SCIENCE CHINA
Life Sciences

• RESEARCH PAPER •

June 2012 Vol.55 No.6: 516–520
doi: 10.1007/s11427-012-4327-6

Bivariate genome-wide association study suggests that the *DARC* gene influences lean body mass and age at menarche

HAI Rong[1,2], ZHANG Lei[1,3], PEI YuFang[1,3], ZHAO LanJuan[3], RAN Shu[1], HAN YingYing[1], ZHU XueZhen[1], SHEN Hui[3], TIAN Qing[3] & DENG HongWen[1,3*]

[1]*Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, China;*
[2]*The Affiliated Hospital of Inner Mongolia Medical College, Huhhot 010050, China;*
[3]*Department of Biostatistics, Tulane University, New Orleans, LA 70112, USA*

Received February 3, 2012; accepted March 20, 2012

Lean body mass (LBM) and age at menarche (AAM) are two important complex traits for human health. The aim of this study was to identify pleiotropic genes for both traits using a powerful bivariate genome-wide association study (GWAS). Two studies, a discovery study and a replication study, were performed. In the discovery study, 909622 single nucleotide polymorphisms (SNPs) were genotyped in 801 unrelated female Han Chinese subjects using the Affymetrix human genome-wide SNP array 6.0 platform. Then, a bivariate GWAS was performed to identify the SNPs that may be important for LBM and AAM. In the replication study, significant findings from the discovery study were validated in 1692 unrelated Caucasian female subjects. One SNP *rs3027009* that was bivariately associated with left arm lean mass and AAM in the discovery samples (P=7.26×10^{-6}) and in the replication samples (P=0.005) was identified. The SNP is located at the upstream of *DARC* (Duffy antigen receptor for chemokines) gene, suggesting that *DARC* may play an important role in regulating the metabolisms of both LBM and AAM.

bivariate genome-wide association study, age at menarche, lean body mass, *DARC* gene

Citation: Hai R, Zhang L, Pei Y F, *et al*. Bivariate genome-wide association study suggests that the *DARC* gene influences lean body mass and age at menarche. Sci China Life Sci, 2012, 55: 516−520, doi: 10.1007/s11427-012-4327-6

Skeletal muscle is an important organ for human health. Loss and functional impairment of skeletal muscle predispose individuals to a series of severe disorders, such as sarcopenia, osteoporotic fracture, obesity, impaired protein balance, dyslipidemia, insulin resistance, and increased mortality [1,2]. Skeletal muscle was measured by lean body mass (LBM) [3], which is the best index for sarcopenia. LBM is a highly inheritable trait, with estimated heritability ranging from 52% to 84% [4–6]. Several candidate loci and/or genes have been identified for LBM. Among them, the gene *GREMLIN1* at 15q13.3 was reported by Wang *et al*. [7] to be linked with LBM through genome-wide linkage studies. Liu *et al*. [8] identified the gene *TRHR* at 8q23.1 and found that it was associated with LBM also through genome-wide association studies (GWAS).

Menarche is the first menstrual cycle in females. It signals the sexual maturation and fertility of a woman. Age at menarche (AAM) is an important complex trait for women's health. AAM is associated with a variety of complex diseases, including breast and endometrial cancer [9], Alzheimer's disease [10], osteoporosis [11,12], and heart disease [13]. AAM is under strong genetic determination. There are highly significant correlations between AAM in mothers and daughters [14], and family history is a strong predictor for early menarche. The estimated heritability for AAM ranges from 50% to 70% [15–18]. Several candidate

*Corresponding author (email: hdeng2@tulane.edu)

genes have been suggested to influence AAM, such as androgen receptor (*AR*), estrogen receptor alpha (*ER-a*) and beta (*ER-b*) [19–21], vitamin D receptor (*VDR*), sex hormone binding globulin (*SHBG*) [22], insulin-like growth factor 1 (*IGF-1*) [23], chemokine (C-C-motif) receptor 3 (*CCR3*) [24], and the *CYP* gene family [25–27].

