



HHS Public Access

Author manuscript

Genet Epidemiol. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Genet Epidemiol. 2015 September ; 39(6): 489–497. doi:10.1002/gepi.21910.

Computing a Synthetic Chronic Psychosocial Stress Measurement in Multiple Datasets and its Application in the Replication of GxE Interactions of the *EBF1* Gene

Abanish Singh^{1,2,3,*}, Michael A. Babyak^{1,3}, Beverly H. Brummett^{1,3}, Rong Jiang^{1,3}, Lana L. Watkins³, John C. Barefoot³, William E. Kraus^{2,4,5}, Svati H. Shah^{2,4}, Ilene C. Siegler^{1,3}, Elizabeth R. Hauser^{2,6,7}, and Redford B. Williams^{1,3}

¹Behavioral Medicine Research Center, Duke University Medical Center, Durham, NC, USA

²Duke Molecular Physiology Institute, 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

⁶Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA

⁷Cooperative Studies Program Epidemiology Center, Durham Veterans Affairs Medical Center, Durham, NC, USA

Abstract

Chronic psychosocial stress adversely affects health and is associated with the development of disease [Williams, 2008]. Systematic epidemiological and genetic studies are needed to uncover genetic variants that interact with stress to modify metabolic responses across the life cycle that are the proximal contributors to the development of cardiovascular disease and precipitation of acute clinical events. Among the central challenges in the field are to perform and replicate gene-by-environment (GxE) studies. The challenge of measurement of individual experience of psychosocial stress is magnified in this context. Although many research datasets exist that contain genotyping and disease-related data, measures of psychosocial stress are often either absent or vary substantially across studies. In this paper, we provide an algorithm to create a synthetic measure of chronic psychosocial stress across multiple datasets, applying a consistent criterion that uses proxy indicators of stress components. We validated the computed scores of chronic psychosocial stress by observing moderately strong and significant correlations with the self-rated chronic psychosocial stress in the MESA Cohort ($Rho = 0.23$, $P < 0.0001$) and with the measures of depressive symptoms in five datasets ($Rho = 0.15 - 0.42$, $P_s = 0.005 - < 0.0001$) and by

*Correspondence: Abanish Singh, Ph.D., Duke Molecular Physiology Institute, 300 N Duke Street, Box 104775, Durham, NC 27701, Tel: +1 919 681 2133, Fax: +1 919 684 0934.

Conflict of interest

Redford Williams is a founder and major stockholder of Williams LifeSkills, Inc. and holds a patent on the use of the 5HTTLPR L allele as a marker of stress-related CVD. Other authors have no conflicts of interest in relation to the work.

comparing the distributions of the self-rated and computed measures. Finally, we demonstrate the utility of this computed chronic psychosocial stress variable by providing three additional replications of our previous finding of gene-by-stress interaction with central obesity traits [Singh et al., 2015].

Keywords

Chronic psychosocial stress; Central obesity; CVD risk gene; *EBF1*; Gene-by-stress interaction

Introduction

Environmental factors, such as psychosocial stress, influence the expression of health behaviors, psychological traits, and neuroendocrine and autonomic functions in ways that alter metabolic, hemostatic, inflammatory and cardiovascular functions [Williams, 2008]. Psychosocial stress is defined as a condition in which environmental demands exceed the resources of the individual [Lazarus, 1966]. The condition may be acute or chronic. The response to acute stress in young and healthy individuals may be adaptive without a burden [Garnezy, 1991; Glantz and Johnson, 1999]. However, chronic psychosocial stress that involves long lasting exposure to challenging environmental stimuli, *i.e.* aversive or demanding conditions over a significant time period [Lazarus, 1966], is associated with adverse human health [Schneiderman et al., 2005]. The INTERHEART study suggests that a higher prevalence of stress and other psychosocial factors, such as depression, account for 34% of the population attributable risk for myocardial infarction, independently of physical risk factors [Rosengren et al., 2004]. Also, stress is a key phenomenon in threats to homeostasis and adaptive responses; the associations between stressors – such as job difficulties, marital problems, and health problems – appear to be mediated by endocrine-immune interactions [Schneiderman et al., 2005]. Given the strong biological basis for the effects of stress on cardiovascular disease risk factors, genetic interaction with stress is a reasonable risk factor model.

The role of chronic psychosocial stress in the development of cardiovascular disease has attracted considerable interest in the recent past [Williams, 2008; Rozanski et al., 1999; Cohen et al., 2007; Kaplan, 2009], including a successful gene-by-stress genome-wide association study (GWAS) [Singh et al., 2015]. However the assessment of chronic psychosocial stress remains an important problem. One solution to the measurement problem has been the use of standard questionnaires to assess the presence or absence of stress [Kamarck, 2012]. A large number of completed studies have vast amounts of phenotypic and genotypic data available for analysis (Framingham Heart Study Cohort, CATHGEN Cohort, etc.), but that did not obtain data expressly assessing chronic psychosocial stress. In these cases, administering questionnaires retrospectively is impractical or impossible. The lack of widely applied stress measures hampers study of this important risk factor as well as hinders the ability to replicate gene-by-stress interaction findings. Identifying a means by which stress can be assessed in a reliable and valid fashion in the absence of a formal measure holds promise to greatly improve our ability to study its role in the association between genes and disease.

