
**BENTHAM
SCIENCE**

A Genome-wide Association Analysis in Four Populations Reveals Strong Genetic Heterogeneity For Birth Weight


 Tiane Luo¹, Xu Liu² and Yuehua Cui^{1,2,*}

¹Division of Health Statistics, School of Public Health, Shanxi Medical University, Shanxi, 030001, China; ²Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

Abstract: Low or high birth weight is one of the main causes for neonatal morbidity and mortality. They are also associated with adulthood chronic illness. Birth weight is a complex trait which is affected by baby's genes, maternal environments as well as the complex interactions between them. To understand the genetic basis of birth weight, we reanalyzed a genome-wide association study data set which consists of four populations, namely Thai, Afro-Caribbean, European, and Hispanic population with regular linear models. In addition to fit the data with parametric linear models, we fitted the data with a nonparametric varying-coefficient model to identify variants that are nonlinearly modulated by mother's condition to affect birth weight. For this purpose, we used baby's cord glucose level as the mother's environmental variable. At the 10^{-5} genome-wide threshold, we identified 33 SNP variants in the Thai population, 26 SNPs in the Afro-Caribbean population, 18 SNPs in the European population, and 7 SNPs in the Hispanic population. Some of the variants are significantly modulated by baby's cord glucose level either linearly or nonlinearly, implying potential interactions between baby's gene and mother's glucose level to affect baby's birth weight. There is no overlap between variants identified in the four populations, indicating strong genetic heterogeneity of birth weight between the four ethnic groups. The findings of this study provide insights into the genetic basis of birth weight and reveal its genetic heterogeneity.

Keywords: Gene-environment interaction, Genetic association, Nonlinear modulation, Varying-coefficient model, Birth weight, Four population.

Received: March 31, 2015

Revised: August 24, 2015

Accepted: August 31, 2015

INTRODUCTION

Low and high birth weight are high risk factors for neonatal morbidity and mortality. In addition, studies have shown that low birth weight is associated with later life metabolic diseases, such as heart disease [1], type 2 diabetes [2], hypertension [3] and renal disease [4, 5]. High birth weight is also associated with health problems for the baby. Large babies face high risk of difficulty of labor and delivery, post neonatal death, obesity, diabetes as well as heart disease over a lifetime [6]. It is thus of paramount importance to understand the genetic basis of birth weight as well as how mother's conditions modulate baby's genes to affect birth weight.

Birth weight is a complex trait involving the function of fetal genes. The estimated heritability of birth weight is 25-40% [7]. In the meantime, fetal growth also depends on the nutrient exchange between mother and fetus. How fetal genes respond to mother's nutrition supply largely determines fetal growth and development, leading to complicated

interactions between fetal genes and mother's environmental conditions. Little and Sing [8] reported the fetal genetic and maternal intrauterine environmental influences on human birth weight adjusting for the effect of external environmental factors. The results show that the interactions between fetal genes and the maternal intrauterine environment play important roles in baby's birth weight. Identifying genetic variants that impacts fetal growth and to determine the interaction of such variants with the intrauterine environment is critical to understand the genetic basis of birth weight. Given the complexity of such modulation effect from the mother's side, it is essential to apply novel statistical strategies to dissect and further quantify the interaction mechanism.

So far the genetic loci that influence birth weight are largely unknown. The results from a meta-analysis of six European cohorts ($n=10,623$) genome-wide association studies (GWAS) show that variants in two loci (ADCY5 gene, near LEKR1 and CCNL1 gene) are associated with fetal growth and birth weight, and the two leading signals were further replicated in 13 replication studies ($n=27,591$) [9]. GWAS in Twins UK cohort ($n=2,997$ female) found variants close to gene NTRK2 are associated with birth weight in female twins [10] A few other loci were also reported in a

*Address correspondence to this author at the Division of Health Statistics, School of Public Health, Shanxi Medical University, Shanxi, 030001, P.R. China; Tel: (517) 432-7098; Fax: (517) 432-1405; E-mail: cui@stt.msu.edu

recent meta-GWAS study of European descents [11]. However, given the large heritability estimated from the family study [7], large efforts are still needed to identify more loci.

To date, there are increasing evidences showing that gene-environment (G×E) interaction plays an important role in complex diseases. G×E interaction is defined as the effect in which genotypic influence on phenotype changes as environment changes [12]. Examples of G×E on disease risk have been broadly reported in literature such as Parkinson's disease [13], type 2 diabetes [14], and mental illness [15]. Owing to the complexity of G×E interaction mechanisms in various diseases, the development of efficient and powerful statistical methods to dissect such effect is crucial. Ma *et al.* [16] first time proposed a nonlinear G×E interaction model for continuous quantitative responses. Wu and Cui [17] extended the model for binary disease responses. These methods are nonparametric in nature, hence have much flexibility to capture the underlying interaction mechanism, especially when nonlinear environmental modulation effects exist.

The aim of the study is to search for and compare genetic variants of birth weight in the four populations, namely Thai, Afro-Caribbean, European, and Hispanic population from the Gene Environment Association Studies initiative (GENEVA, <http://www.genevastudy.org>). In addition to identify genetic variants associated with birth weight, we are also interested in identifying variants that are sensitive to mother's environmental changes to affect birth weight. For this purpose, we focus baby's cord glucose level as the external modulator which directly reflects mother's glucose supply. Identifying genetic determinants of birth weight and testing the gene-environment interaction could give some clue on the prevention of low and high birth weight. Any identified heterogeneity between the four populations could also shed light on the prevention of chronic diseases associated with low or high birth weight.

MATERIALS AND METHODS

Study Populations

The data came from the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study - Maternal Glycemia and Birthweight GEI Study, funded by the trans-NIH Genes, Environment, and Health Initiative (GEI). We performed the genome-wide association analysis to identify common genetic variants of birth weight. The project includes four GWAS datasets that were collected from mothers and their offspring with birth weight as the phenotype. Single nucleotide polymorphisms (SNPs) data were measured on 1,500 infants of European descent, 1,250 Afro-Caribbean infants, 800 Hispanic (Mexican-American) infants, and 1200 Thai infants. The details about the data can be found in literature (HAPO Study Cooperative Research Group 2009).

Quality Control

We extracted the infants' SNP genotypes from the GWAS data. We excluded SNPs with minor allele frequency (MAF) less than 0.05, those with at least 10% missing genotype rate and those failed to pass the Hardy-Weinberg equilibrium test at the 0.001 significance level. After the quality control, the final data contain 512,912 SNPs in the European

set, 876,391 SNPs in the Afro-Caribbean set, 835,583 SNPs in the Hispanic set, and 683,938 SNPs in the Thai set.

