

ARTICLE

Gene by stress genome-wide interaction analysis and path analysis identify *EBF1* as a cardiovascular and metabolic risk gene

Abanish Singh^{*,1,2,3}, Michael A Babyak^{1,3}, Daniel K Nolan², Beverly H Brummett^{1,3}, Rong Jiang^{1,3}, Ilene C Siegler^{1,3}, William E Kraus^{4,5}, Svati H Shah^{2,4}, Redford B Williams^{1,3} and Elizabeth R Hauser^{2,4,6}

We performed gene–environment interaction genome-wide association analysis (G × E GWAS) to identify SNPs whose effects on metabolic traits are modified by chronic psychosocial stress in the Multi-Ethnic Study of Atherosclerosis (MESA). In Whites, the G × E GWAS for hip circumference identified five SNPs within the *Early B-cell Factor 1 (EBF1)* gene, all of which were in strong linkage disequilibrium. The gene-by-stress interaction (SNP × STRESS) term *P*-values were genome-wide significant (*P*s = 7.14E–09 to 2.33E–08, uncorrected; *P*s = 1.99E–07 to 5.18E–07, corrected for genomic control). The SNP-only (without interaction) model *P*-values (*P*s = 0.011–0.022) were not significant at the conventional genome-wide significance level. Further analysis of related phenotypes identified gene-by-stress interaction effects for waist circumference, body mass index (BMI), fasting glucose, type II diabetes status, and common carotid intimal–medial thickness (CCIMT), supporting a proposed model of gene-by-stress interaction that connects cardiovascular disease (CVD) risk factor endophenotypes such as central obesity and increased blood glucose or diabetes to CVD itself. Structural equation path analysis suggested that the path from chronic psychosocial stress to CCIMT via hip circumference and fasting glucose was larger (estimate = 0.26, *P* = 0.033, 95% CI = 0.02–0.49) in the *EBF1* rs4704963 CT/CC genotypes group than the same path in the TT group (estimate = 0.004, *P* = 0.34, 95% CI = –0.004–0.012). We replicated the association of the *EBF1* SNPs and hip circumference in the Framingham Offspring Cohort (gene-by-stress term *P*-values = 0.007–0.012) as well as identified similar path relationships. This observed and replicated interaction between psychosocial stress and variation in the *EBF1* gene may provide a biological hypothesis for the complex relationship between psychosocial stress, central obesity, diabetes, and cardiovascular disease.

European Journal of Human Genetics (2015) 23, 854–862; doi:10.1038/ejhg.2014.189; published online 1 October 2014

INTRODUCTION

Although mortality attributable to cardiovascular disease (CVD) has declined in the United States, the burden of disease remains high.¹ It remains the leading cause of illness and death worldwide. Hypertension, obesity, dyslipidemia, insulin resistance and type II diabetes mellitus, and physical inactivity are among the eight risk factors that account for 61% of cardiovascular deaths. These same risk factors account for over three quarters of ischemic heart disease.² These risk factors are influenced by both environmental exposures and genetic background and the heritability of these risk factors can be as high as 77%,³ making it difficult to clearly separate CVD risk factors into genetic and nongenetic categories. The INTERHEART study has evaluated the effect of both physical and psychosocial factors on the risk of myocardial infarction and has shown that a higher prevalence of psychological stress and other psychosocial factors like depression account for 34% of the population attributable risk for myocardial infarction, independently of physical risk factors.⁴ The co-occurrence of psychosocial risk factors suggests that these factors do not work in isolation, but tend to cluster in the same individuals and risk groups.⁵ For example, female individuals who had high stress at work also

reported high levels of hostility, anger, depression, anxiety, and social isolation.⁶ Although the psychosocial risk factors by themselves may not have a direct effect on the development of CVD, they can contribute to the disease via intermediate mechanisms and biological pathways.⁵ Previous research has shown that psychosocial factors are associated with other factors that can influence these biological pathways.^{7,8} The repeated exposure to psychosocial stress has been linked to the development of visceral obesity.⁹ Methylation studies have also demonstrated that distressing life events can alter the expression of variants through effects on promoter DNA methylation.¹⁰ These findings support the need to understand the relationship between genetic susceptibility, psychosocial stress, and metabolic factors acting in combination to increase the risk of CVD. We hypothesize that using a convergence of gene–environment interaction genome-wide association study (G × E GWAS) and exploration of phenotypic correlation structure will lead us to the identification of novel cardiovascular disease genes and variants. Although the principle of gene–environment interactions in human disease is well established,^{11–13} examples of G × E GWASs with

¹Behavioral Medicine Research Center, Duke University Medical Center, Durham, NC, USA; ²Center for Human Genetics, Duke University Medical Center, Durham, NC, USA; ³Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA; ⁴Department of Medicine, Duke University Medical Center, Durham, NC, USA; ⁵Duke Center for Living, Duke University Medical Center, Durham, NC, USA; ⁶Durham Epidemiologic Research and Information Center, Durham Veterans, Affairs Medical Center, Durham, NC, USA

*Correspondence: Dr A Singh, Center for Human Genetics, Duke University Medical Center, Duke Box 3445, 905 LaSalle Street, Durham, NC 27710, USA. Tel: +1 919 681 2133; Fax: +1 919 684 0934; E-mail: abanish.singh@duke.edu

Received 20 March 2013; revised 11 July 2014; accepted 19 August 2014; published online 1 October 2014

replication of the most significant variants in cardiovascular disease are still rare.

MATERIALS AND METHODS

Discovery data set and analysis

Study population. We chose the publicly available Multi-Ethnic Study of Atherosclerosis (MESA) Cohort¹⁴ data set for testing the G×E interaction hypothesis. The data set can be obtained from the NIH dbGaP data depository through an authorized access (URL https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000209.v12.p3). The selection of this data set was based on the availability of a well-characterized psychosocial stress chronic burden (ie, chronic psychosocial stress) variable, quantified on an ordinal scale of 0 to 5 (0, 1, 2, 3, 4, and 5) based on questionnaires in five domains including questions about ongoing serious health problems, serious health problems with someone close, work-related problems, financial strains, and difficulties in relationships. A total of 5805 individuals – 2460 Whites, 548 Chinese Americans, 1547 Blacks, and 1250 Hispanics – had quality-controlled genotype and phenotype data available. Additional methodological details are shown in the Supplementary Material.