In our previous study [Singh et al., 2015] we applied our proposed method to create a chronic psychosocial stress score in the Framingham Offspring Cohort. In that study we identified a gene-by-stress interaction with common variation in the *EBF1* gene (lead single nucleotide polymorphism (SNP) rs4704963) that contributes to inter-individual differences in human central obesity traits (e.g., hip and waist circumferences and body mass index (BMI)) in the presence of chronic psychosocial stress [Singh et al., 2015]. The *EBF1* gene is a transcription factor with a known hematopoietic function and it has a critical role in the adipogenic transcriptional cascade in multiple cellular models [Jimenez et al. 2007] beside its role in the development of the immune system [Lukin et al. 2008]. In this report, we describe this method in detail, providing a complete generalized algorithm and framework to identify proxy indicators of the components of chronic psychosocial stress and to synthetically infer a consistent chronic psychosocial stress measure using these indicators, along with its validation and application in three additional replications (i.e., Family Heart Study, CATHGEN Cohort, and Duke Caregiver Study) of our gene-by-stress interaction finding [Singh et al., 2015]. We show the utility of this relatively simple method to estimate chronic psychosocial stress that is easily extended to development of other inconsistently-measured environmental covariates for use in GxE studies.

Methods and Material

Computation

Our goal was to develop a measure or a summary score of chronic psychosocial stress for the use in gene-by-stress interaction analyses to allow replication of our GxE finding [Singh et al., 2015]. We generalized our method to develop this measure using proxy indicators of the following five components, if available in a dataset: financial strains, relationship or marital problems, difficulties with job or ability to work, serious health problems of spouse or someone close, and one's own serious health problems. These components were similar to the domains of the formal self-rated chronic psychosocial stress (chronic burden) items in the MESA dataset [Shivpuri et al., 2012] which were derived from a composite stress measure originally developed in the Study of Women's Health Across the Nation [Troxel et al., 2003]. Our approach included four steps to compute a consistent combined chronic psychosocial stress measure across multiple datasets:

1. *Identify the indicators of chronic psychosocial stress components:* If the self-rated specific items of MESA-like chronic psychosocial stress were not available, we searched for other proximal indicators (datapoints) in the protocol, whose language content was equivalent to a given stress component and these equivalent items were treated as primary datapoints. If any of the primary datapoints were not available in the dataset, we attempted to find more distal indicators and treated them as secondary or tertiary datapoints for inferring the stress components. For example, if a dataset did not have a self-rated answer to a financial strain question but included a financial difficulties construct, that construct was used as the financial stress component. However, if this construct was not available, we used the household income to infer the financial difficulties or strain. We collected a list of main questions, primary, secondary and tertiary indicators for the five components of chronic psychosocial stress (Table 1).

2. *Search the availability of indicators in the datasets:* Finding the needed data variables in public access datasets can be quite cumbersome due to inadequate search mechanisms provided on the study web-pages and the non-intuitive variable names. We implemented a computational method to search multiple keywords available in the dataset files as variable names or variable descriptions using Perl v5.10.1. We parsed the text for data variables name and description into hash table indices (*i.e.*, a data structure in computer programming) [Lewis and Cook, 1988] and matched multiple keywords of interest with each entry of the hash data structure using regular expressions (*i.e.*, a method of matching pattern or sequence of characters) [Thompson, 1968]. The search method matched the patterns of all variations of input keywords with the text from dataset files and output the text that wrapped around the patterns. We reviewed the search outputs and identified the relevant data variables that were needed as indicators of chronic psychosocial stress components.
3. *Choose the best available indicators:* We reviewed the variables obtained from the previous step and chose the best available indicators in the dataset based on a priority selection, *i.e.*, starting from primary to tertiary datapoints as described in Table 1.
4. *Generate a combined chronic psychosocial stress score:* After selecting indicators we transformed them to binary (0/1) form in such a way that '0' corresponded to low stress and '1' corresponded to high stress, by splitting continuous or ordinal variables based on their distribution. For example, if we used household "income" for financial strain, we split the income at the median and assigned 0 and 1 to financial strain if income was greater than or equal to the median and less than the median, respectively. If there was more than one distal indicator for a component with dissimilar scales, we converted them to a binary indicator and performed a Boolean "AND" operation (*i.e.*, for the two binary variables X and Y, $X \text{ AND } Y = 1$, if $X=Y=1$, else $X \text{ AND } Y = 0$). If there were multiple same scale questionnaires (such as a job difficulties questionnaires), we summed the responses and then created a binary variable. Finally, we summed the components to get a combined ordinal chronic psychosocial stress score that could range from 0 to maximum 5, if all five components were available for a study.

In MESA the chronicity of self-rated chronic psychosocial stress measure was determined if the problems in five components lasted for six or more months. Here we assumed that the psychosocial, socioeconomic and demographic indicators that we used for computing the measure lasted for six or more months.

Validation

We evaluated polychoric (tetrachoric) correlations between binary components of computed and self-rated measures of chronic psychosocial stress in the MESA dataset, and also performed exploratory and confirmatory factor analyses on the binary variables using Weighted Least Squares Means and Variances (WLSMV) in *Mplus* version 6.11 [Muthen and Muthen, 1998–2010]. We compared the frequency distributions of the computed scores

in the five datasets (i.e., MESA, Framingham Offspring, Family Heart Study, CATHGEN Cohort, and Duke Caregiver Study, as described in Datasets subsection) with the frequency distributions of self-rated chronic stress measure in MESA and evaluated expected correlation (Spearman's correlation coefficient) with a depression measure (such as CES-D, Beck depression inventory), if available in a dataset.

Application

Finally, we performed the ancestry-specific replication of our previously identified gene-by-stress interaction of *EBF1* genetic variations (SNP rs4704963 if available, or LD SNP rs17056278, $R^2 = 1$) with central obesity traits [Singh et al., 2015] in five datasets using the derived chronic stress measure. We performed linear regression on the hip circumference (or BMI, if hip circumference was not available) for the *EBF1* SNP rs4704963 (or SNP rs17056278) under the additive genetic model with age and sex adjustment and population ancestry correction (if the data were available), and clustering family IDs in case the dataset included related individuals such as in the Framingham dataset. We used hip circumference as our primary phenotype based on our correlation-based phenotype selection approach where we observed the strongest correlation between psychosocial stress and hip-circumference in the MESA dataset, as described in the original discovery GWAS work [Singh et al., 2015]. The GxE interaction was tested by including a SNPxSTRESS product term in the model. The ordinal stress variable was treated as a linear variable in the model. The gene-by-stress interaction was considered significant at the threshold P -value < 0.05 for the single SNP analysis. We also plotted the distribution of the mean of hip circumference (or BMI, if hip circumference was not available) against each ordinal value of stress for the two genotype groups of *EBF1* SNP, i.e., major allele homozygotes and minor allele heterozygotes and homozygotes. We used Fisher's combined probability test to obtain the combined GxE P -value for the *EBF1* SNP using the computed chronic psychosocial stress scores in the five datasets. Unless specified otherwise, all analyses were done in STATA SE 11.1.