Phenotype and Covariates Preparation

We converted the unit of birth weight from gram to kg, and excluded any potential outliers using the 1.5IQR (inter-quantile range) rule (values less than $Q_1 - 1.5IQR$ or large than $Q_3 + 1.5IQR$ were removed where Q_1 and Q_3 refer to the first and third quantile). We included covariates such as baby's gender, mother's mean OGTT diastolic blood pressure, gestational age, mother's mean OGTT body mass index (BMI) and baby's cord glucose level in our analysis as these variables were significant in each population when regressing the birth weight against all the covariates. The same outlier removing rule as described above was applied to the quantitative measures. Gender was recoded as 1=male and 0=female. A summary of the phenotype and covariate distribution was shown in (Table 1). The final data contain 1,331 individuals in the European set, with 657 males and 674 females; 1,074 individuals in the Afro-Caribbean set, with 544 males and 530 females; 601 individuals in the Hispanic set, with 297 males and 304 females; and 1,114 individuals in the Thai set, with 542 males and 572 females.

Statistical Methods

Birth weight is a complex trait which is affected by multiple factors. To identify genes associated with newborn baby's birth weight, we first constructed a simple linear regression model while adjusting for the effect of covariates, including mother's mean OGTT diastolic blood pressure, gestational age, mother's mean OGTT BMI and baby's gender. We also did the principal components analysis for SNPs in each population, then applied the first two principle components as covariates to adjust for the effect of potential population stratification. The linear regression model can be expressed as

$$Y = \mu + \beta^T X + \delta_1 Z_1 + \delta_2 Z_2 + \varepsilon,$$

where Y is the birth weight; μ is the intercept term; β are regression parameters of covariates; X are covariates including mother's mean OGTT diastolic blood pressure, gestational age, mother's mean OGTT BMI and baby's gender; Z_1 and Z_2 represent the first two principle components of SNPs in each population with the corresponding effects δ_1 and δ_2 respectively; and ε is the error term with mean 0 and variance σ^2 . After fitting the above regression model, we obtained the residuals (denoted as \tilde{Y}) and treated them as the covariates' adjusted responses to do a genome-wide scan to assess each SNP's effect.

Baby's cord blood glucose level reflects mother's glucose supply to baby, which directly affects fetal growth and development [18]. In addition to identify main SNP effect, in this study, we were also interested in investigating how mother's glucose supply affects fetal growth and further assessing how fetal genes are modulated by mother's glucose level. Thus, we treated baby's cord blood glucose level (denoted as U) as the mother's environmental variable of interest. Let G be the genetic variable of interest which is coded as 0, 1, 2 corresponding to genotype cc, Cc, CC, where allele C is the minor allele.

Table 1. Descriptive characteristics of subjects in the four populations.

Characteristic	European	Afro-Caribbean	Hispanic	Thai	p-value
Sample size	1331	1074	601	1114	-
Ratio (M:F)	657:674	544:530	297:304	542:572	-
No. of SNPs	512,912	876,391	835,583	683,938	-
MBW (kg)	3.42±0.49	3.22±0.43	3.44±0.42	3.09±0.37	<0.001
b_CordPGC(mg)	79.00±15.32	83.47±14.53	77.83±14.80	88.28±19.98	<0.001
Baby Gestage	39.92±1.17	39.79±1.22	39.69±1.15	39.36±1.24	<0.001
m_DBPM_OGTT	71.40±8.10	67.20±7.97	72.00±7.92	67.65±7.68	<0.001
m_BMI	28.51±4.84	27.73±6.09	30.10±5.65	21.82±3.50	<0.001

Data are shown as mean ± SD; MBW=mean birth weight; b_CordPGCmg=Baby's cord glucose level; Baby Gestage: Baby's gestational age; m_DBPM_OGTT=Mother's mean OGTT diastolic blood pressure; m_MBI=mother's body mass index; Ratio: the ratio of male:female. P-value is for testing the mean difference among the four population based on a one-way ANOVA analysis.

We constructed three models to assess the impact of glucose level on baby's birth weight. The first model we considered is a simple linear regression model in which no G×E interaction was assumed, namely the genetic and environmental factors affecting birth weight independently. The model has the following form

$$\tilde{Y} = \alpha_0 + \alpha_1 U + \gamma G + \epsilon, \quad (1)$$

where the covariates' adjusted residual \tilde{Y} represents the new phenotypic response; α_0 is the overall mean; α_1 is the marginal effect of glucose (U); γ is the effect of genetic variable (G); and ϵ is the error term. We test $H_0: \gamma = 0$ to assess a genetic effect. If the null is rejected, then we claim that the genetic marker is associated with the phenotype.

In reality, babies carrying the same genotype may have different birth weights if the amounts of glucose supply from different mothers are different. The phenomenon could be explained partially by interactions between baby's genome and mother's environmental conditions. Assuming there is a linear interaction relationship between baby genes and mother's environment, we constructed the following linear regression model to detect such interaction, i.e.,

$$\tilde{Y} = \alpha_0 + \alpha_1 U + \gamma_0 G + \gamma_1 UG + \epsilon, \quad (2)$$

where γ_0 and γ_1 represent the effects of genetic variable G and the interaction of G and U , respectively. Model (2) can also be written as $\tilde{Y} = \alpha_0 + \alpha_1 U + (\gamma_0 + \gamma_1 U)G + \epsilon$ such that the effect of G on \tilde{Y} is a linear function of U , the so called linear G×E interaction model. We test $H_0: \gamma_0 = \gamma_1 = 0$ to assess a genetic effect. If the null is rejected, then we claim that there exist linear G×E interactions.

Nonlinear G×E interactions are often appeared in nature. When such nonlinear interaction effect exists, fitting a linear interaction model may suffer from power loss, as the simulation studies showed in [16]. To avoid missing potential interaction signals, we relaxed the linearity assumption and fitted a nonlinear interaction model proposed in [16]. Such a model

has the potential to identify nonlinear modulation effect of glucose on baby's gene to affect birth weight. The varying-coefficient (VC) model assuming a nonlinear G×E interaction has the following form

$$\tilde{Y} = \alpha(U) + \gamma(U)G + \epsilon, \quad (3)$$

where $\alpha(U)$ and $\gamma(U)$ are smooth functions; $\alpha(U)$ represents the marginal effect of U and $\gamma(U)$ represents the interaction effect of U which is the parameter of interest. If we further assume a linear intercept function for $\alpha(U)$, i.e., $\alpha(U) = \alpha_0 + \alpha_1 U$, then when $\gamma(U) = \gamma$, i.e., the varying-coefficient is a constant and does not change as U changes, model (3) reduces to model (1). When $\gamma(U) = \gamma_0 + \gamma_1 U$, i.e., the effect of G on \tilde{Y} changes linearly in U , model (3) reduces to the linear interaction model in (2). Thus, both models (1) and (2) are special cases of model (3). By relaxing the structure of $\alpha(U)$ and $\gamma(U)$ which can be estimated with nonparametric techniques, model (3) has much flexibility to capture potential nonlinear interaction effects. We test $H_0: \gamma(\cdot) = 0$ to assess a genetic effect. The details about the testing procedure can be found in [16].