Phenotypes. We used hip circumference (HIPCM), body mass index (BMI), waist circumference (WAISTCM), glucose (GLUC), triglycerides (TRIG), low-density lipoprotein (LDL), high-density lipoprotein (HDL), systolic blood pressure (SYSTBP), diastolic blood pressure (DIABTBP), type II diabetes status (DIABSTAT), common carotid intimal–media thickness (CCIMT), a surrogate marker for atherosclerosis that is a strong predictor of future vascular events,¹⁵ and chronic psychosocial stress (STRESS). Table 1 provides the summary of these traits in the MESA exam 1 data. For all the selected traits we checked for departures from normality using STATA SE 11.1 (StataCorp LP., College Station, TX, USA) and performed a transformation to achieve approximate normality within each ancestry (Supplementary Table S7).

Genotypes. The Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix Inc., Santa Clara, CA, USA) MESA SNPs were subjected to standard quality control procedures as implemented in PLINK¹⁶ v1.07 – that is, minimum minor allele frequency (MAF) of 1%, maximum missing genotypes of 5% per individual, maximum missing individuals per SNP of 5%, and significant departure from Hardy–Weinberg equilibrium (HWE; $P < 0.00001$) – resulting in 528 298 SNPs for the G×E GWAS. We estimated principal components (PCs) based on 162 072 LD-pruned SNPs in each ethnic group using

EIGENSTRAT¹⁷ and used five PCs in Whites, five PCs in Chinese Americans, three PCs in Blacks, and three PCs in Hispanics based on eigenvalue scree plots.

Phenotype selection and G×E GWAS. We selected the metabolic phenotype that showed the strongest correlation with the chronic psychosocial stress variable and performed ancestry-stratified G×E GWAS for the selected pair using a linear regression model (logistic regression for binary phenotypes) under the default additive genetic model with population ancestry correction, age, and sex adjustment as implemented in PLINK,¹⁶ v1.07. The ordinal stress variable was assumed to have a linear relationship with the other variables. We selected significant SNPs applying a conservative Bonferroni correction ($P = 9.46E - 08$) on the gene-by-stress interaction term to identify the SNPs whose effect on metabolic traits may be modified most by stress, and explored G×E association for these SNPs in related metabolic phenotypes in the same data set using a postulated pathway model leading to increased risk of CVD.⁵ The distribution of *P*-values for the interaction term was corrected using genomic control.¹⁸ We also performed association analysis using a conventional SNP-only additive model (ie, without STRESS and SNP×STRESS terms) for the SNPs identified in the G×E GWAS.

Structural equation path analysis. We compared genotypes on the magnitude of zero-order correlations among the untransformed variables under study and used a structural equation path analysis as implemented in *Mplus*,¹⁹ version 6.11 to evaluate possible mediated causal paths²⁰ from stress through risk factors to CCIMT. Path analysis uses a series of simultaneous equations to estimate possible mediating paths. The magnitude of a mediating effect is then calculated by taking the product of all path coefficients along a given proposed path. Further technical details of the structural equation modeling are shown in the Supplementary Material.

Replication data set and analysis

We used the publicly available Framingham Cohort²¹ Generation 2 (or Offspring) data set for our replication analysis. The data set can be obtained from the NIH dbGaP data depository through an authorized access (URL https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v23.p8). The cohort is primarily white. A total of 3157 individuals had both phenotype and quality-controlled genotype data available. We selected replication phenotypes hip girth (HIPGIRTH), waist girth (WAISTGIRTH), BMI, GLUC, DIABSTAT, and CCIMT (Supplementary Table S1). We used a computed ordinal measure of chronic psychosocial stress (STRESS, ranged 0–3) as the sum of dichotomized indicators of financial strain (total family income;

Table 1 Summary of phenotypic traits in the four ethnic groups in MESA data set for the samples having QC genotypes: AGE, SEX, hip circumference (HIPCM), waist circumference (WAISTCM), body mass index (BMI), triglycerides (TRIG), glucose (GLUC), low-density lipoprotein (LDL), high-density lipoprotein (HDL), systolic blood pressure (SYSTBP), diastolic blood pressure (DIABTBP), type II diabetes status (DIABSTAT), common carotid intimal–media thickness (CCIMT), and chronic psychosocial stress measure (STRESS)

Ancestry → Trait	WHITES (N = 2460)		CHINESE AMERICAN (N = 548)		BLACKS (N = 1547)		HISPANICS (N = 1250)	
	Mean (SD)	Min–max	Mean (SD)	Min–max	Mean (SD)	Min–max	Mean (SD)	Min–max
AGE (Years)	62.65 (10.21)	44–84	62.53 (10.13)	44–84	62.29 (10.07)	45–84	61.38 (10.31)	44–84
SEX (M/F)	47.7%/52.3%	Binary (1/2)	65.5%/34.5%	Binary (1/2)	46.6%/53.4%	Binary (1/2)	50.4%/49.6%	Binary (1/2)
HIPCM (CM)	106.08 (10.47)	40.5–160.2	94.84 (6.29)	77–118.7	109.78 (12.20)	83–159	105.07 (10.63)	76.9–160.6
WAISTCM (CM)	97.94 (14.47)	33.5–193.5	87.30 (9.44)	62–116.1	101.45 (14.75)	63–167	100.53 (13.06)	61.8–159.9
BMI (kg/m ²)	27.73 (5.05)	16.9–48.9	24.07 (3.28)	15.36–35.35	30.16 (5.91)	15.87–61.86	29.33 (5.13)	17.57–51.66
TRIG (mg/dl)	133.21 (90.33)	23–2256	142.72 (85.74)	30–831	104.85 (70.87)	21–1540	156.93 (96.38)	33–1670
GLUC (mg/dl)	91.30 (21.77)	38–447	99.75 (29.87)	71–341	100.50 (33.09)	39–377	103.93 (38.86)	38–507
LDL (mg/dl)	117.09 (30.30)	28–252	115.09 (28.79)	41–244	116.32 (32.97)	12–284	119.79 (33.20)	26–315
HDL (mg/dl)	52.39 (15.78)	21–142	48.21 (12.20)	23–138	52.32 (15.16)	15–127	47.19 (12.82)	21–133
SYSTBP (mm Hg)	123.48 (20.49)	75–218	124.46 (20.54)	77.5–204	131.59 (21.80)	67–230.5	126.85 (22.11)	80–211.5
DIABTBP (mm Hg)	70.16 (9.94)	37–108	72.92 (10.13)	43.5–109	74.52 (10.31)	46.5–118.5	71.82 (10.14)	43–108.5
DIABSTAT (N/D) ^a	93.3%/6.7%	Binary (0/1)	83.2%/16.8%	Binary (0/1)	79.3%/20.7%	Binary (0/1)	78.4%/21.6%	Binary (0/1)
CCIMT (mm)	0.87 (0.20)	0.46–2.45	0.82 (0.17)	0.46–2.08	0.90 (0.19)	0.4–2.38	0.86 (0.19)	0.43–2.33
STRESS (Score)	1.23 (1.18)	0–5	0.73 (1.03)	0–5	1.35 (1.24)	0–5	1.24 (1.20)	0–5