Datasets

We applied the foregoing method of computing a chronic psychosocial stress measure and performing subsequent analysis on the following datasets, using the available relevant psychosocial and socioeconomic data, genotypic data of *EBF1* SNP rs4704963 (or LD SNP rs17056278) genotyped in genome-wide array or as individually-genotyped candidate SNPs, and phenotypic data for central obesity trait (hip circumference or BMI):

Multi-Ethnic Study of Atherosclerosis (MESA)—The original GWAS study [Singh et al., 2015] used the MESA dataset because it had a self-rated chronic psychosocial stress (chronic burden) variable [Bild et al., 2002, Shivpuri et al., 2012]. In the current work, we used this dataset to compare the self-rated chronic psychosocial stress measure with computed chronic psychosocial stress scores. In the MESA Cohort a total of 5,805 individuals – 2,460 Whites, 548 Chinese Americans, 1,547 Blacks, and 1,250 Hispanics – had quality controlled genotype and phenotype data available. The genotyping in MESA was done using the Affymetrix Genome-Wide Human SNP Array 6.0.

Framingham Offspring Cohort—We used the Generation 2 (or Offspring) dataset from Framingham Heart Study Cohort [Feinleib et al., 1975] for this work. The cohort is primarily White. A total of 3,157 individuals had both phenotype and quality controlled genotype data available, comprising 1,515 males and 1,642 females. SHARe Illumina genotyping of genome-wide array was provided under an agreement between Illumina and Boston University using A ymetrix Mapping250K (Nsp and Sty) Arrays and Mapping50K (Hind240 and Xba240) Arrays.

Family Heart Study—The study was conducted at Duke University Medical Center under the approval of Duke IRB. A total of 578 participants (220 males and 358 females) were recruited between August 2004 to September 2008 to study the effect of genetic variation on the relationship between psychosocial and cardiovascular risk factors [Brummett et al., 2010]. Genotyping of candidate SNPs was done using the ABI 7900 Taqman@system.

CATHGEN Cohort—The participants in the CATHeterization GENetics (CATHGEN) cohort [Sutton et al., 2008] were recruited through the cardiac catheterization laboratories at Duke University Hospital (Durham, NC, USA). The cohort included a total of 9,181 subjects (5,700 males, 3,481 females) with phenotypic and genotypic data. Three-quarters of the cohort was diagnosed with clinically-significant coronary artery disease (CAD), one-quarter (*i.e.* 1,792 total; 1,196 Whites) of the samples did not have angiographically-defined CAD and were used as controls. Candidate SNPs were genotyped using the Taqman genotyping assay (Life Technologies) and the Type-It Fast Probe PCR kit (Qiagen). CATHGEN has no individual-level psychosocial and socioeconomic data. We used the US National Census's block-wise socioeconomic data (<http://www.census.gov/>) [Ward-Caviness, 2014] for the participant residential address at the time of catheterization to infer the indicator for the financial strain component. We also used the Beck depression inventory (available for 443 CATHGEN participants through separate studies VAGUS [Watkins et al., 2010] and REACH [Barefoot JC, *personal communication*]).

Duke Caregiver Study—This study conducted at the Duke University Medical Center included data from 344 persons, a total of 175 were family caregivers (126 Whites, 49 Blacks) of a relative with Alzheimer's Disease or other dementia and 169 were non-caregiving controls (122 Whites, 47 Blacks) [Kring et al., 2010]. The genotyping of candidate SNPs was done using the ABI 7900 Taqman system (Applied Biosystems, Carlsbad, CA, USA).

Results

Using the MESA definition of chronic stress [Shivpuri et al., 2012], we prepared a matrix of chronic psychosocial stress components, the original question, and the list of indicators that address that component. We listed these indicators as primary, secondary and tertiary datapoints as shown in Table 1. We used the list of these datapoints and their priority (*i.e.* primary to tertiary) as a guide to identify indicators of stress components in case the self-rated answers to related main questions were not available in the datasets.

To compute the chronic psychosocial stress score in the MESA Cohort, we ignored all the main question item level data that were used for the self-rated measure (Figure 1A), and using steps 2 and 3 of our generalized method, we identified total gross family income to derive the per-household indicator (secondary) of the financial strain component; marital separation and divorce for the indicators of marital problems; questions on job security, 'job requires working very fast', and 'asked to do an excessive amount of work' for the indicators of work related difficulties component; and general health status for the indicator (primary) of health problems of one's own (Table 2). Using step 4 as described in the Methods, we created a binary variable for the indicator(s) of each stress component and summed four (out of a total of five) binary variables into one ordinal chronic psychosocial stress measure (Figure 1B).

In the Framingham Offspring dataset (Figure 1C), we identified total family income for the indicator of the financial strain component; job insecurity and physiological job demand scale for the indicators of work related difficulties component; marital disagreement for the indicator of relationship or marital problems; and spouse's heart attack, stroke and heart disease-related death for the indicators of serious health problems of spouse.

In the Family Heart Study (Figure 1D), we identified total income for the indicator of the financial strain component; and the work difficulties component was scored based on 10 questions on job insecurity, lack of career prospects, issues with support at work (supervisor and others), and job dissatisfaction (*i.e.*, a total two components out of five).