The work of [16] has demonstrated that if the data were fitted with a model different from the true model, potential power loss is expected. The case was even worse when the underlying truth is nonlinear as in model (3), but fitted with linear interaction model (2) or no interaction model (1). Given the complex nature of modulation effect of glucose on baby's genes, we expect to identify SNP variants nonlinearly modulated by mother's glucose level in addition to SNPs identified with regular linear regression models. In our analysis, the implementation of the linear models was done with the PLINK software [19]. The implementation of the varying-coefficient nonlinear interaction model was done using the statistical software R. The computational code can be downloaded at <http://www.stt.msu.edu/~cui/Software.html> with the package named **VCGE**.

RESULTS

(Table 1) summarizes the descriptive statistics of the four populations after the quality control process described in the previous section. One way ANOVA test shows that there are mean birth weight differences among the four populations. The mean birth weight in the Thai population is significantly lower than the other populations. ANOVA tests also show significant mean differences between the four populations for other covariates except for baby's gestational age.

We did a systematic evaluation of the genetic basis of birth weight in four different populations, considering potential linear or nonlinear G×E interactions. (Figs. 1-2) show the genome-wide Manhattan plots of the signals in the four populations fitted with the three models. The QQ plots of $-\log_{10}(\text{p-values})$ show no significant deviation from the expected diagonal line (see the supplementary material), indicating no potential inflation of false positives.

Summary of the association analysis results are shown in (Tables 2-5). Columns 1-5 list the SNP ID, the chromosome, corresponding gene symbol, or nearest annotated gene (+/-500kb) and the position each SNP belongs to (according to

the NCBI Build v38), followed by the alleles as well as the minor allele frequency. Column 8 lists the p-values for testing the SNP effect with the corresponding fitted model shown in the last column which was described in the previous section. Specifically, Model 1 is the linear model which assumes no interaction between fetal genes and intrauterine environment. Model 2 refers to model (2) which assumes a linear G×E interaction between fetal genes and intrauterine environment (i.e., cord glucose level in specific); and Model 3 refers to the VC model in model (3) which assumes a non-linear G×E interaction between fetal genes and intrauterine environment. SNPs selected by each model ($\text{p-value} < 10^{-5}$) are ordered by chromosome. Note that the three models test different types of genetic effects. Combining the three sets of p-values increase the number of multiple testing. Thus, we used a less stringent threshold here. If we used the Bonferroni adjusted threshold, none of the SNPs passed this level. Thus, all the listed SNPs are suggestive SNPs rather than statistically significant ones in a Bonferroni adjusted sense. In total, we identified 33 SNPs in the Thai population, 4 SNPs were selected by fitting Model (1), 13 SNPs were selected by fitting Model (2), 16 SNPs were selected by fitting

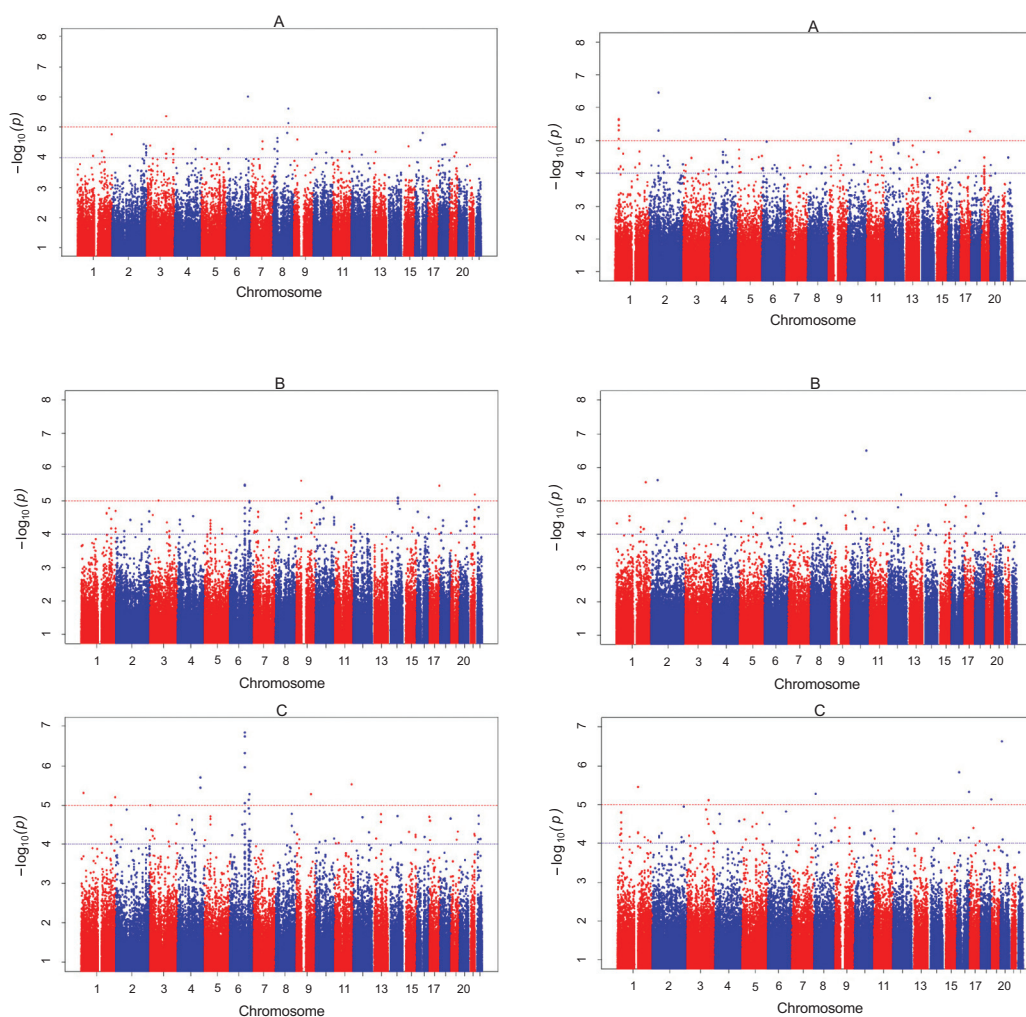


Fig. (1). The Manhattan plot of $-\log_{10}(\text{p-values})$ for the Thai (left panel) and Afro-Caribbean (right panel) population fitted with three models. **A:** $-\log_{10}(\text{p-values})$ fitted with the linear model (1); **B:** $-\log_{10}(\text{p-values})$ fitted with the linear G×E interaction model (2); **C:** $-\log_{10}(\text{p-values})$ fitted with the nonlinear G×E interaction model (3).