^aNormal/diabetes (treated and nontreated).

low income indicated high financial strain), marital problems (marital disagreements scales), work-related difficulties (job insecurity and psychological job demands scales), and health problems of someone close (if spouse ever had heart attack, stroke, or heart disease) from the available psychosocial data. The Framingham Cohort did not include an indicator of self-rated health, and therefore this indicator was not included in the psychosocial stress measure. We checked for departure from normality using STATA SE 11.1 and performed a transformation to achieve approximate normality in each study phenotype (Supplementary Table S8). We performed replication for three of the five most significant MESA *Early B-Cell Factor 1* (*EBF1*) SNPs (rs17056278:C>G, rs17056298:C>G, and rs17056318:T>C) directly genotyped on the Framingham genotyping array, were in high LD ($R^2=0.84-1.0$) with the lead MESA SNP (rs4704963:T>C) and met our quality control filters described above. We performed linear regression (logistic regression for binary phenotypes) under the additive model, including a SNP \times STRESS term for G \times E interaction, and family ID using the Huber-White adjustment for familial clustering as implemented in STATA SE 11.1. We used Fisher's exact test to obtain the discovery and replication combined *P*-values. Structural equation pathway analysis was performed as described above.

RESULTS

Phenotype selection for discovery GWAS

We calculated correlation coefficients between the chronic psychosocial stress measure and metabolic and body weight-related traits for all four ethnic groups in the MESA data set – Whites, Chinese Americans, Blacks, and Hispanics. The strongest correlation between chronic stress and the metabolic traits in MESA was with hip circumference in three of the racial subsets (Table 2), and thus hip circumference was used as the primary metabolic 'outcome' of interest for the gene by stress G \times E GWAS.

Discovery GWAS and CVD-related phenotypes association

Table 3 shows the five SNPs (rs4704963:T>C, rs17056278:C>G, rs17056298:C>G, rs10077799:T>C, and rs17056318:T>C) resulting from the MESA Whites hip circumference G \times E GWAS for which gene-by-stress interactions were genome-wide significant (*P*-values = 7.14E-09 to 2.33E-08 uncorrected, 1.99E-07 to 5.18E-07 GC corrected, Figure 1a, Table 3, and Supplementary Figures S1 and S2). All five SNPs were in strong LD ($R^2=0.84-1.0$) and are located in intronic regions of *EBF1* gene on chromosome 5; that is, rs4704963: T>C between exons 8 and 9, rs17056278:C>G between exons 7 and 8, and the other three between exons 6 and 7 (numbering with 1 for the first exon) of the longest protein coding transcript NM_001290360.1 (RefSeq) that has 16 exons. The SNP-only model (ie, without interaction) *P*-values for the SNPs (*P*s = 0.011–0.022) were not significant at the conventional genome-wide significance level (Table 3). No other SNPs achieved genome-wide significance. The SNPs with the next 5 lowest *P*-values did not achieve genome-wide significance but contained three genes with potential biological significance: *GPRC6A*, *ATP6V1D*, and *SASH1* (Supplementary Table S2). The remaining two SNPs are intergenic with no obvious functional significance. When G \times E GWAS was done in other ancestries, there were no genome-wide significant results for Chinese Americans, Blacks, and Hispanics for the gene-by-stress interaction term (Supplementary Figures S3–S5 and Supplementary Tables S3–S5). Although there were several SNPs with *P*-values < 10E-05, there was no replication of any SNP (or gene) across the four ethnic groups.

Given the strong statistical evidence for the G \times E GWAS associations for the SNPs in *EBF1* (Figure 1a, Table 3, and Supplementary Figure S6), we chose rs4704693 as the lead SNP to explore other CVD-related phenotypes. We observed a similar pattern

Table 2 Pair-wise correlation coefficients of chronic psychosocial stress with central obesity and metabolic traits in the MESA data sets for the samples having quality-controlled genotype and phenotype data

ANCESTRY	CENTRAL OBESITY TRAITS			METABOLIC TRAITS					
	HIPCM	WASTCM	BMI	GLUC	LDL	HDL	TRIG	SYSTBP	DIASTBP
WHITES	0.1053* (<0.0001)	0.0457* (0.023)	0.0990* (<0.0001)	0.0096 (0.63)	-0.0101 (0.62)	0.0356 (0.079)	0.0561* (0.0058)	-0.0682* (0.0008)	-0.0442* (0.029)
CHINESE AMERICANS	-0.0545 (0.21)	-0.0488 (0.27)	-0.0392 (0.36)	0.0188 (0.66)	-0.0412 (0.34)	0.0656 (0.13)	0.0249 (0.56)	-0.0439 (0.31)	-0.0530 (0.022)
BLACKS	0.1377* (<0.0001)	0.0888* (0.0005)	0.1384* (<0.0001)	0.0754* (0.0032)	0.0088 (0.73)	0.0271 (0.29)	0.0343 (0.18)	0.0066 (0.79)	-0.0073 (0.77)
HISPANICS	0.1084* (0.0001)	0.0493 (0.085)	0.0904* (0.0016)	0.0391 (0.17)	-0.0351 (0.22)	0.0618* (0.031)	-0.0105 (0.71)	-0.0826* (0.0019)	-0.0358 (0.21)

The significance levels of correlations are shown in parentheses. The bold entries represent the strongest correlations that were the basis of phenotype selection. *It denotes whether a pair-wise correlation was significant at 5% level.

Table 3 Genome-wide significant SNPs in the MESA Whites discovery set, Framingham Offspring replication set and combined *P*-values

SNP	CHR	POS AND ALLELE CHANGE	Type	Gene	SNP <i>P</i> -val	Discovery			Replication			Combined <i>P</i> -val
						MESA Whites <i>P</i> -val	GC <i>P</i> -val	MAF	Framingham Offspring <i>P</i> -val	MAF	<i>P</i> -val	
rs4704963	5	g.158247378T>C	Intronic	EBF1	0.022	7.14E-09	1.99E-07	0.07	NA	NA	NA	NA
rs17056278	5	g.158252438C>G	Intronic	EBF1	0.022	7.14E-09	1.99E-07	0.07	0.308	0.007	0.07	2.98E-08
rs17056298	5	g.158268679C>G	Intronic	EBF1	0.018	1.30E-08	3.24E-07	0.06	0.230	0.012	0.07	7.92E-08
rs10077799	5	g.158272083T>C	Intronic	EBF1	0.016	1.71E-08	4.05E-07	0.06	NA	NA	NA	NA
rs17056318	5	g.158285953T>C	Intronic	EBF1	0.011	2.33E-08	5.18E-07	0.07	0.152	0.008	0.07	8.41E-08

Abbreviation: NA, not available/not applicable.