In the CATHGEN Cohort (Figure 1E), we did not have a direct measure of financial strain; therefore, to infer it we used multiple block-wise indicators – median house income, per capita income, and median home value – from the US National Census's socioeconomic data of the participant's residential address at the time of catheterization [Ward-Caviness, 2014]. We also did not have a direct measure of marital problems. We inferred marital problems from the marital status information (marital separation, divorce, etc.). The dataset did not have indicators for work difficulties component and health problem related components. The computed score ranged from 0 to 2.

In the Duke Caregiver Study (Figure 1F), we identified total household income for the indicator of financial strain component. We used spouse related hassles for the indicator of marital problems, hassles related to health or well-being of a family member for the indicator of health problems of spouse or someone close, and self-health related hassles for the indicator of one's own serious health problems, The work difficulties component was scored based on hassles related to job security, meeting deadlines or goals (*i.e.*, job demand), and hassles related to supervisor and fellow workers (*i.e.* support at work), to be consistent with other datasets.

As shown in Figure 1, although there were different ordinal levels, the computed chronic psychosocial stress measure in MESA (median=1, mean=0.94, SD=0.85, variance=0.72, skewness=0.60, kurtosis=2.83, N=5,194), Framingham (median=1, mean=0.84, SD=0.78, variance=0.64, skewness=0.54, kurtosis=2.44, N=3,326), CATHGEN (median=0, mean=0.56, SD=0.63, variance=0.40, skewness=0.69, kurtosis=2.48, N=7,016), Family

Heart Study (median=1, mean=0.65, SD=0.69, variance=0.45, skewness=0.53, kurtosis=2.26, N=546), and Duke Caregiver Study (median=1, mean=1.34, SD=1.12, variance=1.25, skewness=0.74, kurtosis=3.11, N=321) demonstrated distributions similar to the self-rated chronic psychosocial stress measure in MESA (median=1, mean=1.21, SD=1.20, variance=1.44, skewness=0.95, kurtosis=3.43, N=5,805). The distributions were flat (kurtosis=2.26–3.43) and slightly skewed towards the right (skewness=0.53–0.95). The overall similarities in frequency distributions for computed (MESA, Framingham Offspring, CATHGEN, Family Heart Study, and Duke Caregiver Study) and self-rated (MESA) chronic psychosocial stress measures support the consistency of our approach.

Table 3 shows the Rho and *P*-values of polychoric correlations (tetrachoric for binary variables) between the computed and self-rated chronic psychosocial stress ordinal measures and the binary component items in MESA Cohort. All correlations were moderate in magnitude but significantly different from zero. The exploratory and confirmatory factor analyses showed that all computed and self-rated binary component items loaded on a single factor (Table 4). As expected, loadings from CFA and EFA were highly similar, as was evidence for unidimensionality. Although some loadings were modest in magnitude, most notably the computed financial strain item, the null hypothesis (loading = 0) was rejected for all indicators.

The Spearman's correlation coefficients and *P*-values of these chronic psychosocial stress measures with depression measures available in MESA, Framingham Offspring, Family Heart Study, CATHGEN, and Duke Caregiver Study are shown in Table 5. As expected, all these correlations were reasonably strong (coefficients 0.15 – 0.42) and significantly different from zero and of the same order as observed for the MESA self-rated measure. The Duke Caregiver Study showed the strongest correlation. In CATHGEN where we used the block-wise socioeconomic data rather than individual-level data and where depression data was available only for a small number of samples (N=443), the correlation was somewhat weaker but still statistically significant. These correlations demonstrate the consistency of our approach of inferring the chronic psychosocial stress using the proxy indicators of its components (Table 1).

The ultimate goal of our work was to perform a replication of our previously reported (MESA Whites (Self-rated) and Framingham Offspring) finding of gene-by-stress interaction with hip circumference [Singh et al., 2015] using computed chronic psychosocial stress measures in MESA Whites and three additional datasets, *i.e.*, the Family Heart Study Whites, CATHGEN Whites, and Duke Caregiver Study Whites. Table 6 shows the *P*-values of the SNP term in the simple regression model and SNPxSTRESS term in the gene-by-stress interaction regression model. In addition to the MESA Whites (Self-rated) and Framingham Offspring [Singh et al., 2015], gene-by-stress interactions were significant in MESA Whites (Computed), Family Heart Study Whites, and at a trend level (*P*=0.064) in Duke Caregiver Study Whites. In CATHGEN Whites we did not observe a significant Gx \times E interaction for the entire cohort. However, when we limited our analysis to only CATHGEN Control Whites (*i.e.* the part of cohort that did not have clinically-significant CAD and, thus, similar to the other datasets), we observed a significant gene-by-stress interaction with BMI. The combined Gx \times E *P*-value for the computed chronic psychosocial stress scores in MESA

Whites, Framingham Offspring, Family Heart Study, Duke Caregiver Study, and CATHGEN Whites was $9.81E-07$. Figure 2 shows the direction of gene-by-stress association, *i.e.*, the mean hip circumference (or BMI) increased with computed chronic psychosocial stress for the rs4704963 minor allele heterozygotes and homozygotes (CT/CC) but not for the major allele homozygotes (TT) in the MESA Whites (computed), Family Heart Study Whites, CATHGEN Control Whites, and Duke Caregiver Study, similar to the MESA Whites (self-rated) and Framingham Offspring, *i.e.*, our original finding [Singh et al., 2015]. Gene-by-stress interactions with *EBF1* SNPs were not significant (GxE $P_s > 0.42$, data not shown) among Blacks in the original MESA GWAS analysis nor in the three replication datasets (Family Heart Study, CATHGEN and Duke Caregiver Study).