Model (3); 26 SNPs in the Afro-Caribbean population, 12 SNPs were selected by fitting Model (1), 7 SNPs were selected by fitting Model (2), 7 SNPs were selected by fitting Model (3); 18 SNPs in the European population, 5 SNPs were selected by fitting Model (1), 6 SNPs were selected by fitting Model (2), 7 SNPs were selected by fitting Model (3); and 7 SNPs in the Hispanic population, 5 SNPs were selected by fitting Model (1), no SNPs were selected by fitting Model (2), 2 SNPs were selected by fitting Model (3). In the Thai population, most identified SNPs show linear or nonlinear modulation effect by glucose, indicating strong gene \times glucose interaction effect in affecting birth weight. Similar pattern was found in the European population. Those signals identified by model (3) could be missed if only linear models were considered. For comparison purpose, we also listed the p-values in other populations for SNPs showing suggestive significance in the tested population (see Table S1-S4 in the Supplementary file for details).

Among the SNPs listed in the tables, some have been reported in literature. For example, SNPs in gene methionine sulfoxide reductase A (MSRA) identified in the Afro-Caribbean population (Table 3) show nonlinear G \times glucose interactions under the VC model. This gene has been observed to show high level of expression in fetal liver tissue

[20]. Scherag *et al.* [21] discovered two new obesity loci in extremely obese children and adults, one is SDCCAG8 and the other is between genes TNKS and MSRA. We observed a few consistent SNP signals in gene TMEM57 in the Afro-Caribbean population. These SNPs are not sensitive to glucose level and show no sign of G \times glucose interaction. We also identified SNP markers rs1040193 in gene IL1A in the Hispanic population with nonlinear interaction effect (Table 5). Association study has shown that SNPs in IL1A are associated with preterm birth and low birth weight in a Japanese population [22]. One SNP (rs10924366) identified in gene SMYD3 in the European population almost reached the genome-wide Bonferroni adjusted threshold with a p-value of 8.99×10^{-8} (Table 4). Gene SMYD3 is a histone methyltransferase and it plays a role in transcriptional regulation. Such a regulation role may be associated with glucose metabolism and further affect birth weight. Further biological investigation is needed to verify its biological function.

We plotted the estimated varying-coefficient function $\hat{\nu}(U)$ against baby's cord glucose level (U) for SNP rs1040193 located in gene *IL1A* in the Hispanic population (see Fig. 3), as an example to show the nonlinear modulation effect of cord glucose level on genetic influence on birth

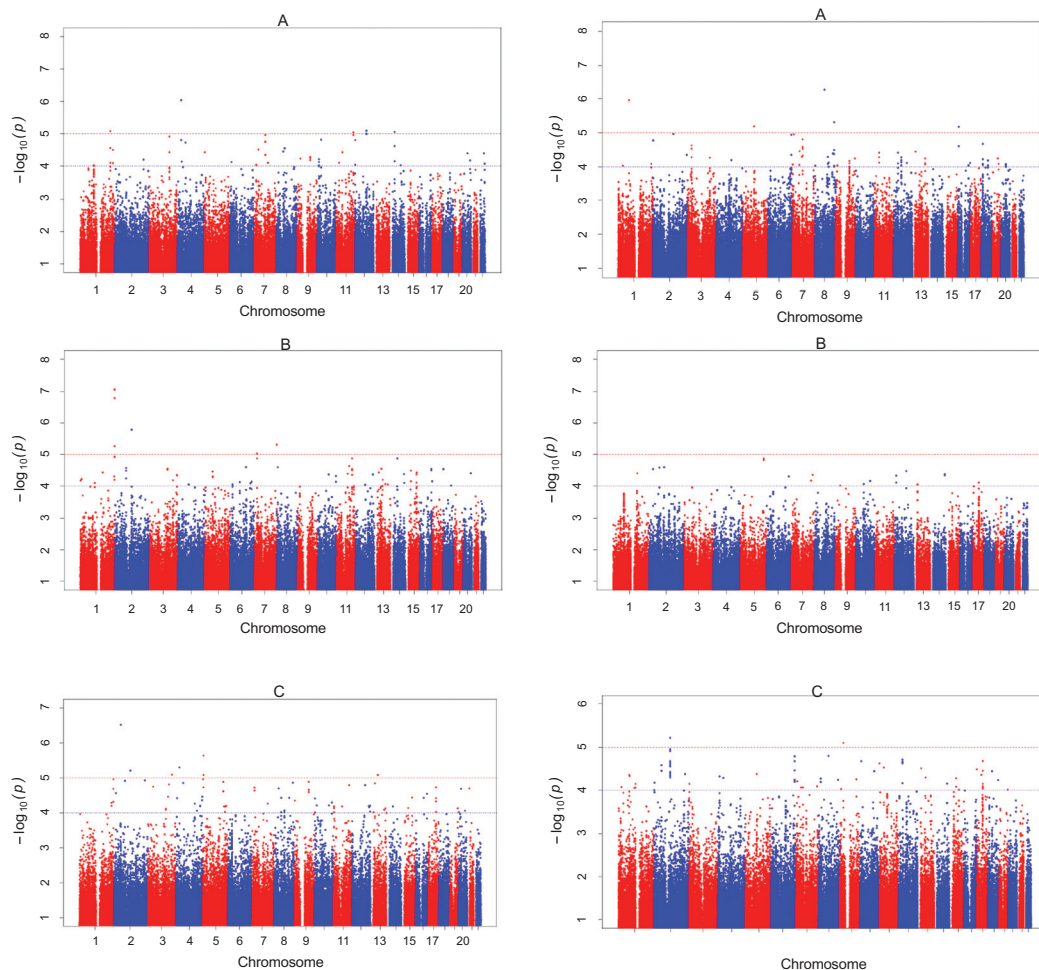


Fig. (2). The Manhattan plots of $-\log_{10}(p)$ -values for the European (left panel) and Hispanic (right panel) population fitted with three models. **A:** $-\log_{10}(p)$ -values fitted with the linear model (1); **B:** $-\log_{10}(p)$ -values fitted with the linear G \times E interaction model (2); **C:** $-\log_{10}(p)$ -values fitted with the nonlinear G \times E interaction model (3).

Table 2. List of SNPs with p-values $<1 \times 10^{-5}$ in the Thai population.