(*Discovery*) The table includes *P*-values for the SNP-only model (SNP *P*-val), that is, the additive genetic model with no interaction term, G×E *P*-values (*P*-val), genomic control G×E *P*-values (GC *P*-val), and MAF of the five SNPs with genome-wide significant *P*-values from the MESA Whites G×E GWAS on Hip Circumference (HIPCM). The environment (E) component in this GWAS was chronic psychosocial stress (STRESS). (*Replication*) The table includes *P*-values of SNP-only Model (SNP *P*-val), G×E *P*-values (*P*-val), and MAF of three out of the five SNPs available in Framingham Offspring Cohort. (*Combined*) The combined *P*-values were obtained using Fisher's exact test on MESA G×E GC-controlled *P*-values (GC *P*-val) and Framingham G×E *P*-values (*P*-val). All genomic locations were numbered based on Genome Reference Consortium Human Build 37 patch release 10 (GRCh37.p10).

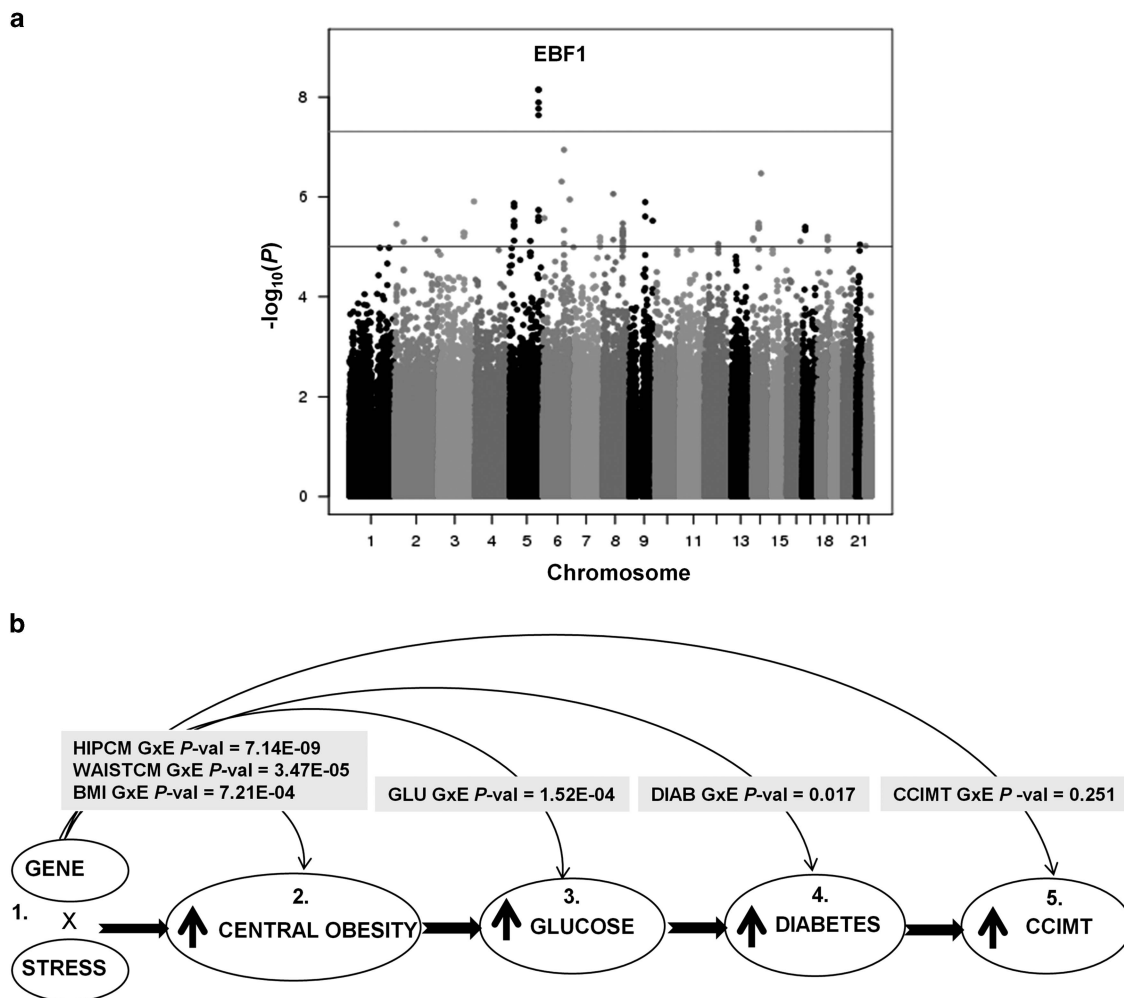


Figure 1 (a) Manhattan plot of G×E GWAS on hip circumference (HIPCM) in MESA Whites where the environmental factor (E) was chronic psychosocial stress (STRESS). (b) G×E model: a schematic diagram showing the effect of *EBF1* SNP rs4704963:T>C on central obesity, glucose, diabetes status, and common carotid IMT, and modified by chronic psychosocial stress in MESA Whites. 'Pheno G×E *P*-val' denotes the *P*-values for the gene-by-stress interaction term for a given phenotype.

(Supplementary Tables S6 and S7) with evidence for G×E association with waist circumference, BMI, fasting glucose, and diabetes status; modest evidence of association with CCIMT, diastolic and systolic blood pressures, and HDL; and no evidence of

association with triglycerides and LDL. As hip circumference, BMI, and waist circumference were strongly correlated, they show the same relative order of the significance of association with the top *EBF1* SNPs.