Despite these findings, the genetic background underlying both LBM and AAM remains largely unknown. This lack of knowledge is partially because of the limited statistical power of the small sample sizes and locus effects. Alternative analytical approaches with improved power are needed to help identify additional responsible loci. Both LBM and AAM are regulated by sex hormones, and humoral and cellular factors. Therefore, these two traits may share common genetic determinants. In the presence of such pleiotropic genes, bivariate analysis, which can analyze the two traits jointly [28], is an alternative analytical approach to the existing univariate ones. By explicitly modeling genetic and environmental correlations between traits, bivariate analysis can theoretically extract more information from the data, consequently providing the test with greater power [29]. Thus, it is of great interest to study LBM and AAM using bivariate analysis with the purpose of identifying potential pleiotropic genes for both traits.

With this aim, we performed a bivariate GWAS of the LBM and AAM traits. Using the Affymetrix human SNP 6.0 array, we successfully genotyped and analyzed ~900000 SNPs in 801 unrelated female Han Chinese subjects. We replicated the significant findings in another sample of 1692 unrelated Caucasian female subjects. Through bivariate association analyses, we identified Duffy antigen receptor for chemokines (*DARC*) as a potential pleiotropic gene underlying both LBM and AAM.

1 Materials and methods

1.1 Subjects

1.1.1 The discovery GWAS subjects

The discovery study sample consisted of 825 unrelated female Han Chinese subjects, living in the cities and neighboring areas of Changsha and Xi'an in China. The study was approved by the Changsha and Xi'an institutional review boards. After signing an informed consent, all subjects received assistance in completing a structured questionnaire that included questions regarding anthropometric variables, lifestyle, diet, family information, and medical history.

Detailed medical information, such as menstrual history, was recorded using a nurse-administered questionnaire. AAM was calculated as the date of menarche minus the date of birth (in years, rounded to the first decimal place). The AAM followed approximately a normal distribution as verified by the Kolmogorov-Smirnov test with the software Minitab (Minitab, Inc., State College, PA, USA).

Lean mass was measured using a dual-energy X-ray absorptiometry (DXA) scanner Hologic QDR 4500W (Hologic Inc., Bedford, MA, USA), following the manufacturer's protocol. A DXA scan can accurately measure total body and regional bone mass, fat mass, and fat-free mass. Lean mass was calculated by subtracting bone mass from fat-free mass. After removal of all metals, the subject lay on a bed and was scanned from head to toe. Whole body and sub-region compositions, such as head, trunk, and limb, were measured by the DXA scanner. To ensure the quality of collected data, all scans were conducted, reviewed, and analyzed by a clinical expert. Body weight, height, and age were obtained on the same visit. In this study, lean limb mass was analyzed.

After excluding subjects with missing values for either the AAM or LBM phenotypes, the final sample that was used in the association analysis consisted of 801 unrelated Chinese female subjects.

1.1.2 The replication GWAS subjects

Significant findings from the discovery study were validated in a replication study. The replication sample consisted of 1728 female subjects of a Caucasian population recruited from the cities and neighboring areas of Omaha, NE and Kansas, MO, in the US. The recruitment was approved by the Institutional Review Boards of Creighton University and the University of Missouri-Kansas City. Signed informed consent documents were obtained from all study participants before they entered the study.

In the replication subjects, the AAM and LBM were both measured in the same way as in the discovery subjects. After excluding subjects with missing phenotypes, the sample size for data analysis was 1692.

1.2 Genome-wide genotyping and quality controls

1.2.1 Discovery sample

Genomic DNA was extracted from peripheral blood leukocytes using standard protocols. The Affymetrix genome-wide human SNP array 6.0 (Affymetrix, Santa Clara, CA, USA), which includes 906600 SNPs and 940000 copy number probes, was used to genotype the SNPs in the genome of each of the subjects, according to the manufacturer's protocol. Briefly, approximately 250 ng of genomic DNA was digested with a restriction enzyme, either *Nsp* I or *Sty* I. The digested DNA from each sample was adaptor-ligated and PCR-amplified. The PCR products were labeled with biotin, denatured, and hybridized to the arrays. Arrays were then washed and stained using phycoerythrin on an Affymetrix Fluidics Station, and scanned using the GeneChip Scanner 3000 7G to quantitate fluorescence intensities. Data management and analyses were conducted using the Genotyping Command Console. Only samples with a minimum call rate of 95% were included. Because

repeated experiments were performed, all of the subjects (n=801) met this criteria and the final mean call rate reached the high level of 98.93%.

