NIH Public Access
Author Manuscript
Genet Epidemiol. Author manuscript; available in PMC 2013 April 01.

Published in final edited form as:
Genet Epidemiol. 2012 April ; 36(3): 206–213. doi:10.1002/gepi.21613.

Gene–Environment Interactions on Growth Trajectories

Shuang Wang[1,*], **Wei Xiong**[1], **Weiping Ma**[2], **Stephen Chanock**[3], **Wieslaw Jedrychowski**[4,5], **Rongling Wu**[6], and **Frederica P. Perera**[5]

[1]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York [2]School of Mathematical Sciences, Fudan University, Shanghai P.R., China [3]National Cancer Institute, Bethesda, Maryland [4]Department of Epidemiology and Preventive Medicine College of Medicine, Jagiellonian University, Krakow, Poland [5]Columbia Center for Children's Environmental Health Mailman School of Public Health, Columbia University, New York, New York [6]Center for Statistical Genetics, The Pennsylvania State University, Hershey, Pennsylvania

Abstract

It has been suggested that children with larger brains tend to perform better on IQ tests or cognitive function tests. Prenatal head growth and head growth in infancy are two crucial periods for subsequent intelligence. Studies have shown that environmental exposure to air pollutants during pregnancy is associated with fetal growth reduction, developmental delay, and reduced IQ. Meanwhile, genetic polymorphisms may modify the effect of environment on head growth. However, studies on gene–environment or gene–gene interactions on growth trajectories have been quite limited partly due to the difficulty to quantitatively measure interactions on growth trajectories. Moreover, it is known that assessing the significance of gene–environment or gene–gene interactions on cross-sectional outcomes empirically using the permutation procedures may bring substantial errors in the tests. We proposed a score that quantitatively measures interactions on growth trajectories and developed an algorithm with a parametric bootstrap procedure to empirically assess the significance of the interactions on growth trajectories under the likelihood framework. We also derived a Wald statistic to test for interactions on growth trajectories and compared it to the proposed parametric bootstrap procedure. Through extensive simulation studies, we demonstrated the feasibility and power of the proposed testing procedures. We applied our method to a real dataset with head circumference measures from birth to age 7 on a cohort currently being conducted by the Columbia Center for Children's Environmental Health (CCCEH) in Krakow, Poland, and identified several significant gene–environment interactions on head circumference growth trajectories.

Keywords

gene–environment interactions; growth curves; Wald test; parametric bootstrap

INTRODUCTION

Several studies in children have suggested that children with larger brains, measured as head circumference or with MRI, tend to score higher on IQ tests or cognitive function tests [Fisch et al., 1976; Gale et al., 2004, 2006; Reiss et al., 1996; Weinberg et al., 1974]. It has

© 2012 Wiley Periodicals, Inc.

*Correspondence to: Shuang Wang, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 630, New York, NY 10032. sw2206@columbia.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

also been suggested that the extent of prenatal head growth and head growth in infancy is the most important for subsequent intelligence [Gale et al., 2006]. Impaired brain growth in utero and in infancy may lead to poorer cognitive function in childhood [Fisch et al., 1976; Gale et al., 2004, 2006; Lundgren et al., 2001; Reiss et al., 1996; Weinberg et al., 1974]. It is also known that brain volume is usually achieved at its maximal value between the ages of 5 and 10 years, but rates of brain growth are highest in the last part of gestation and the first year of life [Lemire et al., 1975]. Therefore, the factors that affect brain growth during these crucial periods might influence subsequent intelligence.

Laboratory studies exposing experimental animals to polycyclic aromatic hydrocarbons (PAH), a widespread class of combustion-related pollutants commonly found in air, food, and drinking water [International Agency for Research on Cancer, 1993], during the prenatal and neonatal periods have reported neurodevelopmental and behavioral effects, including impairment of memory and ability to learn [Brown et al., 2007; Wormley et al., 2004]. In humans, as reported previously, in the Columbia Center for Children's Environmental Health New York City (CCCEH NYC) cohort, transplacental PAH exposure has been shown to be associated with fetal growth reduction, including reduced birth weight and birth head circumference and/or small size for gestational age [Choi et al., 2006; Dejmek et al., 2000; Perera et al., 1998; Tang et al., 2006], and has been associated with developmental delay at age 3 and reduced IQ at age 5 [Perera et al., 2006, 2009]. In the CCCEH cohorts, marked inter-individual variation in response to the same level of PAH exposures was observed, indicating possible involvement of genetic components that interact with environmental factors to determine the outcome of the response.

Research has been done to identify specific genes that mediate the growth and developmental trajectories of a trait [Li et al., 2009; Wu and Hou, 2006; Wu et al., 2004; Zhao et al., 2005]. However, there has been no adequate exploration of how to map genes that modify environmental effects/genetic effects on growth trajectories. The challenge in studying gene–gene or gene–environment interactions on growth trajectory, a special type of “trait” with longitudinal measurements, lies in two aspects: how to quantitatively measure interaction effects on growth trajectories, and how to test the significance of the interactions on growth trajectories. We and other researchers have noticed that some proposed permutation tests being applied to assess the significance of gene–gene or gene–environment interactions on cross-sectional outcomes may bring substantial errors in the tests [Buzkova et al., 2011]. In this paper, we first proposed a score that quantitatively measures interactions on growth trajectories. We then proposed a parametric bootstrap procedure to assess the significance of gene–gene or gene–environment interactions on growth trajectories under the likelihood framework. We also derived a Wald statistic to test for interactions on growth trajectories and compared it to the proposed parametric bootstrap procedure. Note that in situations when a growth trajectory is modeled using a complicated nonlinear function involving multiple parameters, it may be difficult to obtain a test statistic with a closed-form expression to test for gene–gene or gene–environment interactions. In that case, the proposed parametric bootstrap procedure, although less computationally efficient, will be able to handle the more complicated growth functions with multiple parameters. In modeling head circumference measured over time, we applied a well-developed mathematical function on physical growth [Von, 1957]. We demonstrated the feasibility and power of the proposed method through extensive simulation studies and an application to a real data set with head circumference measures from birth to age 7 from the Polish cohort currently being conducted by CCCEH in Krakow, Poland.