We focused our further analysis on the central obesity traits (hip circumference, waist circumference, and BMI), fasting glucose, and diabetes status along with CCIMT that we used as a proxy measure of atherosclerosis,¹⁵ given the low prevalence of clinically significant cardiovascular disease in the MESA data set. We observed that the effect of *EBF1* SNP rs4704963:T>C on central obesity, glucose, diabetes status, and CCIMT was modified by chronic stress in Whites, and also that the magnitude of evidence for modification became smaller from central obesity to glucose to diabetes status to CCIMT (Supplementary Table S7). However, as might be expected given the correlations among the phenotypes, the pattern of modification in the effect of SNP on the phenotypes remained consistent. In order to frame these association results and the magnitude of interaction in multiple CVD-related phenotypes as a system, we hypothesized a pathway model starting with the gene-by-stress interaction through central obesity, high glucose, diabetes, and CCIMT (Figure 1b).⁵ In this model, chronic psychosocial stress modifies the effect of gene variants on CVD risk phenotypes; however, the *P*-value of modification was attenuated as we moved from left to right in the pathway. Figure 2 shows the direction of effect of major and minor alleles of rs4704963:T>C on central obesity, glucose, diabetes status, and CCIMT phenotypes that we included in the model (Figure 1b). We found that mean hip circumference was similar (mean = 106.61 cm, SD = 1.16) across various stress levels for those homozygous for the major allele (TT), but increased linearly (from 104 to 124 cm, mean = 112.71, SD = 8.26) with increasing chronic stress for the

carriers of minor allele (CT/CC, Figure 2). Similarly, in the high-stress group, individuals carrying the minor allele (CT/CC) had high risk factor levels compared with individuals homozygous for the major allele (TT, Figure 2). We also confirmed this direction of effect in all six traits (from Figure 1b) by the sign of the regression coefficient (β) for gene-by-stress interaction term (Supplementary Table S7) that represents the effect of each additional copy of the minor allele; that is, a positive (negative in case of inverse transformation) regression coefficient means that the minor allele increases risk/phenotype.

Replication of discovery GWAS

We identified three SNPs (rs17056278:C>G, rs17056298:C>G, and rs17056318:T>C) genotyped in the Framingham data set out of the five genome-wide significant SNPs in *EBF1* from the MESA GWAS. The minor allele frequencies of these SNPs in Framingham were similar to the MESA data set (Table 3). The SNP-only model and gene-by-stress interaction term *P*-values for hip girth are shown in Table 3, and the *P*-values for additional replication phenotypes, waist girth, BMI, glucose, diabetes status, and CCIMT, equivalent to MESA phenotypes included in Figure 1b, are shown in Supplementary Table S8. The gene-by-stress interaction term *P*-values for the central obesity traits were significant at the threshold 0.05 (*P*s = 0.001–0.012). The direction of effect, as shown in Figure 3, was similar to the discovery results in the MESA data set (Figure 2), although the magnitude of the effect was smaller. In both data sets, obesity-related

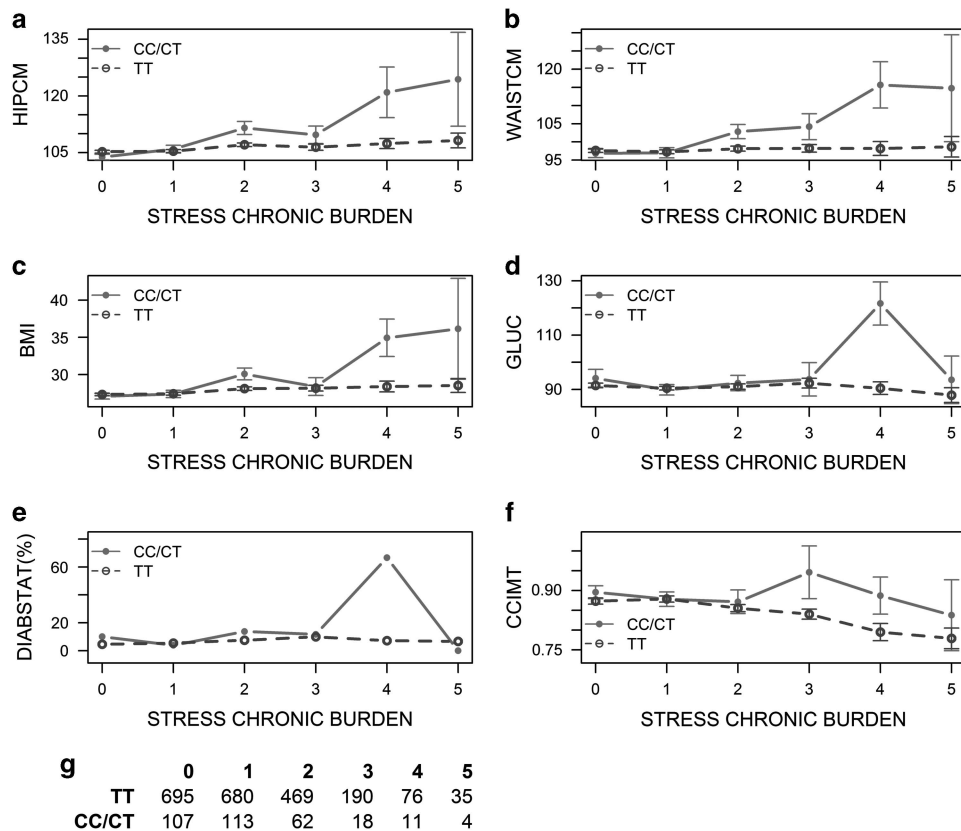


Figure 2 The direction of $G \times E$ interaction effect in MESA Whites. These plots show the mean and SE bars of the raw values of the phenotypes from the Figure 1b pathway model (hip circumference (a), waist circumference (b), BMI (c), glucose (d), diabetes status (e), common carotid intimal–media thickness (f)) vs chronic stress for the two genotype groups of the most significant SNP rs4704963:T>C for the major allele homozygotes (TT) and the minor allele carriers (CC/CT). The plot for diabetes status (DIABSTAT) shows the percent of individuals having type II diabetes mellitus. Panel (g) shows the number of individuals for each genotype-by-stress level.

