No peaks were replicated across the different studies.

We did not have data on body fat or body mass index taken around the time of menarche and were therefore unable to account for these in our analyses. It is possible that the regions showing suggestive linkage to AAM actually underpin variation in these or any other unobserved covariates. If these regions are confirmed, further work is needed to ascertain whether the loci suggested here act directly on AAM or through correlated phenotypes. The same is true for any genetic linkage or association study because association does not imply causation.

Despite our large number of sibling pairs, the combined analysis (with or without bivariate outliers) revealed no significant linkage peaks, suggesting that the genetic architecture of AAM is complex. If a single quantitative trait locus (QTL) explained the variation in AAM, then we would expect to detect a gene of this effect size in a study of this magnitude. The fact that we have not detected a significant LOD score suggests that multiple loci of



**FIG. 1.** Genome-wide variance component LOD scores for Australian, Dutch, and U.K. samples plus the combined analysis. IBD estimates were calculated at 5-cM intervals, and variance-component linkage analysis was carried out using MERLIN. Genome-wide significance levels of LOD = 3.3 and LOD = 1.9 are indicative of statistically significant and suggestive loci, respectively. Candidate genes are indicated where appropriate. The centromere position is marked (▲). X chromosome marker genotypes were available only for the Australian sample, and therefore, a combined analysis of the X chromosome was not performed.

small or modest effect are involved in the heritability of AAM. With a sample of 4899 pips, we have 78% power to detect a QTL (at  $\alpha = 0.0001$ , genome-wide significance) that explains 15% of the trait variation (assuming a heritability of 70% and a recombination fraction of 0.0) (35, 36). In addition, our scan did not replicate peaks reported in previous scans (12, 13). This suggests that AAM variation is underpinned by multiple QTL of small effect and is similar to the genetic architecture of other complex traits. Large numbers of sibling pairs are needed to identify QTL with small effects using linkage analysis (37). If a QTL existed that explained 10% of the trait variation, for a trait with a heritability of 70% such as AAM, then approximately 11,390 pips would be required to detect the QTL with 80% power. Rare alleles with large effects on AAM could be segregating within the population, but these would explain little of the population variation in the trait.

We carried out our analysis after excluding bivariate outliers. The combined analysis with the bivariate outliers included showed some differences (data not shown). Ideally a LOD score peak is positively influenced by many sibling pairs and families, but this is seldom the case because the extremely concordant and discordant sibling pairs provide the majority of the power to detect linkage. For a marker in which an extremely discordant sibling pair share no alleles IBD, the sibling pair will have the effect of increasing the LOD score, and where they share two alleles IBD, they will decrease the LOD score at that locus. For this reason, it is difficult to know how to correctly handle bivariate outliers in linkage analysis. It seems reasonable to suggest that one may have more confidence in a linkage peak that is still present in the absence of bivariate outliers than those that are seen only in their presence. Furthermore, the aim of this study was to identify loci underlying the normal variation in AAM, and therefore, removing bivariate outliers would seem the logical approach. From data alone one cannot distinguish between the hypotheses that bivariate outliers are caused by (rare) alleles of large effect or by nongenetic effects, for example measurement errors. We recently discussed the issue of bivariate outliers in detail in a separate study on height (38).

We found no evidence for genetic linkage around several genes thought to play critical roles in AAM. Association between AAM and *ESR1* has been reported in several studies (14, 17, 39). We found no evidence for linkage to the region of the *ESR1* gene in any of the three populations studied. Recently the *KISS1/KISS1R* pathway has been shown to be critical for puberty. Both the ligand and receptor are expressed in the mediobasal hypothalamus of primates and play important roles in the regulation of GnRH neurons and the reproductive axis (40). We found no evidence for linkage to either *KISS1* or *KISS1R*, suggesting that normal variation in the timing of menarche is not explained by highly penetrant variants of large effect in these genes. We had sufficient power to detect such major loci, but the genes could still harbor common loci of small effect or very rare alleles of large effect because our linkage study is underpowered to detect such variants.