We also want to emphasize that although the proposed method was motivated and applied to detect gene–environment interactions on head circumference growth trajectories, it is readily

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

applied to other growth trajectories and to detect gene–gene interactions on growth trajectories.

MATERIALS AND METHODS

MODEL

The phenotypic value of head circumference of individual i measured at time t with the consideration of the interaction between a gene and a dichotomous environmental exposure can be modeled as:

$$y_i(t) = \sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk} g_{jk}(t) + e_i(t), \quad (1)$$

where ξ_{ijk} is the indicator variable defined as 1 if individual i has genotypic value j (here we consider a dominant genetic model with $j = 1$ for genotypes AA and Aa, and $j = 0$ for genotype aa under the assumption of a bi-allelic marker having a high-risk allele A and a low-risk allele a, where the high-risk allele is the minor allele in the cohort) and environmental exposure level k (k 0,1, 1 as exposed or high exposure, 0 as unexposed or low=exposure); $g_{jk}(t)$ is a function of time representing the mean growth curve of head circumference of individuals having genotypic value j and environmental exposure level k; $e_i(t)$ is the individual residual at time t.

We model the head circumference growth using the following logistic function:

$$g_{jk}(t) = \frac{a_{jk}}{1 + b_{jk} e^{-c_{jk} t}}, \quad (2)$$

with parameters a_{jk}, b_{jk}, and c_{jk}, $j = 0, 1$; $k = 0, 1$. Here a_{jk} is the limiting value of g_{jk} when $t \to \infty$, which can be viewed as the head circumference of grown ups; $a_{jk}/(1 + b_{jk})$ is the initial value of g_{jk} when $t = 0$, which can be viewed as the head circumference of newborns; and c_{jk} is the relative growth rate [Von, 1957]. This logistic equation has been widely applied to model physical growth including weight, height and head circumference [Li et al., 2009; Wu and Hou, 2006; Wu et al., 2004; Zhao et al., 2005]. Note that there are four sets of parameters with each parameter set specifying one of the four growth curves, corresponding to four possible gene and environment combinations. This model setting can also be used for interactions involving the recessive genetic model, but not additive genetic model when there are three genotypic values.

The individual residuals for the same subject are assumed to be correlated over time following an autoregressive AR(1) model, that is, $e_i(t) = \rho e_i(t-1) + \varepsilon_i(t)$, where ρ is the autoregressive parameter and $\varepsilon_i(t)$ are independent errors normally distributed with mean 0 and variance σ^2.

With these assumptions, the likelihood function for n children's head circumference growth data is:

$$L(y) = \prod_{i=1}^{n}\left[\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\frac{1}{(2\pi)^{\frac{L_i}{2}}|\Sigma_i|^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2}\left(y_i - \mu_{jk}\right)'\Sigma_i^{-1}\left(y_i - \mu_{jk}\right)\right)\right], \quad (3)$$

where y_i is the data vector of head circumferences of individual i measured over time; L_i is the length of y_i; μ_{jk} is the mean vector of head circumferences for individuals with genotypic value j and environmental exposure level k computed from g_{jk}, and Σ_i is the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

covariance matrix of y_i. In practice, such as in the CCCEH Polish cohort, children were measured at ages 0, 3, 4, 5, 6, and 7. As there is possible missingness for each subject at different time points, the length of y_i, L_i, may vary for different children. So does the length of μ_{jk} and the dimension of Σ_i.

ESTIMATING PROCEDURE

We used the maximum likelihood method to estimate the unknown parameters, $n = (a_{00}, b_{00}, c_{00}, a_{01}, b_{01}, c_{01}, a_{10}, b_{10}, c_{10}, a_{11}, b_{11}, c_{11}, \sigma^2, \rho)$. It could be difficult to derive a close form for the maximizer of (3) with constraint; therefore, we used numerical analysis to obtain the maximum likelihood estimator (MLE). We applied the SOLNP algorithm [Murtagh and Saunders, 1987; Robinson, 1972; Ye, 1987] to estimate parameters under nonlinear constraints for both constrained (under the null hypothesis) and unconstrained optimization (under the alternative hypothesis).

SDA AND HYPOTHESIS TESTING

We are interested in testing the interaction effects between genes and environmental exposures on growth trajectories. We first quantitatively defined a growth curve outcome of a child using the area under the growth curve. We then introduced the concept of Second-order Difference of Areas-under-the-curve (SDA) to quantitatively capture the strength of gene–environment interactions on growth trajectories based on the definition of interactions, that is, the difference of the effects of environment on growth curves with different genotypic values:

$$SDA = \left[\int_0^T g_{00}(t)\,dt - \int_0^T g_{01}(t)\,dt\right] - \left[\int_0^T g_{10}(t)\,dt - \int_0^T g_{11}(t)\,dt\right], \quad (4)$$

where T is the endpoint of the study. Here “first-order difference” was referred to as the effect of environment on growth curves for a given genotypic value, that is, the difference between the areas under the two growth curves with two environmental conditions for a given genotypic value. The “second-order difference” was then referred to as the difference between the two “first-order differences” for the two genotypic values. In (4), the first term on the right side in the bracket represents the environmental effect with genotypic value 0, where a positive value means children with environmental condition 0 have bigger area under the growth curve than those of the children with environmental condition 1. These children with environmental condition 0 could either be growing faster, or growing bigger in the end. While the second term in the bracket represents that with genotypic value 1. The second-order difference, SDA, therefore, reflects the strength of the interaction effect on growth curves similarly as the regression coefficient of the cross-product term that defines interaction in a regression framework. When SDA = 0, that means the difference of the effects of environment on growth curves with different genotypic values is 0. That is, genes do not modify the effect of the environment on growth trajectories. The bigger the SDA value, the stronger the interaction effect on growth trajectories. Thus, we test the following hypotheses: $H_0 : SDA = 0$ vs. $H_1 : SDA \neq 0$ to test gene–environment interactions on growth trajectories.