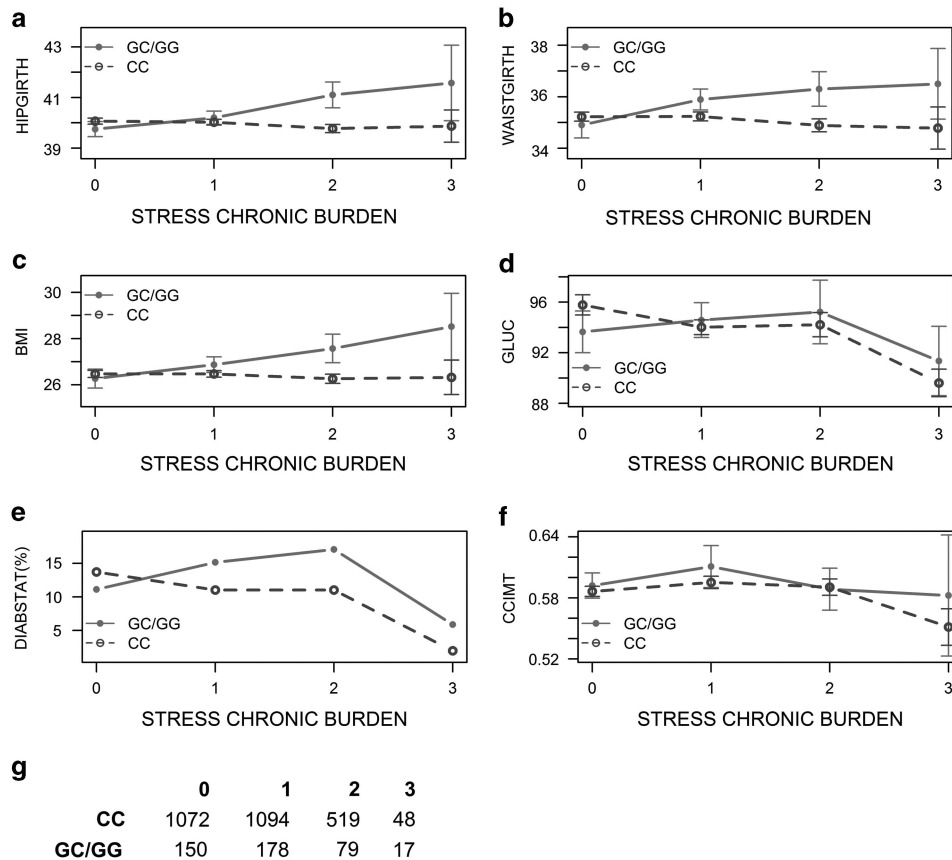


Figure 3 The direction of $G \times E$ interaction effect in the Framingham Offspring Cohort. These plots show the mean and SE bars of the raw values of the phenotypes from the Figure 1b pathway model (hip girth (a), waist girth (b), BMI (c), common carotid intimal-media thickness (f)) vs chronic stress for the two genotype groups for major allele homozygotes (GC/GG) and the minor allele carriers (CC) for EBF1 SNP rs17056278:C>G. The plot for diabetes status (DIABSTAT) shows the percent of individuals having type II diabetes mellitus. Panel (g) shows the number of individuals for each genotype-by-stress level.

risk factors increased with the psychosocial chronic stress for the minor allele groups.

Structural equation path analysis

We examined the pathway model shown in Figure 1b more closely by comparing the raw (zero-order) correlation between each pair of study variables included in the model in the two genotype groups – homozygote major allele (TT) and minor allele heterozygotes and homozygotes (CT/CC) – of the SNP rs4704963:T>C in MESA Whites. We observed significant differences in the zero-order correlations between the genotype groups in stress-central obesity traits (Supplementary Table S9). The P -value for the test of differences in correlations between chronic stress and hip circumference in the TT and CT/CC genotype groups was <0.0001 , with their correlations in the two genotype groups of 0.06 and 0.27 respectively. Figure 4 displays the results of the path analysis for the continuous phenotypes only, with unstandardized path coefficients. Path coefficients can be interpreted in a manner analogous to regression slope coefficients, where the coefficient represents the expected change in the dependent variable (the variable to which an arrow is directed) for each one unit change in the predictor (the variable that is pointing an arrow to the dependent variable). In MESA Whites (Figure 4a), consistent with the observed correlations (Supplementary Table S9), the direct path from chronic stress to hip circumference was statistically significant in both groups, but substantially larger in the CT/CC group compared with the same path in the TT group. In the TT group, a one point increase

on the stress scale was associated with an expected increase of 0.42 cm in hip circumference. In contrast, a one point increase in stress in the CT/CC group was associated with a 3.57 cm increase in hip circumference. In the CT/CC group, the path from chronic stress to CCIMT via hip circumference and glucose (STRESS \Rightarrow HIPCM \Rightarrow GLUCOSE \Rightarrow CCIMT) was larger (estimate = 0.26, $P = 0.033$, 95% CI = 0.02–0.49) than the same path that was not significant in the TT group (estimate = 0.004, $P = 0.340$, 95% CI = -0.004–0.012). The paths from chronic stress to glucose and from glucose to CCIMT were also considerably larger in the CT/CC group compared with the TT group.

Replication of structural equation path analysis

Testing the same path model in the Framingham Offspring Cohort data set, we observed a similar genotype difference in the path from the chronic stress score to hip circumference (Figure 4b), relatively small and nonsignificant in the major allele (C/C) group (estimate = -0.23, $P = 0.382$, 95% CI = -0.75–0.29), but larger and significant in the minor allele carrier group (G/C, G/G) (estimate = 1.65, $P = 0.007$, 95% CI = 0.45–2.84). However, unlike the MESA result, the indirect path from stress to CCIMT via hip circumference and glucose was extremely small and not statistically significant in either group. The estimate of this mediated effect was virtually zero in each group: 0.00008 ($P = 0.395$, 95% CI = -0.0003–0.0001) in the CC group and 0.0003 ($P = 0.346$, 95% CI = -0.0003–0.0008) in the GC/CC group.

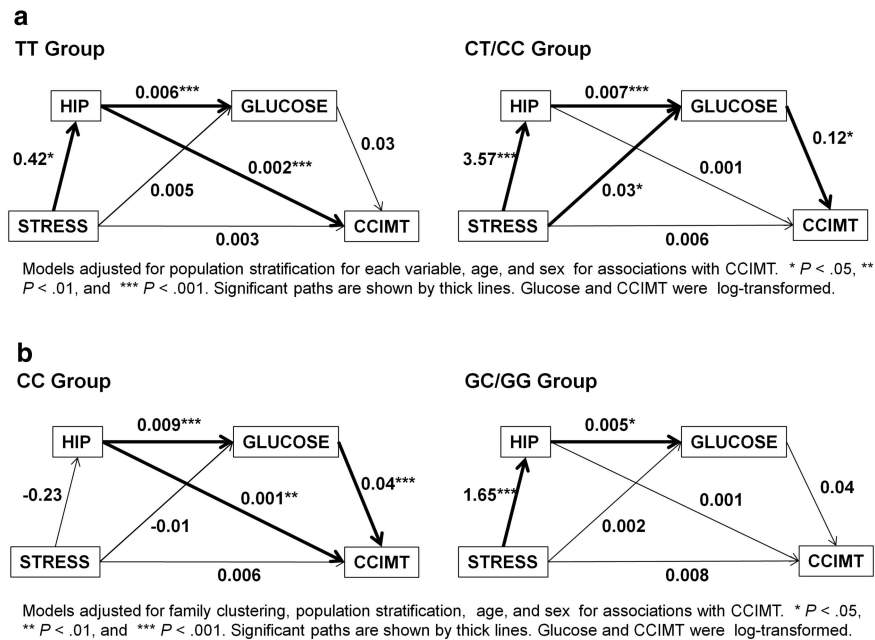


Figure 4 Structural equation path models of proposed direct and indirect effects among chronic psychosocial stress, hip circumference, fasting glucose, and common carotid intimal–medial thickness (CCIMT) for the two genotype groups (TT and CT/CC) of *EBF1* SNP rs4704963:T>C in MESA Whites (a), and for the two genotype groups (CC and GC/GG) *EBF1* SNP rs17056278:C>G in Framingham Offspring Cohort (b). Values represent unstandardized path (slope) coefficients. Data were adjusted for ancestry stratification in a preliminary step. Not depicted in the figure are age and sex, which were included as covariates, adjusting all paths leading to hip circumference, glucose, and CCIMT. Glucose and CCIMT were transformed using the natural logarithm.

