Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in caucasian females

Jian Zhao1*, Dong-Hai Xiong2*, Yan Guo1*, Tie-Lin Yang1, Robert R.Recker2 and Hong-Wen Deng1,2,3,4

1The Key Laboratory of Biomedical Information Engineering of Ministry of Education and Institute of Molecular Genetics, School of Life Science and Technology, Xi’an Jiaotong University, Xi’an 710049, People’s Republic of China; 2Osteoporosis Research Center and Department of Biomedical Sciences, Creighton University, Omaha, NE 68131, USA; 3Department of Orthopedic Surgery, School of Medicine, University of Missouri-Kansas City, 2411 Holmes Street, Kansas City, MO 64108, USA
4To whom correspondence should be addressed at: School of Medicine, Department of Basic Medical Science, University of Missouri/Kansas City, Room: M3-CO3, 2411 Holmes Street, Kansas City, MO 64108-2792, USA. Tel: +816-235-5354, Fax: +816-235-6517, E-mail: dengh@umkc.edu

BACKGROUND: The insulin-like growth factor 1 (IGF1) gene, which plays a crucial role in hypothalamic–pituitary–ovarian hormone-controlled metabolic processes, may influence the onset of menarche. Our study aimed to test association between IGF1 polymorphisms with the variation of age at menarche (AAM) in Caucasian females. METHODS: We recruited a sample of 1048 females from 354 Caucasian nuclear families and genotyped 19 single-nucleotide polymorphisms (SNPs) spanning the entire IGF1 gene. Pairwise linkage disequilibrium among SNPs was measured, and the haplotype blocks were inferred. Both single SNP markers and haplotypes were tested for association with AAM using the quantitative transmission disequilibrium test. RESULTS: Significant association ($P = 0.0153$) between AAM and SNP3 (rs6214) in block1 was detected. CONCLUSIONS: Our results suggested a potential effect of SNP3 in the IGF1 gene on AAM variation in Caucasian women for the first time. However, further independent studies are needed to confirm our findings.

Keywords: association; insulin-like growth factor 1; age at menarche; haplotype

Introduction

Menarche, the beginning of menstruation, is the cornerstone in women’s lives (Ku et al., 2006). Early menarche is associated with elevated risk of ovarian tumors and breast cancer (Geschwind and Galaburda, 1985; Trichopoulos, 1990). Delayed menarche may increase the risk of osteoporosis (Eastell, 2005). Thus, it is valuable and necessary to explore the potential factors influencing the variation of age at menarche (AAM).

AAM is determined by different environmental and genetic factors (van Noord et al., 1997; Worda et al., 2004). It was reported that 53–74% of the variation in AAM could be due to genetic effects (Kaprio et al., 1995; Sharma, 2002). For example genetic linkage and association studies have suggested several quantitative trait loci (Guo et al., 2006) and candidate genes for AAM (Gorai et al., 2003; Xita et al., 2005). However, these findings only partly explain the genetic architecture underlying AAM variation, and more AAM-influencing genetic factors wait to be found.

Insulin-like growth factor 1 (IGF1) plays a crucial role in hypothalamic–pituitary–ovarian hormone-controlled metabolic processes (Jones and Clemmons, 1995) and is the major effector of bone growth (Khosla, 1997; Kawai et al., 1999). It is known that GnRH is the key regulator of the reproductive system, directly regulating the release of LH and FSH that are essential for the normal function of the gonads (Chandrashekar and Bartke, 2003). Hypothalamic GnRH mRNA expression increased at 3.5 months and declined by 6 months of age during puberty in the gilt (Barb et al., 2006). The increment in IGF1 might be involved in the stimulation of GnRH activity in the normal children (Belgorosky and Rivarola, 1998). Furthermore, previous studies have shown that IGF1 is a metabolic signal capable of activating the GnRH/LH-releasing system at the time of puberty in rats (Hiney et al., 1991) and IGF1 is capable of enhancing FSH-stimulated LH receptor expression (Adashi et al., 1985c). IGF1 and FSH can synergize to stimulate progesterone production in primary cultures of maturing human granulosa cells (LaVoie et al., 1999). IGF1 also enhances FSH-mediated steroidogenesis, including estradiol (E2) and progesterone production (Bergh et al., 1991 Adashi et al., 1985a;
Veldhuis and Demers, 1985; Veldhuis and Rodgers, 1987; Erickson et al., 1989, 1991). All of these findings suggest that the \textit{IGF1} may play significant roles in pubertal development, menarche, the menstrual cycle, fertility and reproduction system (Hiney et al., 1996; Huang et al., 1998; Spiliotis, 2003). Menarche depends on the maturation of female reproductive system. Early Menarche is associated with early onset of ovulatory cycles (Hickey and Balen, 2003). It has been reported that hypothalamic–pituitary–ovarian axis controls the age of first ovulation and occurrence of menarche (Beitins, 1981; Hickey and Balen, 2003; Loucks, 2006). It is therefore possible that \textit{IGF1} may have an effect on the onset of menarche.