With ($\hat{a}_{00}, \hat{b}_{00}, \hat{c}_{00}, \hat{a}_{01}, \hat{b}_{01}, \hat{c}_{01}, \hat{a}_{10}, \hat{b}_{10}, \hat{c}_{10}, \hat{a}_{11}, \hat{b}_{11}, \hat{c}_{11}$) estimated from data, we can compute $\hat{g}_{00}, \hat{g}_{01}, \hat{g}_{10}$, and $\hat{g}_{11}$, and estimate SDA. In our case, SDA is a closed-form function of the parameters, so we used the Wald test to test SDA = 0 (for detailed derivation of the Wald statistic see Appendix). However, for other types of growth functions, there may not be a closed-form expression for SDA and it will be difficult to obtain the theoretical distribution of the estimated SDA under the null hypothesis. Also, existing proposed permutation tests being applied to assessing the significance of gene–environment

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

interactions on cross-sectional outcomes may bring substantial errors in the tests [Buzkova et al., 2011]. In such situations, we proposed a parametric bootstrap procedure to assess the significance of gene–environment interactions.

Step 1: Estimate the unconstrained MLE of parameters under H_1, denoted as η_{alt}, and compute the estimated SDA, which we denote as SDA_{obs}.

Step 2: Estimate the constrained MLE of parameters under H_0, denoted as η_{null}.

Step 3: Generate $y^*_{i,b}$, vectors of data on all possible time points, based on model (1) with parameters η_{null}, $i = 1, \ldots, n$, and $b = 1, \ldots, B$. Here n is the number of children, and B is the number of bootstrap samples. For each bootstrap sample $y^*_b = \left(y^*_{1,b}, \ldots, y^*_{n,b}\right)$, compute the estimated SDA, denoted as SDA^*_b.

Step 4: Compute the P-value for testing $H_0 : SDA = 0$ vs. $H_1 : SDA \neq 0$ as the percentage of $(|SDA^*_1|, \ldots, |SDA^*_B|)$ that are greater than $|SDA_{obs}|$, where we take the absolute value of SDA.

SIMULATION STUDIES

We compared the performance of the Wald test and the proposed testing procedures to test gene–environment interactions on head circumference growth trajectories with extensive simulation studies and a real data application. All simulation studies and real data analysis were conducted using Matlab (7.8.0).

CONSTRAINED AND UNCONSTRAINED MLES

We first evaluated the performance of the SOLNP algorithm on unconstrained (under H_1) and constrained (under H_0) optimizations under nonlinear constraints. Specifically, for a given set of parameters η_0, we generated 1,000 datasets with $n = 500$ children, each having 10 head circumference measures from ages 0 to 9. The head circumference measure at each time point was generated using the model (1). The genotypic value of each child was generated from a Bernoulli distribution Bernoulli (0.4) to mimic the distribution of some of the genetic polymorphism data in the CCCEH Polish cohort. That is, for some genetic polymorphisms, about 40% of the cohort has genotypes AA or Aa, corresponding to a minor allele frequency of 0.22. The environmental exposure level of each child was generated from Bernoulli (0.6). For each simulated dataset, we applied the SOLNP algorithm to obtain the unconstrained MLE $\widehat{\eta}_{alt}$ under $H_1 : SDA \neq 0$ and the constrained MLE $\widehat{\eta}_{null}$ under $H_0 : SDA = 0$. The SOLNP algorithm requires an upper and a lower bound, as well as an initial value for each parameter. All parameters a_{jk}, b_{jk}, and c_{jk} $j = 0,1$, and $k = 0,1$ are positive. In addition, since a_{jk} is the head circumference of adults, and the largest known normal head circumference is 63.5 cm [Bushby et al., 1992], we specified the upper bounds for a_{jk} to be 65 cm. The bounds for σ^2 and ρ are $(0, \infty)$ and $(-1, 1)$. We used two types of initial values in our simulation studies and compared their estimation results. In the first type, the true parameter values were set as the initial values. In the second type, we obtained the initial values based on a crude estimation with sample means or based on the scientific meanings of the parameters. We denoted the initial values of parameters by an upper check, for example, the initial values of a_{jk}, $\check{a}_{jk}$ is the largest head circumference in each gene–environment combination inflated by a small factor of 1.01. We took the mean of head circumferences measured at age 0 in each combination to be $\check{a}_{jk}/(1 + \check{b}_{jk})$, and solved for $\check{b}_{jk}$. We then regressed $-\log\left(\left(\frac{\check{a}_{jk}}{y_i(t)} - 1\right)/\check{b}_{jk}\right)$ against t and took the slope to be $\check{c}_{jk}$. After plugging

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$\check{a}_{jk}$, $\check{b}_{jk}$, and $\check{c}_{jk}$ in $g_{jk}(t)$, we could obtain residuals $\check{e}_i(t) = y_i(t) - \check{g}_{jk}(t)$, and $\check{\sigma}^2$ and $\check{\rho}$ are the sample variance and lag 1 autocorrelation of $\check{e}_i(t)$. The optimization results using these two types of initial values are presented in Table I using one hypothetical parameter setting.