Discussion

We developed a 4-step method to create a consistent MESA-like chronic psychosocial stress measure using proxy indicators (Table 1) across multiple datasets for use in GxE replication studies. We applied the method in the MESA Cohort, Framingham Offspring Cohort, Family Heart Study, CATHGEN Cohort, and Duke Caregiver Study. Although our original finding of gene-by-stress interaction in MESA Whites and Framingham Offspring using self-rated and computed chronic psychosocial stress, respectively, were included in our earlier paper [Singh et al., 2015], we reused these datasets to demonstrate the method and to provide a comparison of self-rated and computed chronic psychosocial stress along with the replication outcomes in the three additional datasets available to us.

We assessed the validity of our measure in three ways. We found distributions of computed scores in the five datasets similar to the distribution of self-rated measure in the MESA dataset (Figure 1). We observed strong to moderate correlations between the computed and self-rated chronic psychosocial stress and its binary components in MESA (Table 3) that were loaded on the same factor (Table 4). The loading and correlation for computed financial strain derived from household income were comparatively weak. This may be due, in part, to the fact that sample collection in MESA was done at six different sites with varied income distribution; however, adjusting analysis for site did not significantly alter the results. Although a more esoteric analysis -- testing for equivalence of loadings and error variances -- showed that the loadings were not equivalent, the computed components were still interpretable as arising from the same underlying latent variable. Finally, we assessed the well-known correlation between the computed score of chronic psychosocial stress and a depression measure (e.g., CES-D or Beck Inventory, Table 5) in the datasets.

In addition to our earlier finding of gene-by-stress interaction with central obesity in the MESA Whites and Framingham Offspring [Singh et al., 2015], we observed a similar significant interaction in the Family Heart Study Whites and CATHGEN Control Whites and at a trend level interaction in Duke Caregiver Study Whites (Table 6) with the similar direction of association, *i.e.*, central obesity increased with the chronic psychosocial stress only for minor allele groups of *EBF1* SNP rs4704963 (or SNP rs17056278, Figure 2). We did not observe a significant interaction in the CATHGEN cohort Whites, the only group with a high proportion of subjects with clinically-significant CAD. This result could be due, in part, to three possible reasons: 1) use of non-specific block-level socioeconomic data for

inferring the stress measure did not capture the chronic stress construct, 2) the high obesity -- mean BMI for the CATHGEN samples was the highest among all at the available stress levels and disease prevalence in all genotypic groups of the CATHGEN cohort obscures the interaction, or 3) the original interaction was a false-positive association. The analysis on CATHGEN samples indicates that relatively healthy and non-obese populations may be more susceptible to a gene-by-stress interaction on central obesity. The small sample size in the Duke Caregiver Study may explain the lack of replication of the GxE interaction at the conventional 0.05 significance level, as this dataset was the smallest. Consistent with our original results [Singh et al., 2015], we did not observe a significant gene-by-stress interaction in Black samples that were part of three replication datasets.

Our approach offers additional value to datasets that did not include a self-reported chronic psychosocial stress measure. In such datasets, a computed chronic psychosocial stress score can be used to study gene-by-stress interactions to discover cardiometabolic disease genes that are moderated by chronic psychosocial stress. Evaluation of gene-by-stress interactions can be critical to fully understand the mechanisms and pathways of underlying complex diseases [Williams, 2008]. However, this method is not a substitute for the formal measures of self-rated chronic psychosocial stress and there are several limitations. While we attempt to maximize the set of inferred indicators, it may not always be possible to achieve a complete set of these indicators due to missing variables in individual datasets. A limitation of our approach is that we have chosen not to impute or adjust for missing indicators. Developments in methods for behavioral science research offers several approaches for improving the construction of a measure of chronic psychosocial stress including integrative data analysis [Curran and Hussong, 2009] or multiple imputation methods [Sterne et al., 2009], which will be incorporated in future work. The observed polychoric correlations between the MESA chronic stress self-rated score and computed scores using worst case two ($Rho = 0.15$), three ($Rho=0.19$) and all four available ($Rho= 0.23$) components were close and suggested that the synthetic score developed using incomplete set of inferred indicators could still be useful. Another area of improvement in the approach could be in the selection of split-points for indicating the high-stress category. The choice of the median for a split-point (i.e., in income) increases the sample size in the high-stress group however dividing at the median may not be optimal. Sensitivity analysis to the choice of this split-point or development of optimal split-points remains to be evaluated in future work.

In conclusion, we have demonstrated that even when a self-rated chronic psychosocial stress score is not collected when a study is initially performed, it is possible to use available data to compute this score retrospectively. This method worked well with small studies such as Family Heart Study and the Duke Caregiver Study and large studies such as the MESA, Framingham, and CATHGEN Cohorts. Importantly, we have also strengthened our earlier finding on gene-by-stress interaction with central obesity traits [Singh et al., 2015] by replicating it in three additional datasets. These replications provide further confirmation that the common variation in *EBF1* contributes to inter-individual differences in human obesity in the presence of chronic psychosocial stress.

Acknowledgments

This work was supported by NIH/NHLBI grant HL036587 (Williams). The MESA and Framingham public access datasets were obtained from NIH dbGaP under the standard data user agreement. CATHGEN Cohort, Family Heart Study, and Duke Caregiver Study datasets were obtained from Duke data repositories. The block-wise socioeconomic data for CATHGEN analysis was obtained from US National Census Bureau (<http://www.census.gov/>) [Ward-Caviness, 2014], and Beck depression inventory data for 443 CATHGEN participants was obtained from separate studies VAGUS [Watkins LL et al., 2010] and REACH [Barefoot JC, *personal communication*]. We thank the investigators, staff, and participants of the public access and in-house datasets for their valuable contributions.