Functional annotation of *EBF1* SNPs

Given that all genome-wide significant *EBF1* SNPs were intronic and in strong LD, we examined these SNPs further for any putative functional role. We observed that one of the two most significant SNPs rs17056278:C>G was located on a highly conserved site in the genome across multiple vertebrate species. This site includes the motif sequence *CCAT* that has been identified as a high-affinity binding motif at other places in the genome. Thus, this site is a candidate for conserved transcription factor binding site (tfbscons). Additional evidence indicated similar annotation (ie, tfbscons) for the SNP (Supplementary Figure S6).

DISCUSSION

We performed a genome-wide study to discover gene-by-stress interactions associated with cardiometabolic risk factors in the MESA data set. We observed five SNPs with significant interaction terms (uncorrected $P < 2.33E-08$ and GC corrected $P < 5.18E-07$) in the White ethnic group. The GC corrected P -values were not genome-wide significant. These SNPs are all in the gene *EBF1*, which is a transcription factor with a known hematopoietic function. We have demonstrated independent replication of this association ($P < 0.012$, combined $P < 8.41E-08$) in the Framingham Cohort data set for the three SNPs present in the Framingham genotype data set.

The role of *EBF1* in the development of the immune system has been well described.²² It has been implicated in other studies of cardiovascular disease.²³ The most significant *EBF1* SNP from our analysis (rs4704963:T>C) was also associated with early-onset coronary artery disease (CAD) in a genetic linkage and family-based association study^{23–25} with the same direction of effect, that is, the minor allele carriers have higher BMI and CAD severity.²³ This result increases confidence in validity of the finding in MESA that *EBF1* plays

a significant role in cardiovascular disease risk. Two large GWASs of European Whites identified genome-wide significant variants in *EBF1* associated with hypertension, systolic blood pressure, diastolic blood pressure, and mean arterial pressure.^{26,27} The *EBF1* gene was also identified as putative key regulatory gene for the coronary heart disease causal differential modules based on the global tissue-specific Bayesian and protein–protein interaction networks,²⁸ confirming the complex relationship between CVD and *EBF1*. However, the large consortia of GWASs – GIANT, MAGIC, DIAGRAM, and CARDIoGRAM – did not identify SNPs in *EBF1*. The genetic main effect P -values of the five *EBF1* SNPs for similar traits (ie, BMI, $P > 0.38$; fasting glycemic traits, $P > 0.08$; insulin resistance, $P > 0.38$; type II diabetes, $P > 0.63$; and CAD, $P > 0.54$) were not significant.^{29–32}

More directly documenting the moderation by *EBF1* rs4704963:T>C of stress effects on intermediate phenotypes that stimulate atherogenesis in the present study, path analysis revealed a stronger effect of chronic stress on hip circumference in C allele carriers compared with TT homozygotes. The indirect path from increased stress to hip circumference to glucose to CCIMT was significant in C carriers but not TT homozygotes. The replication of parts of the path model – that is, stress to hip circumference – in the Framingham Offspring data set further supports these relationships. However, the difference in the indirect path from stress to CCIMT in the MESA Whites and Framingham Offspring results could be because of age differences between the two cohorts (Table 1 and Supplementary Table S1). The Framingham Offspring cohort was younger (mean age = 47.76 years) compared with the MESA Whites (mean age = 62.65 years). In both cohorts, the mean CCIMT was greater with increasing age (Supplementary Figure S7) but there was a systematic difference – that is, Framingham had lower mean CCIMT compared with MESA in all age groups – that could result from cohort

differences or differences in protocols or equipment used for CCIMT measurements. The younger age and relatively lower CCIMT in the Framingham study would result in lower power to detect relationships with CCIMT.

Studies in model systems support the association of *EBF1* and the cardiometabolic risk factors observed in human studies. The *EBF1* knockout mouse has a striking metabolic phenotype characterized by lipodystrophy, hypotryglyceridemia, and hypoglycemia, while having an increased metabolic rate and decreased leptin levels.³³ *EBF1* has a critical role in the adipogenic transcriptional cascade in multiple cellular models. It binds to and activates the *C/EBPα* promoter that exerts positive feedback on *C/EBPα* expression, and in turn induces the expression of *PPARγ* that joins with *EBF1* and *C/EBPα* to induce adipocyte gene expression.³⁴ The mouse lipodystrophy is characterized by an increase in yellow adipose tissue in bone marrow and a marked decrease in white adipose tissue, relative to wild-type controls.

We did not observe significant associations ($P > 0.5$, Supplementary Table S2) with the five *EBF1* SNPs in any ethnic group other than in Whites, which was the largest ethnic group within the MESA data set. In addition, we observed that in all other ancestries in the MESA data set (ie, Chinese Americans, Blacks, and Hispanics), mean hip circumference was similar across various stress levels for both the major allele homozygote (TT) and minor allele (CT/CC) carriers for the lead *EBF1* SNP rs4704963:T>C (Supplementary Figure S8). Furthermore, there were no observations of high stress in Blacks for the individuals carrying a minor allele (CT/CC). The inability to replicate the gene-by-stress interaction in other ethnic groups could be because of a number of factors including type I error or winner's curse in the Whites, low power owing to smaller sample sizes in the other ethnicities, differential effects of stress in the different ethnic groups, or allelic heterogeneity across the *EBF1* gene. Ethnic-specific genetic interactions with physiological measures in other settings have also been observed^{35,36} and additional studies will be required to understand these differences.

The functional role of *EBF1* might be anticipated given its role as a transcription factor; however, the mechanism by which stress would modify the *EBF1* genetic risk is unknown. The results of this study suggest that evaluation of the impact of stress in the mouse model system may elucidate the function of *EBF1*, especially in view of a SNP from our findings being part of a putative conserved transcription factor binding site (tfbscons). Specifically, gene regulation and methylation would be a worthwhile topic for further research. The pattern of expression of *EBF1* and its networked genes strongly suggest that, if the association of *EBF1* variants with CAD holds after additional validation, the developmental mechanism²⁸ may be through the association of obesity and diabetes as risk factors for CAD. These observations further suggest that a stress-reduction intervention aimed at reducing risk of CAD may be most effective in individuals with the rs4704963 CC/CT genotypes. We would also postulate that CAD risk reduction would include reductions in central adiposity and related modification of metabolic risk factors.