As we can see, parameter estimates using both types of initial values are very close to the true values. As expected, the scenario with the true parameter values as the initial values has slightly more accurate estimates than the scenario with the crude estimates as the initial values. Thus, the estimation procedure with the SOLNP algorithm is satisfactory.

TYPE I ERROR

We used two sets of parameters to assess the Type I error of the proposed testing procedures (Table II), where the theoretical SDA is zero under the null hypothesis of no gene–environment interactions on head circumference growth trajectories. We simulated 1,000 datasets each having $n = 500$ children and applied both the Wald statistic and the parametric bootstrap method previously described to each simulated dataset. With each simulated dataset for the parametric bootstrap procedures, we repeated B = 1,000 times. The Type I errors were estimated as the percentage of times the null hypothesis was rejected, that is, the asymptotic *P*-value using the Wald statistic and the empirical *P*-value based on the 1,000 parametric bootstraps is less than 0.05. Both types of initial value estimation were considered in the simulation studies.

Table III displays estimated Type I error rates to detect gene–environment interactions on head circumference growth trajectories as well as the mean and SD of SDA_{obs} for each parameter set. With both parameter sets, the nominal Type I error rate 0.05 was well controlled with both types of initial values examined using the parametric bootstrap procedure. However, the Type I error rate was inflated by about 2% when we used the Wald test and the crude estimates as the initial values. This might be due to the estimation bias in SDA caused by the less-accurate initial values of the crude estimates. The parametric bootstrap procedure is not sensitive to the estimation bias caused by the initial values because all bootstrap samples have this bias and the *P*-values are obtained through comparisons. The mean SDA_{obs} was estimated more accurately with the true parameter values being the initial values as expected. The SDs of SDA_{obs} are fairly stable with both types of initial values.

POWER

To examine the power of the proposed testing procedure, we considered different effect sizes of gene–environment interactions, SDA = 2.5, 3.5, 5, and 10 on head circumference growth trajectories, where we varied either a_{jk}, the limiting value of head circumference, or b_{jk} and c_{jk}, the relative growth rate. That is, some children may end up having bigger heads in adulthood than others; some children may grow faster than others. The different parameter settings considered are summarized in Table IV. Table V presents estimated power under the corresponding parameter settings considered, together with the mean and SD of SDA_{obs} using two types of initial values. As we expected, the power increases as SDA increases. When SDA reaches 5, the power is greater than 80%. It is also shown that using initial values from a crude estimation provided comparable results as using the true parameter values as the initial values. Therefore, in the subsequent real data analysis, we feel confident that it is appropriate to use the proposed crude estimation procedure for initial values. We note that the power of the Wald test is very close to that of the parametric bootstrap procedures.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

REAL DATA APPLICATION

The proposed method was applied to detect gene–environment interactions on head circumference growth trajectories from birth through age 7 with the Krakow Polish cohort currently being conducted by CCCEH. Eligible pregnant women entered the study at the beginning stage of their pregnancy. During the second or third trimesters, the women carried a backpack containing a portable personal exposure air monitor during the day and kept it near the bed at night during a consecutive 48-hr period for PAH measurements. Fifteen common genetic polymorphisms in candidate genes that play important roles in the metabolic activation of PAHs and PAH detoxification were selected (Table VI). Some of the genetic polymorphisms were previously analyzed on PAH-DNA adducts, an indicator of DNA damage [Wang et al., 2008, 2010]. The data set consists of 356 children who have at least two head circumference measurements at ages 0, 3, 4, 5, 6, and 7 and have environmental exposure information. The cohort is currently being followed up and will have more head circumference measurements as study continues. PAH measures were dichotomized at median to obtain a binary PAH exposure, defined as PAH high or PAH low similar to our previous study [Wang et al., 2008]. The number of bootstrap samples was set at 1,000, and we used the crude estimation method previously described to obtain initial values of the parameters in the optimization procedures.

Table VII presents the estimated SDAs, which are the measures of the strength of the gene–environment interactions on head circumference growth trajectories, and the corresponding *P*-values from the parametric bootstrap procedures and the Wald test on the 15 markers using the proposed method. Although the two sets of results using the Wald test and the parametric bootstrap are very close to each other, three markers, *rs162549, rs1056836*, and *GSTT1*, deletion significantly interact with the environmental exposure at the 0.05 significance level on head circumference growth trajectories using the parametric bootstrap procedures, while only the last two markers significantly interact with the environmental exposure using the Wald test. We plotted the estimated head circumference growth curves of the four gene–environment combinations together with the observed head circumference measurements for the significant *rs1056836/GSTT1* deletion and environment interactions (Figure 1). For example, the estimated effect size of *GSTT1 deletion*-PAH interaction is almost SDA of 11. More specifically, within the PAH low exposure group, children who have *GSTT1* deleted have a much slower growth rate in head circumference compared to that of children with *GSTT1* not deleted; while within the PAH high exposure group, children who have *GSTT1* deleted have a similar growth rate in head circumference compared to that of children with *GSTT1* not deleted.

DISCUSSION

It is known that most human traits are likely under the control of several genetic factors as well as environmental factors, which interact among each other to influence these traits. To study childhood growth is an important research topic. However, studies on gene–gene or gene–environment interactions on childhood growth have been quite limited. There are two major challenges in study interactions on growth trajectories: how to quantitatively define interactions on growth trajectories, and how to assess the significance of the interactions on growth trajectories.