Based on this hypothesis, we tested \textit{IGF1} gene single nucleotide polymorphisms (SNPs) comprehensively for an association with AAM variation using the quantitative transmission disequilibrium test (QTDT) in a large sample of Caucasian females.

Materials and Methods

Subjects

This study was approved by the Creighton University Institutional Review Board. Signed informed-consent documents were obtained from all participants before they entered the study. Women with chronic diseases and conditions that might potentially affect bone mass, structure, or metabolism were excluded. These diseases/conditions included chronic disorders involving vital organs (heart, lung, liver, kidney and brain), serious metabolic diseases (diabetes, hypo- and hyper-parathyroidism, hyperthyroidism, etc.), other skeletal diseases (Paget disease, osteogenesis imperfecta, rheumatoid arthritis, etc.), chronic use of drugs affecting bone metabolism (hormone replacement therapy, corticosteroid therapy and anti-convulsant drugs) and malnutrition conditions (such as chronic diarrhea, chronic ulcerative colitis, etc.). Initially, all of the 1873 participants from 405 nuclear families were US Caucasians of European origin recruited for complex diseases or traits research. For menarche study, we selected families that contain at least two informative female subjects (≥1 daughters) with AAM data. Therefore, 50 families having only male offspring and 1 family without enough AAM data for daughters were excluded. At last, 1051 female subjects from 354 nuclear families were studied, with the ages ranging from 19 to 85 years old.

AAM data

For each study subject, the detailed medical information including menstrual history was recorded by nurse-administered questionnaire. AAM was calculated as the date of menarche following the onset of menses minus the date of birth (in years rounded to the first decimal place). After excluding three outliers whose AAM were 18 years or older (diagnosed as primary amenorrhea), our AAM data followed normal distribution as verified by the Kolmogorov–Smirnov test implemented in the software Minitab (Minitab, Inc., State College, PA, USA). The final sample used in the association analysis consisted of 1048 females and their AAM ranged from 8.5 to 17.0 years with a mean of 13.0 (SD = 1.4).

SNP selection and genotyping

A total of 19 SNPs in and around \textit{IGF1} gene were selected on the basis of the following criteria: (i) validation status, namely, SNPs we selected exist in Caucasian, (ii) an average density of 1 SNP per 4 kb, (iii) degree of heterozygosity, i.e. minor allele frequencies (MAF) >0.05, (iv) functional relevance and importance and (v) reported to SNP database by various sources. Genomic DNA was extracted from whole blood using a commercial isolation kit (Gentra Systems, Minneapolis, MN, USA) following the procedure detailed in the kit. DNA concentration was assessed by a DU530 UV/VIS spectrophotometer (Beckman Coulter, Inc., Fullerton, CA, USA). All the subjects were genotyped, yielding a large set of genotype data that are further subject to Mendelian inheritance checking, allele frequency estimation and haplotype reconstruction. Nineteen SNPs were successfully genotyped using the high-throughput BeadArray SNP genotyping technology of Illumina Inc (San Diego, CA, USA). The average rate of missing genotype data was reported by Illumina to be ~0.05%. The average genotyping error rate estimated through blind duplicating was reported to be less than ~0.01%. The nineteen SNPs were spaced ~85 kb apart on average and covered the full transcript length of the \textit{IGF1} gene.