In conclusion, in addition to the known functional role of *EBF1* in the development of adipose tissue, we have shown that in the presence of chronic psychological stress, the common variation in this gene contributes to interindividual differences in human obesity, diabetes, and cardiovascular disease risk factors in two population-based cohorts. Although more functional work is needed to characterize the exact role of *EBF1* in development of type II diabetes and CVD risk, the multiple lines of evidence suggest that the *EBF1* gene plays an important role in the pathways leading to type II diabetes and CVD. Thus, in this study, we identified *EBF1* as a gene that plays a role in

connecting chronic psychosocial stress, central obesity as a risk factor for CVD, and ultimately CVD disease development.

CONFLICT OF INTEREST

Redford B Williams is a founder of and stockholder in Williams LifeSkills, Inc. The other authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by NIH/NHLBI grants HL095987 (to Shah), HL073389-01 (to Hauser), and HL036587 (to Williams). The MESA and Framingham data sets were obtained from NIH dbGaP.

- Go AS, Mozaffarian D, Roger VL *et al*. Heart disease and stroke statistics-2013 update: a report from the American Heart Association. *Circulation* 2013; **127**: e6–e245.
- World Health Organization: Global health risk: mortality and burden of disease attributable to selected major risks. *Tech. Rep. ISBN 978 92 4 156387 1*. Geneva, Switzerland: WHO Press, 2009.
- Elder SJ, Lichtenstein AH, Pittas AG *et al*. Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. *J Lipid Res* 2009; **50**: 1917–1926.
- Rosengren A, Hawken S, Ōunpuu S *et al*. Association of psychosocial risk factors with risk of acute myocardial infarction in 11,119 cases and 13,648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* 2004; **364**: 953–962.
- Williams RB: Psychosocial and biobehavioral factors and their interplay in coronary heart disease. *Annu Rev Clin Psychol* 2008; **4**: 349–365.
- Williams RB, Barefoot JC, Blumenthal JA *et al*. Psychosocial correlates of job strain in a sample of working women. *Arch Gen Psychiatry* 1997; **54**: 543–548.
- Scherwitz LW, Perkins LL, Chesrtey MA, Hughes GH, Sidney S, Manolio TA: Hostility and health behaviors in young adults: the CARDIA Study. *Coronary Artery Risk Development in Young Adults Study*. *Am J Epidemiol* 1992; **136**: 136–145.
- Siegler IC, Peterson BL, Barfoot JC, Williams RB: Hostility during late adolescence predicts coronary risk factors at mid-life. *Am J Epidemiol* 1992; **136**: 146–154.
- Speaker KJ, Fleshner M: Interleukin-1 beta: a potential link between stress and the development of visceral obesity. *BMC Psychiatry* 2012; **12**: 1–15.
- Labonte B, Yerko V, Gross J *et al*. Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol Psychiatry* 2012; **72**: 41–48.
- Hunter DJ: Gene-environment interactions in human diseases. *Nat Rev Genet* 2005; **6**: 287–298.
- Thomas D: Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet* 2010; **11**: 259–272.
- Bornstein SR, Schuppenies A, Wong ML, Licinio J: Approaching the shared biology of obesity and depression: the stress axis as the locus of gene-environment interactions. *Mol Psychiatry* 2006; **10**: 892–902.
- Bild DE, Bluemke DA, Burke GL *et al*. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 2002; **156**: 871–881.
- Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M: Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis. *Circulation* 2007; **115**: 459–467.
- Purcell S, Neale B, Todd-Brown K *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 560–575.
- Price AL, Patterson NJ, Plenge RM *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
- Devlin B, Bacanu SA, Roeder K: Genomic control to the extreme. *Nat Genet* 2004; **36**: 1129–1130.
- Muthen LK, Muthen BO: *Mplus User's Guide Sixth Edition* Muthen & Muthen: Los Angeles, CA, 1998–2010.
- Streiner DL: Finding our way: an introduction to path analysis. *Can J Psychiatry* 2005; **50**: 115–122.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP: The Framingham Offspring Study. Design and preliminary data. *Prev Med* 1975; **4**: 518–525.
- Lukin K, Fields S, Hartley J, Hagman J: Early B cell factor: regulator of B lineage specification and commitment. *Semin Immunol* 2008; **20**: 221–227.
- Nolan DK, Sutton B, Haynes C *et al*. Fine mapping of a linkage peak with integration of lipid traits identifies novel coronary artery disease genes on chromosome 5. *BMC Genet* 2012; **13**: 12.
- Hauser ER, Crossman DC, Granger CB *et al*. A genomewide scan for early-onset coronary artery disease in 438 families: the GENECARD study. *Am J Hum Genet* 2004; **75**: 436–447.
- Shah SH, Kraus WE, Crossman DC *et al*. Serum lipids in the GENECARD study of coronary artery disease identify quantitative trait loci and phenotypic subsets on chromosomes 3q and 5q. *Ann Hum Genet* 2006; **70**: 738–748.
- International Consortium for Blood Pressure Genome-Wide Association Studies: Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**: 103–109.

- 27 Wain LV, Verwoert GC, O'Reilly PF *et al*. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet* 2011; **43**: 1005–1011.
- 28 Huan T, Zhang B, Wang Z *et al*. A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb Vasc Biol* 2013; **33**: 1427–1434.
- 29 Speliotes EK, Willer CJ, Berndt SI *et al*. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
- 30 Manning AK, Hivert MF, Scott RA *et al*. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**: 659–669.
- 31 Morris AP, Voight BF, Teslovich TM *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981–990.
- 32 Schunkert H, König IR, Kathiresan S *et al*. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; **43**: 333–338.
- 33 Fretz JA, Nelson T, Xi Y *et al*. Altered metabolism and lipodystrophy in the early B-cell factor 1-deficient mouse. *Endocrinology* 2010; **151**: 1611–1621.
- 34 Jimenez MA, Akerblad P, Sigvardsson M, Rosen ED: Critical role for Ebf1 and Ebf2 in the adipogenic transcriptional cascade. *Mol Cell Biol* 2007; **27**: 743–757.
- 35 Williams RB, Marchuk DA, Gadde KM *et al*. Serotonin-related gene polymorphisms and central nervous system serotonin function. *Neuropsychopharmacology* 2003; **28**: 533–541.
- 36 Gong M, Hubner N: Molecular genetics of human hypertension. *Clin Sci* 2006; **110**: 315–326.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)