1.2.2 Replication sample

Genotyping of the replication samples was also performed using the Affymetrix SNP 6.0 genotyping array following the same protocols as those described in Section 1.2.1 for the discovery samples. All of the subjects (n=1692) met the 95% minimum call rate criterion and the final mean call rate reached the high level of 98.22%.

1.3 Statistical analyses

For each AAM and lean mass measurement, a stepwise regression analysis was performed to screen the covariate effects of age, age^2, height, and weight. Age was significant for AAM, while age, height, and weight were significant for lean mass (P<0.05). The significant covariates were used to adjust the raw phenotype values.

To correct for potential population stratification, we performed principal component analysis (PCA) on the genome-wide SNP data with EIGENSTRAT [30]. The top five principal components (PCs) were used to adjust the phenotype to minimize the effect of population stratification.

In the discovery sample, the association analyses between genotype and the two adjusted phenotypes were performed with a bivariate linear regression model. The model can be represented as

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\beta} x_i + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{y}_i$ is the vector of the two phenotypes for individual i; $\boldsymbol{\mu}$ is the vector of grand means; x_i is the genotype score for individual i, and $\boldsymbol{\beta}$ is the vector of its effects. Finally, $\boldsymbol{\varepsilon}_i$ is the vector of residues following a multivariate normal distribution with mean zero. The genotype score x_i was encoded with an additive mode of inheritance. The association was examined by testing $\boldsymbol{\beta}$ from zero, and the test was performed with the linear regression model in R package *lm*.

For comparison purposes, we also performed univariate association analyses for the tested phenotypes, using the univariate linear regression model in the same R package *lm*.

In the replication sample, SNPs that were suggestively associated with the AAM and LBM phenotypes (P<1×10^{-5}) were further tested for association. To quantify the overall evidence of association achieved in the discovery and replication samples, Fisher's method was used to combine the two individual P-values.

2 Results

The basic phenotypic characteristics of the subjects in the discovery and replication samples are listed in Table 1. In the discovery sample, a total of 44 SNPs had suggestive bivariate association signals (P<1×10^{-5}). One of the SNPs, *rs3027009*, was also found to be significant in the replication sample; for the AAM and left-arm lean mass phenotypes, P=7.26×10^{-6} in the discovery sample, and P=0.005 in the replication sample (Table 2).

The SNP *rs3027009* is located 1.2 kb upstream from the Duffy blood group, chemokine receptor (*DARC*) gene. Compared with the bivariate associations for this SNP, the univariate associations had much higher P-values, although they were significant (Table 2) too at the nominal level 0.05. The combined bivariate P-value was 7.14×10^{-7}, close to the genome-wide significance level (Table 2).

3 Discussion

Here, we performed the first genome-wide bivariate associ-

Table 1 Basic characteristics of the study subjects[a)]

	Discovery	Replication
Trait	n=801	n=1692
Age (years)	37.68 (13.83)	51.59 (12.92)
Height (cm)	158.34 (5.21)	163.28 (6.27)
Weight (kg)	54.68 (8.17)	71.45 (16.04)
Left-arm lean (kg)	1.74 (0.32)	2.34 (0.52)
Right-arm lean (kg)	1.91 (0.37)	2.47 (0.53)
Left-leg lean (kg)	6.01 (0.79)	7.64 (1.37)
Right-leg lean (kg)	6.09 (0.82)	7.76 (1.39)
Age at menarche (years)	13.91 (1.61)	12.92 (1.58)

a) Data are presented as mean (SD).