In this study, we first proposed a score SDA to quantitatively measure interaction effects on the special type of “trait,” growth trajectories. We then proposed a parametric bootstrap procedure to empirically assess the significance of gene–gene or gene–environment interactions on growth trajectories under the likelihood framework. We applied a well-developed and biologically meaningful mathematical function to model growth curves,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

which incorporates the initial and limiting values as well as the relative growth rate. With this growth function adopted, SDA, which measures the strength of the interaction, is a closed-form function of the parameters. Therefore, we also derived the Wald test for testing the interactions. We note that for other types of growth functions when there may not be a closed-form expression for SDA, it will be difficult to obtain the theoretical distribution of the estimated SDA under the null hypothesis. In such cases, the proposed parametric bootstrap will be a natural choice. The proposed method is demonstrated to be feasible and robust; and provides accurate estimates as well as good powers through extensive simulations. A real data application on head circumference measured through birth to age 7 from a birth cohort study currently being conducted in Krakow, Poland by CCCEH identified three markers, *rs162549, rs1056836*, and *GSTT1*, deletion to significantly modify the effects of the environmental exposure on head circumference growth trajectories. We note that *GSTT1* deletion is an important gene deletion and has found to modify the risk of colorectal cancer [Csejtei et al., 2008]. As the Polish cohort is being followed up to an older age, more head circumference measurements will be collected. With which we expect to have more accurate parameter estimates and a better power.

Although we focused on detecting gene–environment interactions on head circumference growth trajectories, the proposed method is readily applied to other growth curves, such as weight and height growth trajectories. In the current study, we assumed the genotyped markers are putative quantitative trait loci (QTL). The proposed method to assess significance of interactions could be readily extended by considering flanking markers with the well-developed QTL methods.

Acknowledgments

This work was supported by the National Institute of Environmental Health Sciences (P01 ES09600, R01 ES08977, P50ES015905). The authors declare no conflict of interest regarding this work.

APPENDIX: DERIVATION OF THE WALD TEST STATISTIC

Let $\theta = (a_{00}, b_{00}, c_{00}, a_{01}, b_{01}, c_{01}, a_{10}, b_{10}, c_{10}, a_{11}, b_{11}, c_{11})$. Then for the specific growth function (2), SDA has a closed-form as a function of θ:

$$SDA = R(\theta) = (A_{00} - A_{01}) - (A_{10} - A_{11}),$$

where

$$A_{jk} = \int_0^T g_{jk}(t)\,dt = \frac{a_{jk}}{c_{jk}} \ln \frac{\exp\left(c_{jk}T\right) + b_{jk}}{1 + b_{jk}}.$$

We will show later that $\widehat{\theta} = \left(\widehat{a}_{00}, \widehat{b}_{00}, \widehat{c}_{00}, \widehat{a}_{01}, \widehat{b}_{01}, \widehat{c}_{01}, \widehat{a}_{10}, \widehat{b}_{10}, \widehat{c}_{10}, \widehat{a}_{11}, \widehat{b}_{11}, \widehat{c}_{11}\right)$ and $(\widehat{\sigma}, \widehat{\rho})$ are asymptotically independent. Therefore, the Wald statistic for testing

$$H_0: SDA = 0 \quad \text{vs.} \quad H_1: SDA \neq 0 \quad \text{is}$$

$$W_n = \frac{R\left(\widehat{\theta}\right)^2}{\left\{\left[C\left(\widehat{\theta}\right)\right]' \left[I_n\left(\widehat{\theta}\right)\right]^{-1} \left[C\left(\widehat{\theta}\right)\right]\right\}}.$$

Here $I_n(\theta)$ is the information matrix corresponding to θ, and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$C(\theta)=\frac{\partial R(\theta)}{\partial\theta}=\left[S_{00}(\theta)',-S_{01}(\theta)',-S_{10}(\theta)',S_{11}(\theta)'\right]',$$

where

$$S_{jk}(\theta)=\left[\frac{\partial A_{jk}}{\partial a_{jk}},\frac{\partial A_{jk}}{\partial b_{jk}},\frac{\partial A_{jk}}{\partial c_{jk}}\right]'.$$

Let $r_{ijk}=y_i-\mu_{jk}=(r_{ijk}(0)),\ldots,r_{ijk}(L_{i-1}))$, then the log likelihood

$$l(\eta)=\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[-\log\left((2\pi)^{\frac{L_i}{2}}|\Sigma_i|^{\frac{1}{2}}\right)-\frac{1}{2}r_{ijk}'\Sigma_i^{-1}r_{ijk}\right].$$

The information matrix for η is $I_n(\eta)=-\mathrm{E}\left[\frac{\partial^2 l(\eta)}{\partial\eta^2}\right]$. First, we have

$$\frac{\partial l(\eta)}{\partial a_{jk}}=\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[-r_{ijk}'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial a_{jk}}\right].$$

Note that $\frac{\partial r_{ijk}}{\partial a_{jk}}=-\frac{\partial\mu_{ijk}}{\partial a_{jk}}$ is a function of the parameters, and (σ, ρ) only exists in Σ_i^{-1}, so $\frac{\partial^2 l(\eta)}{\partial a_{jk}\partial\sigma}$ and $\frac{\partial^2 l(\eta)}{\partial a_{jk}\partial\rho}$ are both linear functions of $(r_{ijk}(0),\ldots,r_{ijk}(L_{i-1}))$. Since $\mathrm{E}[r_{ijk}]=0$, we have

$$-\mathrm{E}\left[\frac{\partial^2 l(\eta)}{\partial a_{jk}\partial\sigma}\right]=-\mathrm{E}\left[\frac{\partial^2 l(\eta)}{\partial a_{jk}\partial\rho}\right]=0.$$