References

- Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacobs DR Jr, Kronmal R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *American Journal of Epidemiology*. 2002; 156(9):871–881. [PubMed: 12397006]
- Brummett BH, Boyle SH, Ortel TL, Becker RC, Siegler IC, Williams RB. Associations of depressive symptoms, trait hostility, and gender with C-reactive protein and interleukin-6 response after emotion recall. *Psychosom Med*. 2010; 72(4):333–9. [PubMed: 20190126]
- Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. *JAMA*. 2007; 298(14):1685–1687. [PubMed: 17925521]
- Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*. 2009; 14:81–100. [PubMed: 19485623]
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 1975; 4:518–525. [PubMed: 1208363]
- Garnezy N. Resiliency and vulnerability to adverse developmental outcomes associated with poverty. *Am Behav Sci*. 1991; 34:416–430.
- Glantz, MD.; Johnson, JL. *Resilience and Development: Positive Life Adaptations*. New York: Kluwer Acad./Plenum Publishers; 1999.
- 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. [PubMed: 17060461]
- Kamarck T. Psychosocial stress and cardiovascular disease: An exposure science perspective. *Psychological Science Agenda*. 2012; 26(4)
- Kaplan JR, Chen H, Manuck SB. The Relationship between Social Status and Atherosclerosis in Male and Female Monkeys as Revealed by Meta-Analysis. *American Journal of Primatology*. 2009; 71(9):732–741. [PubMed: 19452517]
- Kring SI, Brummett BH, Barefoot J, Garrett ME, Ashley-Koch AE, Boyle SH, Siegler IC, Sorensen TI, Williams RB. Impact of psychological stress on the associations between apolipoprotein E variants and metabolic traits: findings in an American sample of caregivers and controls. *Psychosomatic Medicine*. 2010; 72:427–433. [PubMed: 20467002]
- Lazarus, RS. *Psychological stress and the coping process*. New York: McGraw-Hill, Inc; 1966.
- Lewis TG, Cook CR. Hashing for dynamic and static internal tables. *Computer*. 1988; 21(10):45–56.
- Lukin K, Fields S, Hartley J, Hagman J. Ealy B cell factor: regulator of B lineage specification and commitment. *Semin Immunol*. 2008; 20:221–227. [PubMed: 18722139]
- Muthen, LK.; Muthen, BO. *Mplus User's Guide*. 6. Muthen & Muthen; Los Angeles, CA: 1998–2010.
- Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, Blackett KN, Sitthi-amorn C, Sato H, Yusuf S. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet*. 2004; 364(9438):953–962. [PubMed: 15364186]
- Rozanski A, Blumenthal JA, Kaplan J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation*. 1999; 99:2192–2217. [PubMed: 10217662]
- Schneiderman N, Ironson G, Siegel SD. Stress and Health: Psychological, Behavioral, and Biological Determinants. *Annual Review of Clinical Psychology*. 2005; 1:607–628.

- Shivpuri S, Gallo LC, Crouse JR, Allison MA. The association between chronic stress type and C-reactive protein in the Multi-Ethnic Study of Atherosclerosis (MESA): Does gender make a difference? *J Behav Med.* 2012; 35(1):74–85. [PubMed: 21503709]
- Singh A, Babyak MA, Nolan DK, Brummett BH, Jiang R, Siegler IC, Kraus WE, Shah SH, Williams RB, Hauser ER. Gene by stress genome-wide interaction analysis and path analysis identify EBF1 as a cardiovascular and metabolic risk gene. *European Journal of Human Genetics.* 2015; 23:854–862. [PubMed: 25271088]
- Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009; 338 :b2393. [PubMed: 19564179]
- Sutton BS, Crosslin DR, Shah SH, Nelson SC, Bassil A, Hale AB, Haynes C, Goldschmidt-Clermont PJ, Vance JM, et al. Comprehensive genetic analysis of the platelet activating factor acetylhydrolase (PLA2G7) gene and cardiovascular disease in case-control and family datasets. *Hum Mol Genet.* 2008; 17(9):1318–28. [PubMed: 18204052]
- Thompson K. Programming Techniques: Regular expression search algorithm. *Communication of the ACM.* 1968; 11(6):419–422.
- Troxel WM, Matthews KA, Bromberger JT, Sutton-Tyrrell K. Chronic stress burden, Discrimination, and subclinical carotid artery disease in African American and caucasian women. *Health Psychology.* 2003; 22(3):300–309. [PubMed: 12790258]
- Ward-Caviness, CK. Gene-Environment Interactions in Cardiovascular Disease [dissertation]. Durham, NC: Duke University; 2014.
- Watkins LL, Blumenthal JA, Babyak MA, Davidson JR, McCants CB, O'Connor C, Sketch MH. Prospective association between phobic anxiety and cardiac mortality in individuals with coronary heart disease. *Psychosomatic Medicine.* 2010; 72(7):664–671. [PubMed: 20639390]
- Williams RB. Psychosocial and biobehavioral factors and their interplay in coronary heart disease. *Annual Review of Clinical Psychology.* 2008; 4:349–365.