The region of suggestive linkage on chromosome 12 contains the IGF-I (*IGF1*) gene. *IGF1* plays an important role in control of the female reproductive system including pubertal development and menarche (41). A recent family-based association

study including 1048 females in 354 Caucasian nuclear families identified significant association between AAM and a common SNP (rs6214) in exon 4 of the *IGF1* gene (42). The SNP rs6214 is common with a minor allele frequency of 0.41, and there was also significant association with a haplotype including rs6214 and rs5742694. The mean AAM for subjects carrying the GA haplotype was 13.13 yr and the mean AAM for noncarriers was 12.86 yr (41). Association with common variants in *IGF1* and our evidence for suggestive linkage to chromosome 12 suggest the *IGF1* locus may contribute to variation in AAM and warrants further investigation. However, the association does not reach the level of genome-wide significance and has not been replicated in other samples. Given our finding of linkage to this region is only suggestive; additional studies must be carried out to establish whether this locus contributes to variation in AAM.

In summary, we have carried out the largest genome-wide scan for loci influencing variation in AAM. We report one region of suggestive linkage on chromosome 12q supported by evidence from two populations. The *IGF1* gene lies directly under this suggestive linkage peak and significant association with AAM was reported for SNP rs6124 in exon 4 of *IGF1*. Further studies are needed to confirm the role of *IGF1* and/or other variants in this region of chromosome 12 on AAM. A hypothesis-free whole genome association study may reveal other common polymorphisms of small effect on AAM.

## Acknowledgments

We thank to all the individuals and families who took part in the study. We also thank Ann Eldridge and Marlene Grace for ascertaining the adolescent sample age at menarche data, Australian Twin Registry staff for recruitment of the older Australian twin cohorts, Barbara Haddon for coordinating the Australian endometriosis study recruitment, Jacki Wicks and Olivia Zheng for preparation and management of Australian endometriosis study genome scan data, Anjali Henders and Megan Campbell for managing sample processing and DNA preparation, and Scott Gordon for his work merging and maintaining the Australian genome-scan data. We also thank Professor Weber (Mammalian Genotyping Service, Marshfield, WI) and Professor Slagboom (Molecular Epidemiology Section, Leiden University Medical Centre) for genotyping of The Netherlands samples.

Address all correspondence and requests for reprints to: Grant W. Montgomery, Molecular Epidemiology Laboratory, Queensland Institute of Medical Research, Post Office, Royal Brisbane Hospital, Queensland 4029, Australia. E-mail: grant.montgomery@qimr.edu.au.

This work was supported by the Australian Government's Cooperative Research Centre Program and the Grant 339430 from the National Health and Medical Research Council (NHMRC) of Australia. The adult twin studies were supported by Grants 941177 and 971232 from the NHMRC (Australia) and Grants AA07535, AA07728, and AA10249 from the National Institute of Alcoholism and Alcohol Abuse. The studies of twins born 1964–1971 were supported, in addition to NHMRC grants, by National Institutes of Health (NIH) Grants AA07728, AA09022, AA10249, AA11998, AA10333, DA00272, DA12854, and AA00277. The adolescent twin studies were supported by the Australian Research Council Grants A79600334, A79801419, A79906588, and DP0212016; the Human Frontier Science Program Grant RG0154.1998-B; NHMRC Grants 950998, 981339, and 241944; and the NIH Grant CA88363. The work was also supported by NHMRC Medical Bioinformatics Program Grant 389982. Studies in The Neth-

erlands were supported by Spinozapremie Grant NWO/SPI 56-464-14192; Genetic-epidemiology of anxious depression Grant ZonMW 940-37-024; Twin research focusing on behavior and depression Grant NWO 400-05-717; Quantitative genetic modeling of variation in cognitive brain maturation Grant NWO 051.02.060; Center for Medical Systems Biology (NWO Genomics); Twin-family database for behavior genetics and genomics studies Grant NWO-MagW 480-04-004; Genome-wide analyses of European twin and population cohorts to identify genes predisposing to common diseases Grant EU/QLRT-2001-01254. G.W.M. and P.M.V. were supported by NHMRC fellowships, and C.A.A. was supported by an MRC (United Kingdom) postgraduate studentship.

Disclosure Statement: The authors have nothing to disclose.