Statistical analysis

PedCheck (O’Connell and Weeks, 1998) was used to check Mendelian consistency of SNP genotype data and any inconsistent genotypes were removed. Then the error checking option embedded in Merlin (Abecasis et al., 2002) was run to identify and disregard the genotypes flanking excessive recombinants, thus further reducing genotyping errors. Allele frequencies for each SNP were calculated by allele counting, and the Hardy–Weinberg equilibrium was tested using the PEDSTATS procedure embedded in Merlin. Population haplotypes and their frequencies were inferred for each of the nineteen genes using PHASE v2.1.1 software among 703 unrelated parents. Linkage disequilibrium (LD) structure for each gene was defined, using GOLD (http://www.sph.umich.edu/csg/abecasis/GOLD/), to chart pairwise D’ statistics derived from haplotype data. HaploBlockFinder (http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi/) was used to identify block structures and select haplotype-tagging SNPs (htSNPs) of each candidate gene. To infer haplotypes defined by the tagging SNPs within each block of the gene for all of the subjects among 405 families, we adopted the algorithm of integer linear programming implemented in PedPhase V2.0 (http://www.cs.ucr.edu/~kzili/haplotyping.html), which is based on LD assumption and able to recover phase information at each marker locus with great speed and accuracy even in the presence of 20% missing data (Li and Jiang, 2005). The QTDT software (http://www.sph.umich.edu/csg/abecasis/QTDT/) (Allison, 1997; Abecasis et al., 2000) was used to test each SNP and haplotype with estimated frequencies >5% for association with AAM variation. Using QTDT, we also performed 1000 permutations to adjust for the potential multiple-testing problem.

Results

Allele frequencies, LD and haplotype reconstruction

The information of all the \textit{IGF1} SNPs is summarized in (Table 1). The MAFs of these SNPs ranged from 0.06 to 0.41. All the SNPs were in Hardy–Weinberg equilibrium (\(P > 0.01\)). LD block structure of \textit{IGF1} is shown in Figure 1. We identified three significant LD blocks, with the sizes of 13.6, 53.9 and 17.3 kb, respectively. Block 1 ranged from the 3′ untranslated region to exon4 with SNP2 and SNP3 selected as htSNPs. Block 2 extended from intron3 to intron2 with SNP4 and SNP13 selected as htSNPs. Block 3 covered from intron2 to the promoter region with SNP16, SNP17 and SNP19 selected as htSNPs. SNP15 and rs5742629 had no strong LD with any other SNPs and thus cannot be assigned
to any of the three blocks. The detailed information for blocks is presented in (Table 2).

**QTDT analyses**

Table 3 summarizes association results of IGF1 gene with AAM in nuclear families by QTDT. We tested all htSNPs and haplotypes for association with AAM. We did not find any evidence of population stratification for either single SNP markers or haplotypes. For single locus analysis, we detected significant association for SNP3 \((P = 0.0153)\), which lies in block 1. We also detected evidence of association with AAM for haplotype GA \((P = 0.0243)\) in block 1. These two associations both reached the permutation-established experiment-significance level of \(a = 0.026\). Thus the association evidence matched at both the single SNP and haplotype levels. We further assessed the effect of different haplotypes in block 1 on AAM. The mean AAM for subjects carrying the GA haplotype in block 1 is 13.13 years (SD = 1.44), and for the non-carriers it is 12.86 years (SD = 1.39) \((P = 0.0243)\).

**Discussion**

Our study is the first to suggest that IGF1 may associate with the variation of AAM. The protein encoded by the IGF1 gene is involved in the biosynthesis of ovarian hormones and