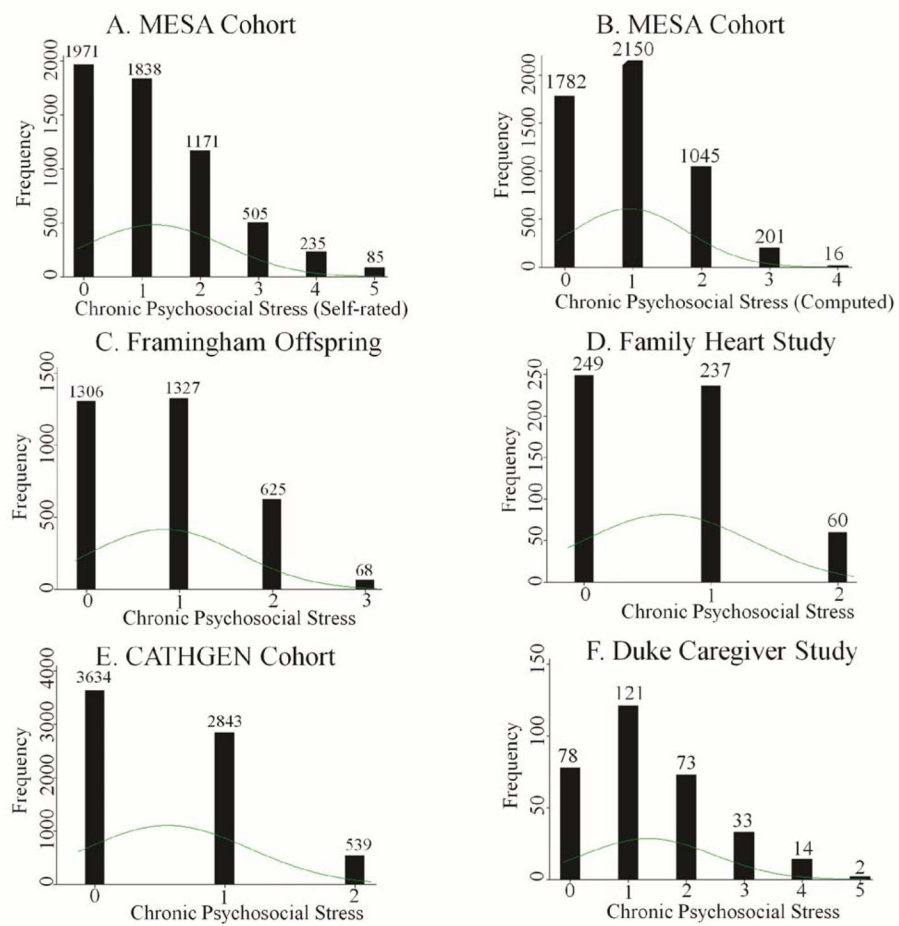


Figure 1. The histograms of chronic psychosocial stress scores in A. MESA Cohort (Self-rated), B. MESA Cohort (Computed), C. Framingham Offspring Cohort, D. Family Heart Study, E. CATHGEN Cohort, and F. Duke Caregiver Study datasets.

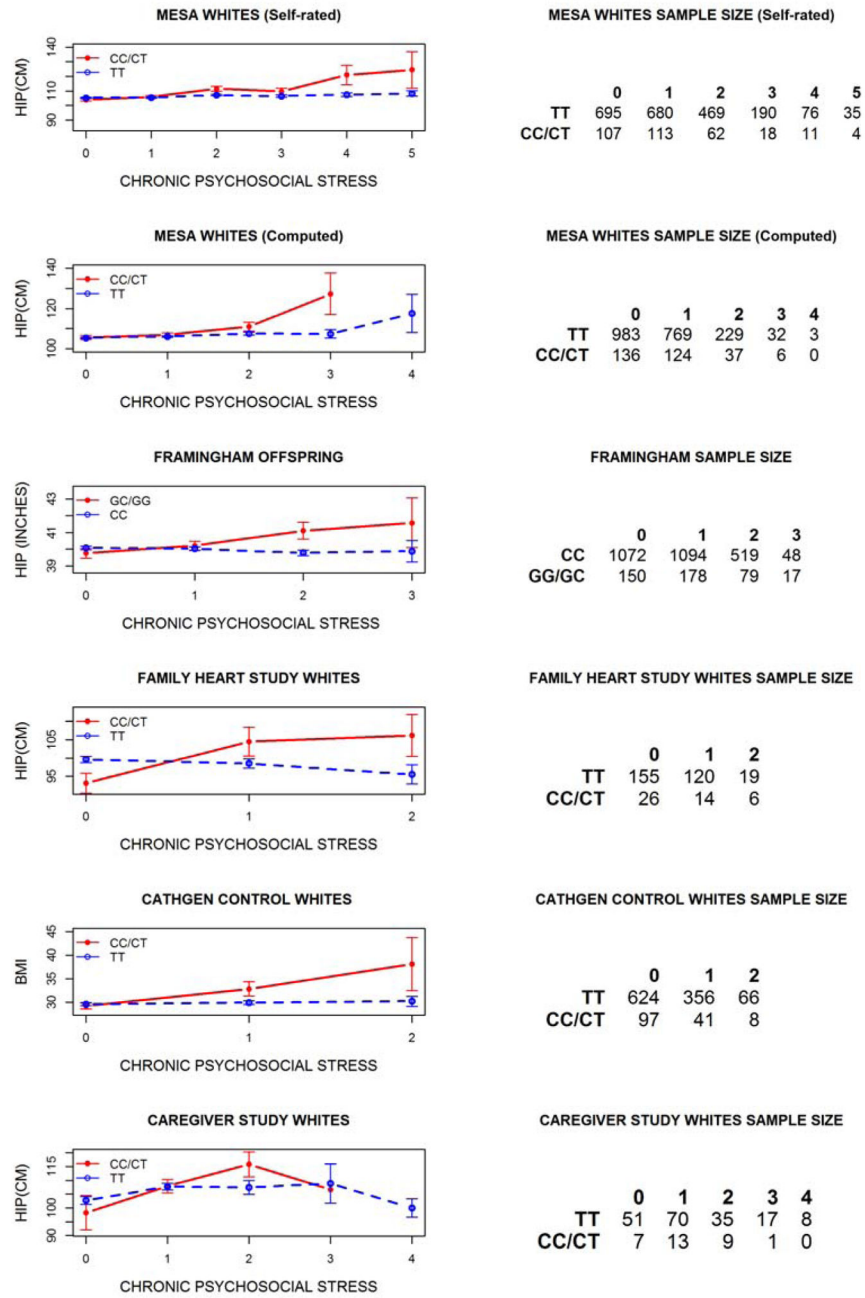


Figure 2. The mean of hip circumference or BMI vs. chronic psychosocial stress levels for two genotype groups of *EBF1* SNPs.