## References

- Kaltiala-Heino R, Kosunen E, Rimpela M 2003 Pubertal timing, sexual behaviour and self-reported depression in middle adolescence. *J Adolesc* 26:531–545
- Kaltiala-Heino R, Rimpela M, Rissanen A, Rantanen P 2001 Early puberty and early sexual activity are associated with bulimic-type eating pathology in middle adolescence. *J Adolesc Health* 28:346–352
- Velie EM, Nechuta S, Osuch JR 2006 Lifetime reproductive and anthropometric risk factors for breast cancer in postmenopausal women. *Breast Dis* 24:17–35
- Naves M, Diaz-Lopez JB, Gomez C, Rodriguez-Rebollar A, Cannata-Andia JB 2005 Determinants of incidence of osteoporotic fractures in the female Spanish population older than 50. *Osteoporosis Int* 16:2013–2017
- Cooper GS, Ephross SA, Weinberg CR, Baird DD, Whelan EA, Sander DP 1999 Menstrual and reproductive risk factors for ischemic heart disease. *Epidemiology* 10:255–259
- Morabia A, Costanza MC, The World Health Organization Collaborative Study of Neoplasia and Steroid Contraceptives 1998 International variability in ages at menarche, first livebirth and menopause. *Am J Epidemiol* 148:1195–1205
- Thomas F, Renaud F, Benefice E, De Meeus T, Guegan J-F 2001 International variability of ages at menarche and menopause: Patterns and main determinants. *Hum Biol* 73:271–290
- Tena-Sempere M 2006 GRP54 and kisspeptin in reproduction. *Hum Reprod Update* 12:631–639
- Anderson CA, Duffy DL, Martin NG, Visscher PM 2007 Estimation of variance components for age at menarche in twin families. *Behav Genet* 37:668–677
- Towne B, Czerwinski SA, Demerath EW, Blangero J, Roche AF, Siervogel RM 2005 Heritability of age at menarche in girls from the Fels Longitudinal Study. *Am J Phys Anthropol* 128:210–219
- Treloar SA, Martin NG 1990 Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. *Am J Hum Genet* 47:137–148
- Guo Y, Shen H, Xiao P, Xiong DH, Yang TL, Guo YF, Long JR, Recker RR, Deng HW 2006 Genomewide linkage scan for quantitative trait loci underlying variation in age at menarche. *J Clin Endocrinol Metab* 91:1009–1014
- Rothembuhler A, Fradin D, Heath S, Lefevre H, Bouvattier C, Lathrop M, Bougneres P 2006 Weight-adjusted genome scan analysis for mapping quantitative trait loci for menarchal age. *J Clin Endocrinol Metab* 91:3534–3537
- Stavrou I, Zois C, Ioannidis JP, Tsatsoulis A 2002 Association of polymorphisms of the oestrogen receptor  $\alpha$  gene with the age of menarche. *Hum Reprod* 17:1101–1105
- Guo Y, Xiong D-H, Yang T-L, Guo Y-F, Recker RR, Deng H-W 2006 Polymorphisms of estrogen-biosynthesis genes CYP17 and CYP19 may influence age at menarche: a genetic association study in Caucasian females. *Hum Mol Genet* 15:2401–2408
- Xita N, Tsatsoulis A, Stavrou I, Georgiou I 2005 Association of SHBG gene polymorphism with menarche. *Mol Hum Reprod* 11:459–462
- Stavrou I, Zois C, Chatzikyriakidou A, Georgiou I, Tsatsoulis A 2006 Combined estrogen receptor  $\alpha$  and estrogen receptor  $\beta$  genotypes influence the age of menarche. *Hum Reprod* 21:554–557
- Zhu G, Duffy DL, Eldridge A, Grace M, Mayne C, O’Gorman L, Aitken JF, Neale MC, Hayward NK, Green AC, Martin NG 1999 A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: a maximum-likelihood combined linkage and association analysis in twins and their sibs. *Am J Hum Genet* 65:483–492
- Wright M, Geus Ed, Ando J, Luciano M, Posthuma D, Ono Y, Hansell N, Baal Cv, Hiraishi K, Hasegawa T, Smith G, Geffen G, Geffen L, Kanba S, Miyake A, Martin N, Boomsma D 2001 Genetics of cognition: outline of a collaborative twin study. *Twin Res* 4:48–56
- Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, Phillips JW, Sachs A, Schadt EE 2004 Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 75:1094–1105
- Heath AC, Howells W, Kirk KM, Madden PA, Bucholz KK, Nelson EC, Slutske WS, Statham DJ, Martin NG 2001 Predictors of non-response to a questionnaire survey of a volunteer twin panel: findings from the Australian 1989 twin cohort. *Twin Res* 4:73–80