Therefore, $\hat{a}_{jk}$ are asymptotically independent of $\widehat{\sigma}$ and $\widehat{\rho}$. It is also true for $\hat{b}_{jk}$ and $\hat{c}_{jk}$. Also note that

$$\frac{\partial^2 l(\eta)}{\partial a_{jk}^2}=-\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[r_{ijk}'\Sigma_i^{-1}\frac{\partial^2 r_{ijk}}{\partial a_{jk}^2}+\left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial a_{jk}}\right];$$

and

$$\frac{\partial^2 l(\eta)}{\partial a_{jk}b_{jk}}=-\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[r_{ijk}'\Sigma_i^{-1}\frac{\partial^2 r_{ijk}}{\partial a_{jk}b_{jk}}+\left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial b_{jk}}\right];$$

so

$$-\mathrm{E}\left[\frac{\partial^2 l(\eta)}{\partial a_{jk}^2}\right]=\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[\left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial a_{jk}}\right];$$

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

and

$$-\mathrm{E}\left[\frac{\partial^2 l(\eta)}{\partial a_{jk} b_{jk}}\right]=\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\left[\left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial b_{jk}}\right].$$

It is also true for other second derivatives with respect to the mean parameters.

Therefore

$$I_n(\theta)=\begin{bmatrix} D_{00}(\theta) & 0 & 0 & 0 \\ 0 & D_{01}(\theta) & 0 & 0 \\ 0 & 0 & D_{10}(\theta) & 0 \\ 0 & 0 & 0 & D_{11}(\theta) \end{bmatrix},$$

where

$$D_{jk}(\theta)=\sum_{i=1}^{n}\sum_{j=0}^{1}\sum_{k=0}^{1}\xi_{ijk}\times\begin{bmatrix} \left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial a_{jk}} & \left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial b_{jk}} & \left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial c_{jk}} \\ \left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial b_{jk}} & \left(\frac{\partial r_{ijk}}{\partial b_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial b_{jk}} & \left(\frac{\partial r_{ijk}}{\partial b_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial c_{jk}} \\ \left(\frac{\partial r_{ijk}}{\partial a_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial c_{jk}} & \left(\frac{\partial r_{ijk}}{\partial b_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial c_{jk}} & \left(\frac{\partial r_{ijk}}{\partial c_{jk}}\right)'\Sigma_i^{-1}\frac{\partial r_{ijk}}{\partial c_{jk}} \end{bmatrix}$$

and

$$W_n=R(\theta)^2/\left\{\sum_{j=0}^{1}\sum_{k=0}^{1}S_{jk}(\theta)'D_{jk}(\theta)^{-1}S_{jk}(\theta)\right\}.$$

REFERENCES

Brown LA, Khousbouei H, Goodwin JS, Irvin-wilson CV, Ramesh A, Sheng L, Mccallister MM, Jiang GC, Aschner M, Hood DB. Down-regulation of early ionotrophic glutamate receptor subunit developmental expression as a mechanism for observed plasticity deficits following gestational exposure to benzo(a)pyrene. Neurotoxicology. 2007; 28:965–978. [PubMed: 17606297]

Bushby KMD, Cole T, Matthews JNS, Goodship JA. Centiles for adult head circumference. Arch Dis Child. 1992; 67:1286–1287. [PubMed: 1444530]

Buzkova P, Lumley T, Rice K. Permutation and parametric bootstrap tests for gene–gene and gene–environment interactions. Ann Hum Genet. 2011; 75:36–45. [PubMed: 20384625]

Choi H, Jedrychowski W, Spengler J, Camann DE, Whyatt RM, Rauh V, Tsai WY, Pererea FP. International studies of prenatal exposure to polycyclic aromatic hydrocarbons and fetal growth. Environ Health Persp. 2006; 114:1744–1750.

Csejtei A, Tibold A, Varega Z, Koltai K, Ember A, Orsos Z, Feher G, Horvath OP, Ember I, Kiss I. *GSTM*, *GSTT* and *p53* polymorphisms as modifiers of clinical outcome in colorectal cancer. Anticancer Res. 2008; 28:1917–1922. [PubMed: 18630481]

Dejmek J, Solansky I, Benes I, Lenicek J, Sram RJ. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. Environ Health Persp. 2000; 108:1159–1164.

Fisch RO, Bilek MK, Horrobin JM, Chang PN. Children with superior intelligence at 7 years of age. Am J Dis Child. 1976; 130:481–487. [PubMed: 1274898]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Gale CR, O'callaghan FJ, Godfrey KM, Law CM, Martyn CN. Critical periods of brain growth and cognitive function in children. Brain. 2004; 127:321–329. [PubMed: 14645144]

Gale CR, O'callaghan FJ, Bredow M, Martyn CM, The AVON Longitudinal Study of Parents and Children Study Team. The influence of head growth in fetal life, infancy, and childhood on intelligence at the ages of 4 and 8 years. Pediatrics. 2006; 118:1486–1492. [PubMed: 17015539]

International Agency for Research on Cancer. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. International Agency for Research on Cancer; Lyon, France: 1993. Polynuclear Aromatic Compounds. Part l. Chemical, Environmental, and Experimental Data.; p. l-453.

Lemire, RJ.; Loeser, J.; Leech, R.; Alvord, E. Normal and Abnormal Development of the Human Nervous System. Harper & Row; Maryland: 1975.

Li N, Das K, Wu R. Functional mapping of human growth trajectories. J Theor Biol. 2009; 261:33–42. [PubMed: 19632241]

Lundgren EM, Cnattingius S, Jonsson B, Tuvemo T. Intellectual and psychological performance in males born small for gestational age with and without catch-up growth. Pediatr Res. 2001; 50:91–96. [PubMed: 11420424]

Murtagh, BA.; Saunders, MA. MINOS 5.1 user's guide. Department of OR, Stanford University; 1987. Technical Report SOL 83–20R

Perera FP, Whyatt RM, Jedrychowski W, Rauh V, Manchester D, Santella RM, Ottman R. Recent developments in molecular epidemiology: a study of the effects of environmental polycyclic aromatic hydrocarbons on birth outcomes in Poland. Am J Epidemiol. 1998; 47:309–314. [PubMed: 9482506]

Perera FP, Rauh V, Whytt RM, Tsai WY, Tang D, Diza D, Tu YH, Camann D, Kinney P. Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. Environ Health Persp. 2006; 114:1287–1292.