<table>
<thead>
<tr>
<th>SNP</th>
<th>Marker</th>
<th>Minor allele frequency</th>
<th>Location</th>
<th>Alleles</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs2373720</td>
<td>0.0638</td>
<td>3' UTR</td>
<td>A/G</td>
<td>101282429</td>
</tr>
<tr>
<td>2</td>
<td>rs6219</td>
<td>0.0998</td>
<td>Exon4</td>
<td>A/G</td>
<td>101292659</td>
</tr>
<tr>
<td>3</td>
<td>rs6214</td>
<td>0.4052</td>
<td>Exon4</td>
<td>A/G</td>
<td>101296036</td>
</tr>
<tr>
<td>4</td>
<td>rs5742694</td>
<td>0.2521</td>
<td>Intron3</td>
<td>A/C</td>
<td>101301703</td>
</tr>
<tr>
<td>5</td>
<td>rs978458</td>
<td>0.2706</td>
<td>Intron3</td>
<td>A/G</td>
<td>101304706</td>
</tr>
<tr>
<td>6</td>
<td>rs4764883</td>
<td>0.2935</td>
<td>Intron3</td>
<td>C/T</td>
<td>101308772</td>
</tr>
<tr>
<td>7</td>
<td>rs11111267</td>
<td>0.1939</td>
<td>Intron3</td>
<td>A/G</td>
<td>101313706</td>
</tr>
<tr>
<td>8</td>
<td>rs5742671</td>
<td>0.1939</td>
<td>Intron2</td>
<td>A/G</td>
<td>101318225</td>
</tr>
<tr>
<td>9</td>
<td>rs5742667</td>
<td>0.2716</td>
<td>Intron2</td>
<td>A/G</td>
<td>101325945</td>
</tr>
<tr>
<td>10</td>
<td>rs2373721</td>
<td>0.2529</td>
<td>Intron2</td>
<td>A/G</td>
<td>101329512</td>
</tr>
<tr>
<td>11</td>
<td>rs2288378</td>
<td>0.2529</td>
<td>Intron2</td>
<td>A/G</td>
<td>101332475</td>
</tr>
<tr>
<td>12</td>
<td>rs1075380</td>
<td>0.2771</td>
<td>Intron2</td>
<td>A/G</td>
<td>101346703</td>
</tr>
<tr>
<td>13</td>
<td>rs203378</td>
<td>0.0582</td>
<td>Intron2</td>
<td>A/G</td>
<td>101349534</td>
</tr>
<tr>
<td>14</td>
<td>rs4764697</td>
<td>0.2845</td>
<td>Intron2</td>
<td>A/G</td>
<td>101355639</td>
</tr>
<tr>
<td>15</td>
<td>rs5742629</td>
<td>0.2819</td>
<td>Intron2</td>
<td>A/G</td>
<td>101359730</td>
</tr>
<tr>
<td>16</td>
<td>rs1019731</td>
<td>0.1438</td>
<td>Intron2</td>
<td>A/C</td>
<td>101366892</td>
</tr>
<tr>
<td>17</td>
<td>rs12821878</td>
<td>0.2258</td>
<td>Intron2</td>
<td>A/G</td>
<td>101370134</td>
</tr>
<tr>
<td>18</td>
<td>rs2162679</td>
<td>0.1554</td>
<td>Intron1</td>
<td>A/G</td>
<td>101373726</td>
</tr>
<tr>
<td>19</td>
<td>rs35765</td>
<td>0.1149</td>
<td>Promoter</td>
<td>A/C</td>
<td>101384163</td>
</tr>
</tbody>
</table>

present in many tissues to control cell differentiation, proliferation and apoptosis. It has been proposed that IGF1 functions as an autocrine/paracrine regulator of ovary development. The IGF system plays a role in the formation of the ovary and ovum and participates in the regulation of bone metabolism (Libanati et al., 1999). In this study, we found the significant associations with AAM for SNP3 and the haplotype GA reconstructed by the htSNPs (SNP3 and SNP4). SNP3 (rs6214) is located in exon4, and although itself does not cause any amino acid change, it may be in strong LD with certain amino acid sequence of IGF1 protein. However, all of the above are just hypotheses and wait to be tested by further functional analyses.

In females, different kinds of hormones, growth factors and cytokines are involved in the onset of menarche (Tamura et al., 1998; Behl and Kaul, 2002; Bornstein et al., 2004; Quirk et al., 2004). Lots of evidence showed that the IGF1 system plays a key role in ovarian folliculogenesis (Adashi et al., 1985b; Giudice, 1992). In vivo levels of IGF1 (Cook et al., 1993) and IGF1 binding protein (IGFBP3) are positively correlated with the serum E2 concentration in girls during puberty (Sutel et al., 2000; Blogowska et al., 2003; Kanbur-Oksuz et al., 2001; Kulik-Rechberger and Janiszewska, 2004). More interestingly, there is a higher positive correlation between plasma E2 and IGF1/IGFBP3 ratio, an index of free IGF1. There is a positive correlation between LH and FSH with both IGF1 and IGFBP3, and FSH and LH stimulate some isoform of hepatic IGF1 mRNA (Ruiz et al., 2001). Concurrent treatment with increasing concentrations of IGF1 brought about dose-dependent increases in FSH-induced LH receptor mRNA, a response which was 2.5-fold greater than that induced by FSH alone at the highest concentration in rat granulosa cells (Hirakawa et al., 1999). On average, the production of E2 was ~2.8-fold higher when polycystic ovary disease granulosa cells were treated concomitantly with FSH and IGF1 than that evoked only by FSH (Erickson et al., 1990). In all, these biological and physiological studies lend support to our finding that the IGF1 gene polymorphisms influence the variation of AAM.