Table 1

A matrix of chronic psychosocial stress components construct and indicator datapoint variables.

Chronic psychosocial stress Components	Main question(s) asked	Primary datapoints	Secondary datapoints	Tertiary datapoints
Financial strain	Having ongoing financial strain (economic or money problems)?	Financial strain construct	Family income Socioeconomic status Income to poverty ratio	Health insurance Foreclosure of mortgage or loan Home value Mean home value in neighborhood
Relationship or marital problem	Having ongoing difficulties in a relationship or marriage?	Relationship or Marital difficulties construct Marital disagreement construct	Marital separation/ Divorce Hassles related to spouse	Marital status Marital reconciliation Trouble with in-laws
Work related difficulties	Having ongoing serious difficulties at your work place?	Job related difficulties construct	Job insecurity Psychological job demand Hassles related to work	Job dissatisfaction Problem with boss Problem with co-workers Problem in career prospects
Health problems of someone close	Having serious ongoing health problem of spouse or someone close to you?	Caregiving stress construct	Health status of spouse or close family members Hassles related to health of some close	Wellbeing of family members Caregiving status Death of spouse or close family members due to illness
Health problems of one's own	Having serious ongoing health problems (yourself)?	Self-rated health or well-being	Hassles related to self-health	Not applicable

Table 2

An example of search results for chronic psychosocial stress components indicators in the MESA Cohort dataset.

Chronic psychosocial stress components	Main question(s) asked	Primary datapoints	Secondary datapoints	Tertiary datapoints
Financial strain	Having ongoing financial strain (economic or money problems)?	Not Available	TOTAL GROSS FAMILY INCOME	Not Needed
Relationship or marital problem	Having ongoing difficulties in a relationship or marriage?	Not Available	MARITAL SEPARATION/ DIVORCE	Not Needed
Work related difficulties	Having ongoing serious difficulties at your work place?	Not available	JOB SECURITY, JOB REQUIRES WORKING VERY FAST, ASKED TO DO AN EXCESSIVE AMOUNT OF WORK	Not Needed
Health problems of someone close	Having serious ongoing health problem of spouse or someone close to you?	Not available	Not available	Not available
Health problems of one's own	Having serious ongoing health problems (yourself)?	GENERAL HEALTH STATUS	Not Needed	Not Needed

Table 3

Correlations between the computed and self-rated chronic psychosocial stress and its binary components in the MESA Cohort dataset. The table shows the Rho and *P*-values of polychoric correlation (tetrachoric for binary components).

Stress Variable	Rho	<i>P</i> -value
Financial Strain	0.25	<0.0001
Marital Problems	0.23	<0.0001
Work Difficulties	0.33	<0.0001
Health Problems-Self	0.45	<0.0001
Chronic Psychosocial Stress	0.23	<0.0001

Table 4

Standardized factor loadings from exploratory and confirmatory factor analysis of self-rated and computed components of chronic psychosocial stress in MESA Cohort.

Stress Variable	EFA	CFA	<i>P</i> -value
Financial Strain (SR)	0.72	0.78	< .001
Work Difficulties (SR)	0.64	0.64	< .001
Health Problems-Self (SR)	0.53	0.43	< .001
Marital Problems (SR)	0.50	0.45	< .001
Health Problems-Others (SR)	0.33	0.32	< .001
Financial Strain (C)	0.19	0.17	< .001
Work Difficulties (C)	0.24	0.24	< .001
Health Problems-Self (C)	0.37	0.36	< .001
Marital Problems (C)	0.29	0.32	< .001

SR: Self-report

C: Computed

EFA: Exploratory Factor Analysis

CFA: Confirmatory Factor Analysis

Table 5

Spearman's Correlation (Rho) between chronic psychosocial stress and depression measures.

Dataset	Depression measure	Rho	P-values
MESA Cohort (Self-rated)	CES-D	0.35	<0.0001
MESA Cohort (Computed)	CES-D	0.23	<0.0001
Framingham Offspring Cohort	CES-D	0.23	<0.0001
Family Heart Study	CES-D	0.25	<0.0001
CATHGEN Cohort ¹	Beck Inventory	0.15	0.005
Duke Caregiver Study	CES-D	0.42	<0.0001

¹Depression data available only for a small subset *i.e.* 443 participants. No observations for CATHGEN CONTROL participants.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6

Replication of gene-by-stress interaction effect on central obesity trait hip circumference (or BMI, if hip was not available): SNP *P*-values from the simple (i.e., without an interaction) linear regression model and the interaction term (GxE) *P*-values from GxE interaction linear regression model in MESA Whites and Framingham Offspring cohort datasets, Family Heart Study Whites, CATHGEN Whites, CATHGEN Control Whites, and Duke Caregiver Study for *EBF1* SNP rs4704963 (or LD SNP rs17056278, if rs4704963 was not available).

Dataset	Phenotype	SNP	Minor allele frequency (MAF)	SNP <i>P</i> -value	GxE <i>P</i> -value
MESA Whites (Self-rated) [Singh et al, 2015]	HIP (CM)	rs4704963	0.07	0.021	7.14E-09
MESA Whites (Computed)	HIP (CM)	rs4704963	0.07	0.021	0.0024
Framingham Offspring [Singh et al, 2015]	HIP (IN)	rs17056278	0.069	0.318	0.0074
Family Heart Study Whites	HIP (CM)	rs4704963	0.065	0.709	9.62E-05
CATHGEN Whites	BMI	rs4704963	0.068	0.752	0.5967
CATHGEN Control Whites	BMI	rs4704963	0.062	0.172	0.0019
Duke Caregiver Study Whites	HIP (CM)	rs4704963	0.067	0.467	0.0639