Perera FP, Li Z, Whytt R, Hoepner L, Wang S, Camann D, Rauh V. Prenatal polycyclic aromatic hydrocarbon exposure and child intelligence at age 5. Pediatrics. 2009; 124:e195–e202. [PubMed: 19620194]

Reiss AL, Abrams MT, Singer HS, Ross JL, Denckla MB. Brain development, gender and IQ in children. Brain. 1996; 119:1763–1774. [PubMed: 8931596]

Robinson SM. A quadratically convergent algorithm for general nonlinear programming problems. Math Program. 1972; 3:145–156.

Tang D, Li TY, Liu JJ, Chen YH, Qu L, Perera FP. PAH-DNA adducts in cord blood and fetal and child development in a Chinese cohort. Environ Health Persp. 2006; 114:1297–1300.

Von Bertalanffy L. Quantitative laws in metabolism and growth. Q Rev Biol. 1957; 32:217–231. [PubMed: 13485376]

Wang S, Chanock S, Tang D, Li ZG, Jedrychowski W, Perera FP. An assessment of interactions between PAH exposure and genetic polymorphisms on PAH-DNA adducts in African American, Dominican, and Caucasian Mothers and Newborns. Cancer Epidem Biomar Prev. 2008; 17:405–413.

Wang S, Zheng T, Chanock S, Jedrychowski W, Perera FP. Methods for detecting interactions between genetic polymorphisms and prenatal environment exposure with a mother-child design. Genet Epidemiol. 2010; 34:125–132. [PubMed: 19582785]

Weinberg WA, Dietz SG, Penick EC, Mcalister WH. Intelligence, reading achievement, physical size and social class: a study of St Louis Caucasian boys aged 8.0 to 9.6 years, attending regular schools. J Pediatr. 1974; 85:482–489. [PubMed: 4443855]

Wormley DD, Chirwa S, Nayyar T, Wu J, Johnson S, Brown LA, Harris E, Hood DB. Inhaled benzo(a)pyrene impairs long-term potentiation in the F1 generation rat dentate gyrus. Cell Mol Biol. 2004; 50:715–721. [PubMed: 15641162]

Wu R, Hou W. A hyperspace model to decipher the genetic architecture of developmental processes: allometry meets ontogeny. Genetics. 2006; 172:7–637. [PubMed: 16204212]

Wu R, Ma C, Lin M, Wang Z, Casells G. Functional mapping of quantitative trait loci underlying growth trajectories using a transform-both-sides logistic model. Biometrics. 2004; 60:729–738. [PubMed: 15339296]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Ye, Y. Ph.D. Thesis. Department of EES, Stanford University; 1987. Interior algorithms for linear, quadratic, and linearly constrained nonlinear programming..

Zhao W, Hou W, Little RC, Wu R. Structured antedependence models for functional mapping of multiple longitudinal traits. Stat Appl Genet Mol Biol. 2005; 4 article 33.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

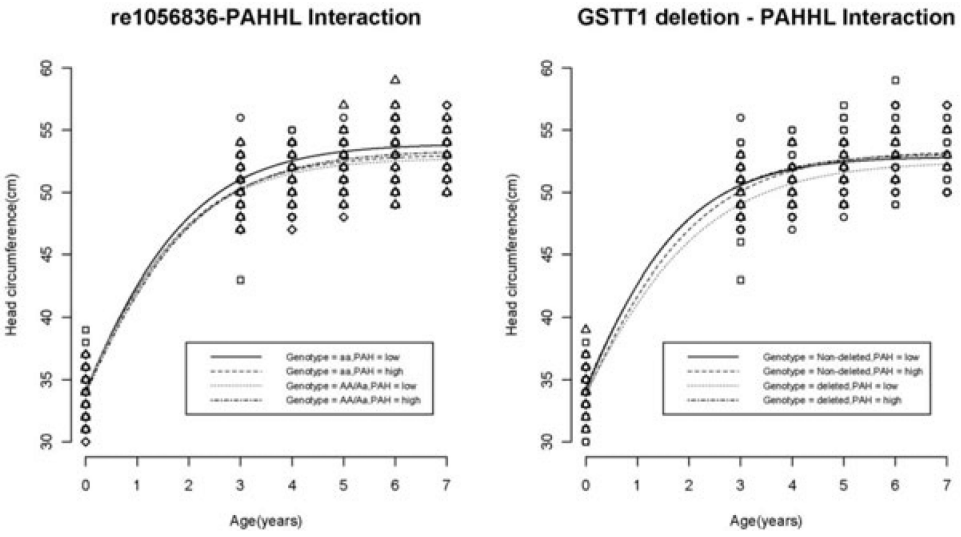

Fig. 1.