Table 2 Characteristics of SNP *rs3027009*[a)]

Sample	SNP	Chr	Position	Alleles	MAF	Univariate P-value		Bivariate P-value
						LA lean	AAM	LA lean AAM
Discovery					0.01	2.51×10^{-3}	6.16×10^{-4}	7.26×10^{-6}
Replication	rs3027009	1	159173887	A/G	0.08	0.64	1.45×10^{-3}	5.45×10^{-3}
Combined						0.01	1.33×10^{-5}	7.14×10^{-7}

a) The univariate and bivariate association results are shown. Combined P-values were calculated using Fisher's method. Chr, Chromosome; MAF, minor allele frequency; LA, left-arm; AAM, age at menarche.

ation study of LBM and AAM in a Chinese population sample. We further validated our findings in an additional Caucasian population sample. Our findings identified the *DARC* gene as a potential pleiotropic gene underlying both the LBM and AAM phenotypes.

Genome-wide association studies have become a popular approach because of their superior resolution compared with linkage studies. With the high levels of phenotypic correlations among many diseases and traits, adequate attentions should be given to multivariate approaches. However, compared with the fruitful methodological developments that have taken place in linkage studies, theoretical research in association studies is quite rare. Previously, Lange *et al*. [31] applied a generalized estimating equation (GEE) based approach to multivariate association analyses in pedigrees. Liu *et al*. [32] developed a population-based bivariate traits (one quantitative and one qualitative) association analysis and Pei *et al*. [33] developed a multivariate haplotype-trend association method. Zhang *et al*. [29] extended the GEE-based method to data containing a mixture of unrelated and familial subjects. However, the application of these methods to experimental data has not been well recognized. In a previous study, Liu *et al*. [34] performed the first genome-wide bivariate association studies on obesity and osteoporosis. The associated gene that they identified, *SOX6*, has been replicated by later studies, indicating the efficacy of their bivariate analyses.

SNP *rs3027009* that we found to be associated with LBM and AAM is located upstream of the *DARC* gene. The gene encodes a transmembrane glycoprotein of 35 kD with 336 amino acid residues. The protein was first reported by Cutbush *et al*. [35] to act as a chemokine receptor. *DARC* has been reported to be closely associated with a variety of human disorders, such as malaria [36], tumors [37], and AIDS [38], and some of the latest studies have suggested that *DARC* influences the growth and transformation of tumors through regulating the generation of tumor-relevant veins [37]. Edderkaoui *et al*. [39] reported that polymorphisms in the mouse *DARC* gene changed bone mineral density and influenced the risk to osteoporosis. Considering the high correlation between lean mass and bone mineral density, *DARC* may play a role in lean mass metabolism, though this relationship has not been reported previously. Our findings imply, for the first time, that the *DARC* gene is associated with LBM; however, the pathway through which *DARC* may regulate lean mass is unknown. The *DARC* protein is the main immune molecule on the surface of red blood cells [40]. Thus, we hypothesize that *DARC* affects the formation of the microvascular system in muscle tissue, which in turn affects the ingestion of the protein required for the synthesis of muscle.

Menarche is the signal of the sexual maturation and fertility in females. The process is controlled by the hypothalamic-pituitary-ovarian axis, and is related to the release of gonadotropin and estrogen. It was reported previously that the estrogen receptor (*ER*) genes and the vitamin D receptor (*VDR*) gene were associated with AAM [19,21]. *DARC* and the *ER*s are closely related. Wang *et al*. [41] reported that the expressions of *DARC* and that of the *ER*s are positively correlated in breast cancer tissue. Therefore, though direct evidence is still lacking, we hypothesize that *DARC* may influence AAM by its chemotaxis effect on the estrogen receptor pathway.

In summary, using a novel bivariate GWAS approach, we identified a gene, *DARC*, which may be important for both LBM and AAM. Replication of the association findings in the Caucasian sample confirmed the association. The *DARC* gene has potential functional implications in both the formation of skeletal muscle and in the regulation of the estrogen receptor, and may be an important pleiotropic gene of LBM and AAM. Further investigations are needed to study the biological function of *DARC* and to replicate the findings in larger-scale studies.