- Cornes BK, Medland SE, Ferreira MAR, Morley KI, Duffy DL, Heijmans BT, Montgomery GW, Martin NG 2005 Sex-limited genome-wide linkage scan for body mass index in an unselected sample of 933 Australian twin families. *Twin Res Hum Genet* 8:616–632
- Treloar S, Hadfield R, Montgomery G, Lambert A, Wicks J, Barlow DH, O’Connor DT, Kennedy S, Group tES 2002 The International Endogene Study: a collection of families for genetic research in endometriosis. *Fertil Steril* 78:679–685
- Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, Dawson G, Mackay IJ, Weeks DE, Bennett ST, Carey A, Ewen-White KR, Duffy DL, O’Connor DT, Barlow DH, Martin NG, Kennedy SH 2005 Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. *Am J Hum Genet* 77:365–376
- Boomsma DI, de Geus EJC, Vink JM, Stubbe JH, Distel MA, Hottenga J-J, Posthuma D, van Beijsterveldt TCEM, Hudziak JJ, Bartels M, Willemsen G 2006 Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 9:849–857
- Vink JM, Willemsen G, Stubbe JH, Middeldorp CM, Ligthart RSL, Baas KD, Dirkzwager HJC, Geus EJC, Boomsma DI 2004 Estimating non-response bias in family studies: application to mental health and lifestyle. *Eur J Epidemiol* 19:623–630
- Berg SMvd, Boomsma DI 2007 The familial clustering of age at menarche in extended twin families. *Behav Genet* 37:661–667
- Min JL, Lakenberg N, Bakker-Verweij M, Suchiman E, Boomsma DI, Slagboom PE, Meulenbelt I 2006 High microsatellite and SNP genotyping success rates established in a large number of genomic DNA samples extracted from mouth swabs and genotypes. *Twin Res Hum Genet* 9:501–506
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR 2002 Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101
- Spector TD, Williams FMK 2006 The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* 9:899–906
- Wilson SG, Reed PW, Bansal A, Chiano M, Lindersson M, Langdown M, Prince RL, Thompson D, Thompson E, Bailey M, Kleyn PW, Sambrook P, Shi MM, Spector TD 2003 Comparison of genome screens for two independent cohorts provides replication of suggestive linkage of bone mineral density to 3p21 and 1p36. *Am J Hum Genet* 72:144–155
- R Development Core Team 2005 R: A language and environment for statistical computing reference index version 2.2.1 R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Sullivan PF, Montgomery GW, Hottenga JJ, Wray NR, Boomsma DI, Martin NG 2006 Empirical evaluation of the genetic similarity of samples from twin registries in Australia and The Netherlands using 359 STRP markers. *Twin Res Hum Genet* 9:600–602
- Lander E, Kruglyak L 1995 Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Sham PC, Cherny SS, Purcell S, Hewitt JK 2000 Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 66:1616–1630
- Purcell S, Cherny SS, Sham PC 2003 Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19:149–150
- Risch N, Merikangas K 1996 The future of genetic studies of complex human diseases. *Science* 273:1616
- Benyamin B, Perola M, Cornes BK, Madden PA, Palotie A, Nyholt DR, Montgomery GW, Peltonen L, Martin NG, Visscher PM 2008 Within-family outliers: segregating alleles or environmental effects? A linkage analysis of height from 5815 sibling pairs. *Eur J Hum Genet* 16:516–524
- Long JR, Xu H, Zhao LJ, Liu PY, Shen H, Liu YJ, Xiong DH, Xiao P, Liu YZ, Dvornyk V, Li JL, Recker RR, Deng HW 2005 The oestrogen receptor  $\alpha$  gene is linked and/or associated with age of menarche in different ethnic groups. *J Med Genet* 42:796–800
- Plant TM 2006 The role of KiSS-1 in the regulation of puberty in higher primates. *Eur J Endocrinol* 155(Suppl 1):S11–S16
- Spiliotis BE 2003 Growth hormone insufficiency and its impact on ovarian function. *Ann NY Acad Sci* 997:77–84
- Zhao J, Xiong DH, Guo Y, Yang TL, Recker RR, Deng HW 2007 Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in Caucasian females. *Hum Reprod* 22:1789–1794