SNP_ID	CHR ^a	Position ^b	Gene Symbol	Nearest Gene ^c	Allele ^d	MAF	p-value	Model ^e
rs17350052	3	140555600	CLSTN2		G/A	0.227	4.33×10^{-6}	1
rs1482394	6	148895942	UST		A/G	0.483	9.80×10^{-7}	1
rs3019943	8	106974707	-	HMGB1P46	A/G	0.113	7.39×10^{-6}	1
rs1564018	8	107038788	-	HMGB1P46	A/G	0.112	2.41×10^{-6}	1
rs17667547	3	64472771	-	LOC101929316	A/G	0.105	9.99×10^{-6}	2
rs949882	6	109264859	LOC100996634		A/G	0.342	3.37×10^{-6}	2
rs9374069	6	109265131	LOC100996634		G/A	0.341	3.53×10^{-6}	2
rs4879913	9	35610915	CD72		A/G	0.395	2.57×10^{-6}	2
rs1547842	10	113641000	NRAP		G/A	0.323	8.08×10^{-6}	2
rs3127106	10	113645856	NRAP		A/G	0.323	8.50×10^{-6}	2
rs3121478	10	113645905	NRAP		G/A	0.323	7.88×10^{-6}	2
rs3121487	10	113648920	NRAP		C/A	0.329	8.66×10^{-6}	2
rs2252463	14	72597321	-	DPF3	A/T	0.145	8.15×10^{-6}	2
rs8018050	14	73340527	NUMB		A/G	0.094	8.45×10^{-6}	2
rs4527079	17	78372680	LOC101928674		A/G	0.285	3.61×10^{-6}	2
rs944422	21	45562421	-	SLC19A1	G/A	0.432	6.66×10^{-6}	2
rs9647239	21	45592539	-	SLC19A1	G/A	0.372	6.65×10^{-6}	2
rs12057431	1	15065954	KAZN		G/A	0.083	4.94×10^{-6}	3
rs7517282	1	15067290	KAZN		A/G	0.083	4.94×10^{-6}	3
rs7531373	1	15067428	KAZN		G/A	0.083	4.94×10^{-6}	3
rs4256853	1	245356816	KIF26B		G/A	0.179	6.27×10^{-6}	3
rs10021001	4	162482639	-	TOMM22P4	A/G	0.497	3.64×10^{-6}	3
rs4521302	4	162484961	-	TOMM22P4	A/C	0.487	1.99×10^{-6}	3
rs9384701	6	109236339	LOC100996634		A/G	0.202	9.04×10^{-6}	3
rs9386780	6	109263207	LOC100996634		G/A	0.363	1.11×10^{-6}	3
rs949882	6	109264859	LOC100996634		A/G	0.342	1.43×10^{-7}	3
rs9374069	6	109265131	LOC100996634		G/A	0.341	1.82×10^{-7}	3
rs13210693	6	109277761	-	LOC100996634	A/G	0.361	4.81×10^{-7}	3
rs6910119	6	109278060	-	LOC100996634	G/A	0.361	4.81×10^{-7}	3
SNP6-138121923	6	-	-		A/G	0.172	7.35×10^{-6}	3
rs199256	6	143007986	LINC01277		A/G	0.290	5.29×10^{-6}	3
rs765399	9	105160922	-	LOC101928609	G/A	0.147	5.30×10^{-6}	3
rs17126029	11	122022674	-	MIR100HG	A/G	0.250	3.01×10^{-6}	3

^aCHR: chromosome; ^bPositions according to Build 38; ^cGenes within ± 500 kb of the lead SNP; ^dAllele (minor/major); ^eModel 1 refers to model (1) with the corresponding p-values obtained by testing $H_0: \gamma = 0$; Model 2 refers to model (2) with the corresponding p-values obtained by testing $H_0: H_0: \gamma_1 = \gamma_2 = 0$; Model 3 refers to model (3) with the corresponding p-values obtained by testing $H_0: \gamma(\cdot) = 0$.

Table 3. List of SNPs with p-value $<1 \times 10^{-5}$ in the Afro-Caribbean population.

SNP_ID	CHR ^a	Position ^b	Gene Symbol	Nearest Gene ^c	Allele ^d	MAF	p-value	Model ^e
rs6699113	1	25451252	TMEM57		A/G	0.222	3.46×10^{-6}	1
rs7554255	1	25451615	TMEM57		A/G	0.227	4.85×10^{-6}	1
rs35614701	1	25471632	TMEM57		G/A	0.214	2.25×10^{-6}	1
rs35589882	1	25480604	TMEM57		A/G	0.214	2.25×10^{-6}	1
rs35225089	1	25487129	TMEM57		A/G	0.214	2.39×10^{-6}	1
rs35886763	1	25502267	TMEM57		A/G	0.212	8.86×10^{-6}	1
rs13391261	2	64429275	-	LGALS1	A/G	0.423	3.46×10^{-7}	1
rs9309360	2	64436035	-	LGALS1	A/G	0.211	4.99×10^{-6}	1
rs7698522	4	105197599	TET2		G/A	0.200	9.07×10^{-6}	1
rs2371228	12	96702711	C12orf55		C/A	0.072	8.92×10^{-6}	1
rs10132619	14	77339033	TMED8		G/A	0.150	5.18×10^{-7}	1
rs8077382	17	76629604	ST6GALNAC1		G/A	0.136	5.22×10^{-6}	1
rs2185385	1	210757467	KCNH1		G/A	0.170	2.77×10^{-6}	2
rs13394954	2	47098890	C2orf61		C/A	0.069	2.41×10^{-6}	2
rs2804613	10	112091491	-	GPAM	A/G	0.112	3.24×10^{-7}	2
rs832508	12	94277635	CCDC41/PLXNC1		G/A	0.420	6.58×10^{-6}	2
rs7195627	16	23269875	-	SCNN1B	G/A	0.484	7.62×10^{-6}	2
rs4141733	20	14583273	MACROD2-IT1		G/A	0.088	5.77×10^{-6}	2
rs6042824	20	14618543	MACROD2		A/C	0.168	7.21×10^{-6}	2
rs12083119	1	145983307	POLR3GL		A/G	0.121	3.53×10^{-6}	3
rs2174747	3	153395183	-	C3orf79	G/A	0.253	7.81×10^{-6}	3
rs13251198	8	10324332	MSRA		A/C	0.405	5.35×10^{-6}	3
rs4148375	16	16118445	ABCC1		G/A	0.405	1.46×10^{-6}	3
rs7350878	16	86633582	-	FOXL1	G/A	0.085	4.67×10^{-6}	3
rs2654179	18	77763190	-	LOC100421527	G/A	0.159	7.46×10^{-6}	3
rs6086702	20	8945555	-	RNU105B	A/G	0.314	2.37×10^{-7}	3

See Table 2 for the explanation of the table header notation.

Table 4. List of SNPs with p-value $<1 \times 10^{-5}$ in the European population.

SNP_ID	CHR ^a	Position ^b	Gene Symbol	Nearest Gene ^c	Allele ^d	MAF	p-value	Model ^e
rs532342	1	216125293	USH2A		A/G	0.177	8.16×10^{-6}	1
rs2309558	4	27308619	-	RP11-415C15.2	C/A	0.108	9.21×10^{-7}	1
rs1012849	11	123478760	GRAMD1B		G/A	0.247	8.83×10^{-6}	1

(Table 4) contd....