This work was supported by the Shanghai Leading Academic Discipline Project (Grant No. S30501) and by a start-up fund from the Shanghai University of Science and Technology, China. Deng HongWen was partially supported by grants from National Institutes of Health (Grant Nos. P50AR055081, R01AG026564, R01AR050496, RC2DE020756, R01AR057049, and R03TW008221) and the Franklin D. Dickson/Missouri Endowment and the Edward G. Schlieder Endowment. Zhao LanJuan was supported by the National Natural Science Foundation of China (Grant No. 31100902).

1 Sipila S, Heikkinen E, Cheng S, *et al*. Endogenous hormones, muscle strength, and risk of fall-related fractures in older women. J Gerontol A Biol Sci Med Sci, 2006, 61: 92–96
2 Karakelides H, Nair K S. Sarcopenia of aging and its metabolic impact. Curr Top Dev Biol, 2005, 68: 123–148
3 Hansen R D, Raja C, Aslani A, *et al*. Determination of skeletal muscle and fat-free mass by nuclear and dual-energy x-ray absorptiometry methods in men and women aged 51–84 y (1–3). Am J Clin Nutr, 1999, 70: 228–233
4 Hsu F C, Lenchik L, Nicklas B J, *et al*. Heritability of body composition measured by DXA in the diabetes heart study. Obes Res, 2005, 13: 312–319
5 Keen-Kim D, Mathews C A, Reus V I, *et al*. Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. Hum Mol Genet, 2006, 15: 3324–3328
6 Nguyen T V, Howard G M, Kelly P J, *et al*. Bone mass, lean mass, and fat mass: same genes or same environments? Am J Epidemiol, 1998, 147: 3–16
7 Wang X L, Deng F Y, Tan L J, *et al*. Bivariate whole genome linkage analyses for total body lean mass and BMD. J Bone Miner Res, 2008, 23: 447–452
8 Liu X G, Tan L J, Lei S F, *et al*. Genome-wide association and replication studies identified TRHR as an important gene for lean body mass. Am J Hum Genet, 2009, 84: 418–423
9 Kaaks R, Lukanova A, Kurzer M S. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. Cancer Epidemiol Biomarkers Prev, 2002, 11: 1531–1543
10 Paganini-Hill A, Henderson V W. Estrogen deficiency and risk of Alzheimer's disease in women. Am J Epidemiol, 1994, 140: 256–261
11 Silman A J. Risk factors for Colles' fracture in men and women: results from the European Prospective Osteoporosis Study. Osteoporos Int, 2003, 14: 213–218
12 Roy D K, O'Neill T W, Finn J D, *et al*. Determinants of incident vertebral fracture in men and women: results from the European Prospective