Our study has several notable strengths. First, our sample is relatively large, which includes 1048 females from 354 Caucasian nuclear families. Therefore, adequate statistical power is guaranteed to find the genetic variants of modest effect sizes on the AAM variation. Second, we selected 19 SNPs distributed throughout the IGF1 gene to test for its association with AAM, rather than just testing only a few SNPs in the gene. These SNPs comprehensively covered the entire IGF gene. Third, we adopted the family-based association analysis to obviate the known population stratification problem in the European American population. All of these approaches made this study a robust and reliable one for the follow-up research.

We estimated the power of our study using the Program Genetic Power Calculator (GPC, http://statgen.iop.kcl.ac.uk/ gpc/qtlassoc.html). Under the condition of the conservative significance level of 0.001 and the strong LD of \(|D'| = 0.9\), our sample can achieve 90% power to detect the causal variants responsible for ~4% variation of AAM.

In general, AAM is acquired by recall and self-reports, which may increase the likelihood of error. However, study suggests a high correction of 0.79 between recalled and actual AAM and it is also found that recall of AAM was generally quite good, both in precision and accuracy (Must et al., 2002). AAM was recalled to within 1 year of the actual event by 84% of the females (Casey et al., 1991). The accuracy of recall of AAM was confirmed further in other previous studies (Bergsten-Brufecors 1976; Bean et al., 1979; Greif and Ulman, 1982; Koo and Rohan, 1997). In addition, study found that important personal experiences can produce clear memories of a person’s circumstance at the time of the event, that is ‘flashbulb memories’ (Pillemer et al., 1987), and menarche was such a significant developmental event that almost all of the women remembered the date of their first menstruation (Golub and Catalano, 1983). Therefore, recalled measures and self-reports should be quite valid to obtain AAM.

In conclusion, our present study does support the hypothesis that IGF1 variants can influence the onset age of menarche in Caucasians, especially in the European Americans. However, the causal functional variants underlying the observed associations are still unknown. Replication of our results and functional studies are necessary to unravel the true associated variants and the corresponding molecular mechanisms.

Acknowledgement

Investigators of this work were partially supported by grants from NIH (R01 AR050496, K01 AR21770-01, R01 AR45349-01, and R01 GM60402-01A1) and an LB595 grant from the State of Nebraska. The study also benefited from grants from Project 30570875 supported by National Natural Science Foundation of China, Xi’an Jiaotong University, and the Ministry of Education of China.

References


Table 3: QTDT results for the associations between IGF1 polymorphisms and AAM variation

<table>
<thead>
<tr>
<th>SNP*</th>
<th>Marker</th>
<th>Stratification</th>
<th>Association*</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs6219</td>
<td>0.2796</td>
<td>0.9871</td>
</tr>
<tr>
<td>3</td>
<td>rs6214</td>
<td>0.6193</td>
<td>0.0153</td>
</tr>
<tr>
<td>4</td>
<td>rs5742694</td>
<td>0.7503</td>
<td>0.9993</td>
</tr>
<tr>
<td>13</td>
<td>rs2033178</td>
<td>NS</td>
<td>0.8037</td>
</tr>
<tr>
<td>15</td>
<td>rs5742629</td>
<td>0.9987</td>
<td>0.1343</td>
</tr>
<tr>
<td>16</td>
<td>rs1019731</td>
<td>NS</td>
<td>0.6233</td>
</tr>
<tr>
<td>17</td>
<td>rs12821878</td>
<td>0.5150</td>
<td>0.6581</td>
</tr>
<tr>
<td>19</td>
<td>rs357656</td>
<td>0.5483</td>
<td>0.7865</td>
</tr>
<tr>
<td>Block1 haplotype</td>
<td>AG</td>
<td>0.2981</td>
<td>0.3312</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>0.9845</td>
<td><strong>0.0243</strong></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>0.4796</td>
<td>0.0545</td>
</tr>
</tbody>
</table>

*SNPs listed in this table are haplotype-tagging SNPs; *P < 0.05 are labeled in bold.


Beitins IZ. Menstrual abnormalities during adolescence. Prim Care 1981;8:3–18.


Khosa S. Idiopathic osteoporosis—is the osteoblast to blame? J Clin Endocrinol Metab 1997;82:2792–4.


Loucks AB. The response of luteinizing hormone pulsatility to 5 days of low energy availability disappears by 14 years of gynecological age. J Clin Endocrinol Metab 2005;90:1318–64.


Submitted on August 10, 2006; resubmitted on October 27, 2006; resubmitted on December 22, 2006; accepted on January 2, 2007