SNP_ID	CHR ^a	Position ^b	Gene Symbol	Nearest Gene ^c	Allele ^d	MAF	p-value	Model ^e
rs2805	12	81256730	ACSS3		A/G	0.247	8.05 × 10 ⁻⁶	1
rs1475067	14	33547565	NPAS3		A/G	0.413	8.66 × 10 ⁻⁶	1
rs9660719	1	245840369	SMYD3		A/G	0.115	5.49 × 10 ⁻⁶	2
rs10924366	1	245864108	SMYD3		G/A	0.081	8.99 × 10 ⁻⁸	2
rs10924373	1	245875929	SMYD3		G/A	0.124	1.66 × 10 ⁻⁷	2
rs17732795	2	116759042	-	LOC100533709	G/A	0.159	1.66 × 10 ⁻⁶	2
rs2158493	7	18999023	HDAC9		G/A	0.406	9.27 × 10 ⁻⁶	2
rs896767	7	158280211	PTPRN2		A/G	0.246	4.82 × 10 ⁻⁶	2
rs860133	2	49351798	-	FSHR	C/A	0.192	3.02 × 10 ⁻⁷	3
rs17732795	2	116759042	-	LOC100533709	G/A	0.159	6.26 × 10 ⁻⁶	3
rs10490783	3	169322406	MECOM		A/G	0.057	8.06 × 10 ⁻⁶	3
rs2309558	4	27308619	-	RP11-415C15.2	C/A	0.108	5.05 × 10 ⁻⁶	3
rs6862164	5	6944154	LOC102724959		A/G	0.276	8.35 × 10 ⁻⁶	3
rs12520925	5	6961740	LOC102724959		G/A	0.401	2.32 × 10 ⁻⁶	3
rs9535618	13	51209178	-	FAM124A	G/A	0.110	8.24 × 10 ⁻⁶	3

See Table 2 for the explanation of the table header notation.

Table 5. List of SNPs with p-value < 1 × 10⁻⁵ in the Hispanic population.

SNP_ID	CHR ^a	Position ^b	Gene Symbol	Nearest Gene ^c	Allele ^d	MAF	p-value	Model ^e
rs2221083	1	76755677	-	TPI1P1	A/G	0.079	1.09 × 10 ⁻⁶	1
rs12186353	5	81880751	-	SHFM1P1	A/C	0.224	6.40 × 10 ⁻⁶	1
rs10110416	8	68268607	-	RPL31P40	A/G	0.053	5.28 × 10 ⁻⁷	1
rs6578225	8	135226451	-	RP11-452N4.1	A/G	0.411	4.82 × 10 ⁻⁶	1
rs11644531	16	5958823	RBFOX1		C/A	0.181	6.56 × 10 ⁻⁶	1
rs1040193	2	112772522	IL1A		A/G	0.228	6.11 × 10 ⁻⁶	3
rs10757553	9	25576242	-	TUSC1	C/A	0.477	7.94 × 10 ⁻⁶	3

See Table 2 for the explanation of the table header notation.

weight. The estimated function implies that cord glucose level has little effect on birth weight when the level of cord glucose is less than 75mg, then nonlinearly modifies genetic influences on birth weight as the cord glucose level increases. Such a nonlinear dynamic effect cannot be depicted by a linear interaction model, showing the relative merit of the nonlinear interaction analysis.

In summary, we identify various numbers of SNPs in the four populations that are either modulated by glucose or

insensitive to glucose supply to affect birth weight at a genome-wide suggestive threshold (10⁻⁵). It is interesting to note that none of the identified SNPs overlap in the four populations, indicating strong genetic heterogeneity in the four populations. The identified SNPs provide a catalog of variants associated with birth weight in different ethnic groups and serve as potential candidate SNPs for further biological validation.

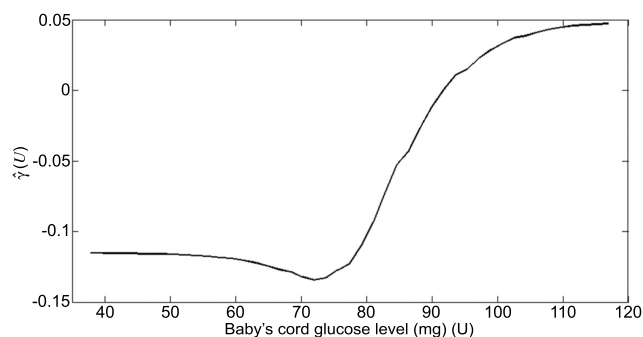


Fig. (3). The plot of the estimated varying-coefficient function $\hat{\gamma}(U)$ against baby's cord glucose level (U) for SNP rs1040193 located in gene *IL1A* in the Hispanic population.

CONCLUSION AND DISCUSSION

Birth weight is a complex trait that is controlled by genes as well as complex interactions between genes and environmental factors. However, the exact number of genes that control fetal growth during pregnancy are largely unknown, and the underlying G×E interaction machinery is poorly understood. Genome-wide association data provide rich resources in the identification of genetic variants associated with birth weight. Furthermore, the identification of G×E interactions offers promise for therapeutic intervention of birth related disease.

The purpose of this work is to reanalyze a GWAS data set which consists of four populations, to identify common genetic variants associated with birth weight, and to reveal the genetic heterogeneity of the four populations on birth weight. We analyzed the data using three different statistical models. At a genome-wide suggestive threshold (10^{-5}), we identified 33 SNPs in the Thai population, 26 SNPs in the Afro-Caribbean population, 18 SNPs in the European population, and 7 SNPs in the Hispanic population. Although none of these SNPs reached a genome-wide Bonferroni adjusted threshold, they provide substantial insights into the genetic heterogeneity of birth weight in different ethnic groups.

In addition to fit regular linear regression models to identify main genetic and linear G×E interaction effects, we applied a nonlinear VC model to identify nonlinear G×E interactions to understand how baby's genes are modulated by mother's glucose supply to affect birth weight. A few papers have reported that mother's glucose directly affects birth weight [23, 18]. Thus, it is our interest to investigate how mother's glucose interacts with baby's genes to affect birth weight. At the 10^{-5} genome-wide threshold, we identified 16 SNPs in the Thai population, 7 SNPs in the Afro-Caribbean population, 7 SNPs in the European population, and 2 SNPs in the Hispanic population that show nonlinear genetic sensitivity to glucose on birth weight. These SNPs cannot be detected by applying regular linear models and they provide novel insights into the genetic basis of birth weight.

To our surprise, the identified SNPs show no overlap across the four populations. The large difference in genetic variants identified in the four populations, on one hand, reveals the genetic heterogeneity of birth weight and provides insight into the genetic architecture of birth weight in differ-

ent ethnic groups. On the other hand, such heterogeneity might be due to the reproducibility issue of single SNP analysis. A few studies have reported the relative merit of gene-based analysis to improve association reproducibility (e.g., [24-26]). Thus, our future investigations will be focused on gene-level analysis which might have better reproducibility given that gene function may be more preservative across populations. In addition, we must admit that there are total 241,637 overlapped SNPs in the four populations after quality control. Some of the SNPs identified in one population might not exist in other populations, leading to the non-overlapping results observed in this analysis. Thus, we should also be cautious to reach the heterogeneity conclusion.