Osteoporosis Study (EPOS). Osteoporos Int, 2003, 14: 19–26
13 Yang T L, Chen X D, Guo Y, *et al*. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am J Hum Genet, 2008, 83: 663–674
14 Treloar S A, Martin N G. Age at menarche as a fitness trait: nonadditive genetic variance detected in a large twin sample. Am J Hum Genet, 1990, 47: 137–148
15 van den Berg S M, Boomsma D I. The familial clustering of age at menarche in extended twin families. Behav Genet, 2007, 37: 661–667
16 Anderson C A, Duffy D L, Martin N G, *et al*. Estimation of variance components for age at menarche in twin families. Behav Genet, 2007, 37: 668–677
17 Anderson C A, Zhu G, Falchi M, *et al*. A genome-wide linkage scan for age at menarche in three populations of European descent. J Clin Endocrinol Metab, 2008, 93: 3965–3970
18 Kaprio J, Rimpela A, Winter T, *et al*. Common genetic influences on BMI and age at menarche. Hum Biol, 1995, 67: 739–753
19 Long J R, Xu H, Zhao L J, *et al*. The oestrogen receptor alpha gene is linked and/or associated with age of menarche in different ethnic groups. J Med Genet, 2005, 42: 796–800
20 Stavrou I, Zois C, Chatzikyriakidou A, *et al*. Combined estrogen receptor alpha and estrogen receptor beta genotypes influence the age of menarche. Hum Reprod, 2006, 21: 554–557
21 Stavrou I, Zois C, Ioannidis J P, *et al*. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. Hum Reprod, 2002, 17: 1101–1105
22 Xita N, Tsatsoulis A, Stavrou I, *et al*. Association of SHBG gene polymorphism with menarche. Mol Hum Reprod, 2005, 11: 459–462
23 Zhao J, Xiong D H, Guo Y, *et al*. Polymorphism in the insulin-like growth factor 1 gene is associated with age at menarche in caucasian females. Hum Reprod, 2007, 22: 1789–1794
24 Yang F, Xiong D H, Guo Y, *et al*. The chemokine (C-C-motif) receptor 3 (CCR3) gene is linked and associated with age at menarche in Caucasian females. Hum Genet, 2007, 121: 35–42
25 Gorai I, Tanaka K, Inada M, *et al*. Estrogen-metabolizing gene polymorphisms, but not estrogen receptor-alpha gene polymorphisms, are associated with the onset of menarche in healthy postmenopausal Japanese women. J Clin Endocrinol Metab, 2003, 88: 799–803
26 Guo Y, Xiong D H, Yang T L, *et al*. Polymorphisms of estrogen-biosynthesis genes CYP17 and CYP19 may influence age at menarche: a genetic association study in Caucasian females. Hum Mol Genet, 2006, 15: 2401–2408
27 Lai J, Vesprini D, Chu W, *et al*. CYP gene polymorphisms and early menarche. Mol Genet Metab, 2001, 74: 449–457
28 Zhang L, Bonham A J, Li J, *et al*. Family-based bivariate association tests for quantitative traits. PLoS ONE, 2009, 4: e8133
29 Zhang L, Pei Y F, Li J, *et al*. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. PLoS ONE, 2009, 4: e6502
30 Price A L, Patterson N J, Plenge R M, *et al*. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet, 2006, 38: 904–909
31 Lange C, Silverman E K, Xu X, *et al*. A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics, 2003, 4: 195–206
32 Liu J, Pei Y, Papasian C J, *et al*. Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. Genet Epidemiol, 2009, 33: 217–227
33 Pei Y F, Zhang L, Liu J, *et al*. Multivariate association test using haplotype trend regression. Ann Hum Genet, 2009, 73: 456–464
34 Liu Y Z, Pei Y F, Liu J F, *et al*. Powerful bivariate genome-wide association analyses suggest the SOX6 gene influencing both obesity and osteoporosis phenotypes in males. PLoS ONE, 2009, 4: e6827
35 Cutbush M, Mollison P L, Parkin D M. A new human blood group. Nature, 1950, 165: 188–189
36 Miller L H, Mason S J, Dvorak J A, *et al*. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood Group determinants. Science, 1975, 189: 561–563
37 Shen H, Schuster R, Stringer K F, *et al*., The Duffy antigen/receptor for chemokines (DARC) regulates prostate tumor growth. FASEB J, 2006, 20: 59–64
38 He W, Neil S, Kulkarni H, *et al*. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. Cell Host Microbe, 2008, 4: 52–62
39 Edderkaoui B, Baylink D J, Beamer W G, *et al*. Identification of mouse Duffy antigen receptor for chemokines (Darc) as a BMD QTL gene. Genome Res, 2007, 17: 577–585
40 Durpes M C, Hardy-Dessources M D, El Nemer W, *et al*. Activation state of alpha4beta1 integrin on sickle red blood cells is linked to the duffy antigen receptor for chemokines (DARC) expression. J Biol Chem, 2011, 286: 3057–3064
41 Wang J, He Q, Shao Y G, *et al*. Duffy antigen receptor for chemokines expression is related with ER expression in primary lesion of breast cancer. Chin J Clinic Med, 2009, 16: 631–633