LIST OF ABBREVIATIONS

BMI	=	Body mass index
G×E	=	Gene-environment interaction
GEI	=	Genes, Environment, and Health Initiative
GENEVA	=	Gene Environment Association Studies initiative
GWAS	=	Genome-wide association study
HAPO	=	Hyperglycemia and Adverse Pregnancy Outcome
MAF	=	Minor allele frequency
OGTT	=	Oral glucose tolerance test
SNPs	=	Single nucleotide polymorphisms
VC	=	Varying-coefficient

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for their insightful comments that greatly improved the presentation of the manuscripts. This work was partially supported by grants from NSF (DMS-1209112) and from National Natural Science Foundation of China (31371336 and 81001294). Funding support for the GWA mapping: Maternal Metabolism-Birth Weight Interactions study was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01HG004415). The study is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01 HG04424). The datasets used for the analyses described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap> through dbGaP accession number phs000096.v4.p1.

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

- [1] Eriksson, J.G.; Forsen, T.; Tuomilehto, J.; Osmond, C.; Barker, D.J. Early growth and coronary heart disease in later life: longitudinal study. *B. M. J.*, **2001**, *322*, 949-953.
- [2] Wei, J.N.; Sung, F.C.; Li, C.Y.; Chang, C.H.; Lin, R.S.; Lin, C.C.; Chiang, C.C.; Chuang, L.M. Low birth weight and high birth weight infants are both at an increased risk to have type 2 diabetes among schoolchildren in taiwan. *Diabetes Care*, **2003**, *26*, 343-348.
- [3] Bray, G.A.; Bellanger, T. Epidemiology trends, and morbidities of obesity and the metabolic syndrome. *Endocrine*, **2006**, *29*, 109-117.
- [4] Lackland, D.T.; Bendall, H.E.; Osmond, C.; Egan, B.M.; Barker, D.J. Low birth weights contribute to high rates of early-onset chronic renal failure in the Southeastern United States. *Arch Intern Med.*, **2000**, *160*, 1472-1476.
- [5] Vikse, B.E.; Irgens, L.M.; Leivestad, T.; Hallan, S.; Iversen, B.M. Low birth weight increases risk for end-stage renal disease. *J. Am. Soc. Nephrol.*, **2008**, *19*, 151-157.
- [6] Rich-Edwards, J.W.; Stampfer, M.J.; Manson, J.E.; Rosner, B.; Hankinson, S.E.; Colditz, G.A.; Willett, W.C.; Hennekens, C.H. Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *B. M. J.*, **1997**, *315*, 396-400.
- [7] Clausson, B.; Lichtenstein, P.; Cnattingius, S. Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *B. J. O. G.*, **2000**, *107*, 375-381.
- [8] Little, R.E.; Sing, C.F. Genetic and environmental influences on human birth weight. *Am. J. Hum. Genet.*, **1987**, *40*, 512-526.
- [9] Freathy, R.M.; Mook-Kanamori, D.O.; Sovio, U.; Prokopenko, I.; Timpson, N.J.; Berry, D.J.; Warrington, N.M.; Widén, E.; Hottenga, J.J.; Kaakinen, M.; Lange, L.A.; Bradfield, J.P.; Kerkhof, M.; Marsh, J.A.; Magi, R.; Chen, C.M.; Lyon, H.N.; Kirin, M.; Adair, L.S.; Aulchenko, Y.S.; Bennett, A.J.; Borja, J.B.; Bouatia-Naji, N.; Charoen, P.; Coin, L.J.; Cousminer, D.L.; de Geus, E.J.; Deloukas, P.; Elliott, P.; Evans, D.M.; Froguel, P.; Genetic Investigation of ATC.; Glaser, B.; Groves, C.J.; Hartikainen, A.L.; Hasanani, N.; Hirschhorn, J.N.; Hofman, A.; Holly, J.M.; Hyppönen, E.; Kanoni, S.; Knight, B.A.; Laitinen, J.; Lindgren, C.M.; Meta-Analyses of G.; Insulin-related traits, C.; McArdle, W.L.; O'Reilly, P.F.; Pennell, C.E.; Postma, D.S.; Pouta, A.; Ramasamy, A.; Rayner, N.W.; Ring, S.M.; Rivadeneira, F.; Shields, B.M.; Strachan, D.P.; Surakka, I.; Taanila, A.; Tiesler, C.; Uitterlinden, A.G.; van Duijn, C.M.; Wellcome, Trust Case Control, C.; Wijga, A.H.; Willemssen, G.; Zhang, H.; Zhao, J.; Wilson, J.F.; Steegers, E.A.; Hattersley, A.T.; Eriksson, J.G.; Peltonen, L.; Mohlke, K.L.; Grant, S.F.; Hakonarson, H.; Koppelman, G.H.; Dedoussis, G.V.; Heinrich, J.; Gillman, M.W.; Palmer, L.J.; Frayling, T.M.; Boomsma, D.I.; Davey Smith, G.; Power, C.; Jaddoe, V.W.; Jarvelin, M.R.; Early Growth Genetics C; McCarthy, M.I. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nat. Genet.*, **2010**, *42*, 430-435.
- [10] Metrustry, S.J.; Edwards, M.H.; Medland, S.E.; Holloway, J.W.; Montgomery, G.W.; Martin, N.G.; Spector, T.D.; Cooper, C. Valdes A.M. Variants close to NTRK2 gene are associated with birth weight in female twins. *Twin Res. Hum. Genet.*, **2014**, *17*, 254-261.
- [11] Horikoshi, M.; Yaghothkar, H.; Mook-Kanamori, D.O.; Sovio, U.; Taal, H.R.; Hennig, B.J.; Bradfield, J.P.; St Pourcain, B.; Evans, D.M.; Charoen, P.; Kaakinen, M.; Cousminer, D.L.; Lehtimäki, T.; Kreiner-Møller, E.; Warrington, N.M.; Bustamante, M.; Feenstra, B.; Berry, D.J.; Thiering, E.; Pfab, T.; Barton, S.J.; Shields, B.M.; Kerkhof, M.; van Leeuwen, E.M.; Tulford, A.J.; Kutalik, Z.; Zhao, J.H.; den Hoed, M.; Mahajan, A.; Lindi, V.; Goh, L.K.; Hottenga, J.J.; Wu, Y.; Raitakari, O.T.; Harder, M.N.; Meirhaeghe, A.; Ntalla, I.; Salem, R.M.; Jameson, K.A.; Zhou, K.; Monies, D.M.; Lagou, V.; Kirin, M.; Heikkinen, J.; Adair, L.S.; Alkuraya, F.S.; Al-Odaib, A.; Amouyel, P.; Andersson, E.A.; Bennett, A.J.; Blakemore, A.I.; Buxton, J.L.; Dallongeville, J.; Das, S.; de Geus, E.J.; Estivill, X.; Flexeder, C.; Froguel, P.; Geller, F.; Godfrey, K.M.; Gottrand, F.; Groves, C.J.; Hansen, T.; Hirschhorn, J.N.; Hofman, A.; Hollegaard, M.V.; Hougaard, D.M.; Hyppönen, E.; Inskip, H.M.; Isaacs, A.; Jørgensen, T.; Kanaka-Gantenbein, C.; Kemp, J.P.; Kiess, W.; Kilpeläinen, T.O.; Klopp, N.; Knight, B.A.; Kuzawa, C.W.; McMahon, G.; Newnham, J.P.; Niinikoski, H.; Oostra, B.A.; Pedersen, L.; Postma, D.S.; Ring, S.M.; Rivadeneira, F.; Robertson, N.R.; Sebert, S.; Simell, O.; Slowinski, T.; Tiesler, C.M.; Tönjes, A.; Vaag, A.; Viikari, J.S.; Vink, J.M.; Vissing, N.H.; Wareham, N.J.; Willemssen, G.; Witte, D.R.; Zhang, H.; Zhao, J.; Meta-Analyses of Glucose- and Insulin-related traits Consortium (MAGIC), Wilson, J.F.; Stumvoll, M.; Prentice, A.M.; Meyer, B.F.; Pearson, E.R.; Boreham, C.A.; Cooper, C.; Gillman, M.W.; Dedoussis, G.V.; Moreno, L.A.; Pedersen, O.; Saarinen, M.; Mohlke, K.L.; Boomsma, D.I.; Saw, S.M.; Lakka, T.A.; Körner, A.; Loos, R.J.; Ong, K.K.; Vollenweider, P.; van Duijn, C.M.; Koppelman, G.H.; Hattersley, A.T.; Holloway, J.W.; Hoche, B.; Heinrich, J.; Power, C.; Melbye, M.; Guxens, M.; Pennell, C.E.; Bonnelykke, K.; Bisgaard, H.; Eriksson, J.G.; Widén, E.; Hakonarson, H.; Uitterlinden, A.G.; Pouta, A.; Lawlor, D.A.; Smith, G.D.; Frayling, T.M.; McCarthy, M.I.; Grant, S.F.; Jaddoe, V.W.; Jarvelin, M.R.; Timpson, N.J.; Prokopenko, I.; Freathy, R.M. Early Growth Genetics (EGG) Consortium. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat. Genet.*, **2013**, *45*(1), 76-82.
- [12] Falconer, D.S. The problem of environment and selection. *Am. Natural*, **1952**, *86*, 293-298.
- [13] Ross, C.A.; Smith, W.W. Gene-environment interactions in Parkinson's disease. *Parkinsonism Relat. Disord.*, **2007**, *13* Suppl 3, S309-315.
- [14] Zimmet, P.; Alberti, K.G.; Shaw, J. Global and societal implications of the diabetes epidemic. *Nature*, **2001**, *414*, 782-787.
- [15] Caspi, A.; Moffitt, T.E. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nat. Rev. Neurosci.*, **2006**, *7*, 583-590.
- [16] Ma, S.J.; Yang, L.J.; Romero, R.; Cui, Y.H. Varying coefficient model for gene-environment interaction: a nonlinear look. *Bioinformatics*, **2011**, *27*, 2119-2126.
- [17] Wu, C.; Cui, Y. A novel method for identifying nonlinear gene-environment interactions in case-control association studies. *Hum. Genet.*, **2013**, *132*, 1413-1425.
- [18] Ong, K.K.; Diderholm, B.; Salzano, G.; Wingate, D.; Hughes, I.A.; MacDougall, J.; Acerini, C.L.; Dunger, D.B. Pregnancy insulin, glucose, and BMI contribute to birth outcomes in nondiabetic mothers. *Diabetes Care*, **2008**, *31*, 2193-2197.
- [19] Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; Sham, P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **2007**, *81*, 559-575.
- [20] Kuschel, L.; Hansel, A.; Schonherr, R.; Weissbach, H.; Brot, N.; Hoshi, T.; Heinemann, S.H. Molecular cloning and functional expression of a human peptide methionine sulfoxide reductase (hMsra). *FEBS Lett.*, **1999**, *456*, 17-21.
- [21] Scherag, A.; Dina, C.; Hinney, A.; Vatn, V.; Scherag, S.; Vogel, C.I.; Muller, T.D.; Grallert, H.; Wichmann, H.E.; Balkau, B.; Heude, B.; Jarvelin, M.R.; Hartikainen, A.L.; Levy-Marchal, C.; Weill, J.; Delplanque, J.; Korner, A.; Kiess, W.; Kovacs, P.; Rayner, N.W.; Prokopenko, I.; McCarthy, M.I.; Schafer, H.; Jarick, I.; Boeing, H.; Fisher, E.; Reinehr, T.; Heinrich, J.; Rzehak, P.; Berdel, D.; Borte, M.; Biebermann, H.; Krude, H.; Rosskopf, D.; Rimbach, C.; Rief, W.; Fromme, T.; Klingenspor, M.; Schurmann, A.; Schulz, N.; Nothen, M.M.; Muhleisen, T.W.; Erbel, R.; Jockel, K.H.; Moebus, S.; Boes, T.; Illig, T.; Froguel, P.; Hebebrand, J.; Meyre, D. Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet.*, **2010**, *6*, e1000916.
- [22] Sata, F.; Toya, S.; Yamada, H.; Suzuki, K.; Saijo, Y.; Yamazaki, A.; Minakami, H.; Kishi, R. Proinflammatory cytokine polymorphisms and the risk of preterm birth and low birthweight in a Japanese population. *Mol. Hum. Reprod.*, **2009**, *15*, 121-130.
- [23] Franks, P.W.; Looker, H.; Kobes, S.; Touger, L.; Tataranni, P.A.; Hanson, R.L.; Knowler, W.C. Gestational glucose tolerance and risk of type 2 diabetes in young Pima Indian offspring. *Diabetes*, **2006**, *55*, 460-465.
- [24] Neale, B.M.; Sham, P.C. The future of association studies: gene-

- based analysis and replication. *Am. J. Hum. Genet.*, **2004**, *75*, 353-62.
- [25] Cui, Y.; Kang, G.; Sun, K.; Qian, M.; Romero, R.; Fu, W. Gene-centric genomewide association study via entropy. *Genetics*, **2008**, *179*, 637-650.
- [26] Liu, J.Z.; McRae, A.F.; Nyholt, D.R.; Medland, S.E.; Wray, N.R.; Brown, K.M.; Investigators, A.; Hayward, N.K.; Montgomery, G.W.; Visscher, P.M.; Martin, N.G.; Macgregor, S. A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.*, **2010**, *87*, 139-45.