Significant gene–environment interactions between markers *rs1056836/GSTT1* deletion and environmental exposure on head circumference growth trajectories with the Polish cohort using the proposed method.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

TABLE I

Parameter estimates*

	True parameter values	Unconstrained optimization		Constrained optimization	
Type of initial values		Truth	Crude	Truth	Crude
a_{00}	51.5	51.50 (0.17)	51.48 (0.17)	51.52 (0.13)	51.39 (0.20)
b_{00}	0.51	0.51 (0.01)	0.51 (0.01)	0.51 (0.01)	0.50 (0.01)
c_{00}	0.82	0.82 (0.02)	0.82 (0.02)	0.82 (0.01)	0.85 (0.03)
a_{01}	53.0	52.99 (0.14)	53.00 (0.14)	52.98 (0.11)	53.12 (0.16)
b_{01}	0.55	0.55 (0.01)	0.55 (0.01)	0.55 (0.01)	0.56 (0.01)
c_{01}	0.84	0.84 (0.02)	0.84 (0.02)	0.84 (0.01)	0.81 (0.02)
a_{10}	52.5	52.51 (0.20)	52.49 (0.21)	52.48 (0.15)	52.45 (0.24)
b_{10}	0.53	0.53 (0.01)	0.53 (0.01)	0.53 (0.01)	0.53 (0.02)
c_{10}	0.85	0.85 (0.02)	0.85 (0.03)	0.85 (0.02)	0.87 (0.05)
a_{11}	54.5	54.50 (0.17)	54.47 (0.17)	54.52 (0.13)	54.33 (0.19)
b_{11}	0.58	0.58 (0.01)	0.58 (0.01)	0.58 (0.01)	0.57 (0.01)
c_{11}	0.75	0.75 (0.02)	0.75 (0.02)	0.75 (0.01)	0.78 (0.03)
σ^2	1.50	1.49 (0.03)	1.48 (0.03)	1.50 (0.03)	1.48 (0.04)
ρ	0.67	0.67 (0.01)	0.66 (0.02)	0.67 (0.01)	0.66 (0.02)

*Mean (SD), based on 1,000 simulations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE II

Parameter settings under the null hypothesis with no gene–environment interactions on head circumference growth trajectories for Type I error analysis

	Parameter set 1	Parameter set 2
a_{00}	52.94	50.00
b_{00}	0.55	0.48
c_{00}	0.84	0.84
a_{01}	52.6	50.58
b_{01}	0.54	0.5
c_{01}	0.82	0.83
a_{10}	53.5	50.44
b_{10}	0.57	0.49
c_{10}	0.75	0.82
a_{11}	52.5	50.95
b_{11}	0.53	0.48
c_{11}	0.85	0.79
σ^2	1.5	1.5
ρ	0 67	0.67

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

TABLE III

Type I error results

Type of initial values	True parameter	Type I error from bootstrap	Type I error from Wald test	Mean (SDA_{obs})	Std (SDA_{obs})
Truth	Set l	0.042	0.049	–0.03	1.54
	Set 2	0.046	0.054	–0.02	1.54
Crude	Set 1	0.050	0.072	–0.15	1.65
	Set 2	0.056	0.068	–0.20	1.61

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE IV

Parameter settings for power analysis

Interaction effect size	SDA = 2.5	SDA = 3.5	SDA = 5	SDA = 10
Part I: Parameter settings by varying a_{jk}				
a_{00}	52.00	52.00	51.50	51.50
b_{00}	0.51	0.51	0.51	0.51
c_{00}	0.82	0.82	0.82	0.82
a_{01}	53.13	53.01	52.33	51.79
b_{01}	0.55	0.55	0.55	0.56
c_{01}	0.84	0.84	0.84	0.84
a_{10}	52.5	52.5	52.5	52.5
b_{10}	0.53	0.53	0.53	0.53
c_{10}	0.85	0.85	0.85	0.85
a_{11}	54.5	54.5	54.5	54.5
b_{11}	0.58	0.58	0.58	0.58
c_{11}	0.75	0.75	0.75	0.75
σ^2	1.5	1.5	1.5	1.5
ρ	0.67	0.67	0.67	0.67
Part II: Parameter settings by varying b_{jk} and c_{jk}				
a_{00}	52.01	51.98	52.00	52.01
b_{00}	0.53	0.53	0.53	0.53
c_{00}	0.8	0.8	0.8	0.8
a_{01}	53	53	53	53
b_{01}	0.55	0.55	0.55	0.55
c_{01}	0.78	0.78	0.78	0.76
a_{10}	52.5	52.5	52.5	52.5
b_{10}	0.53	0.53	0.53	0.53
c_{10}	0.85	0.87	0.9	0.77
a_{11}	53.5	53.5	53.5	53.5
b_{11}	0.57	0.57	0.57	0.57
c_{11}	0.93	1	1.1	1.1
σ^2	1.5	1.5	1.5	1.5
ρ	0.67	0.67	0.67	0.67

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

TABLE V

Power analysis results

Type of initial values	True SDA	Power from bootstrap	Power from Wald test	Mean (SDA_{obs})	Std (SDA_{obs})
Part I: Power results by varying a_{jk}					
Truth	2.5	0.326	0.342	2.48	1.54
	3.5	0.579	0.595	3.48	1.54
	5	0.873	0.896	4.97	1.54
	10	1.0	1.0	9.97	1.55
Crude	2.5	0.291	0.331	2.34	1.63
	3.5	0.541	0.583	3.35	1.61
	5	0.843	0.871	4.89	1.65
	10	1.0	1.0	9.95	1.62
Part II: Power results by varying b_{jk} and c_{jk}					
Truth	2.5	0.321	0.339	2.47	1.55
	3.5	0.590	0.610	3.47	1.54
	5	0.886	0.904	4.98	1.53
	10	1.0	1.0	9.98	1.53
Crude	2.5	0.338	0.355	2.40	1.61
	3.5	0.601	0.598	3.38	1.62
	5	0.861	0.860	4.81	1.68
	10	1.0	1.0	9.93	1.68

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE VI

Chromosomal positions and gene locations of the 15 markers from the selected candidate genes

Gene	SNP rs Num	Alleles	Chr.	Position (bp)
CYP1A1	rs2198843	C/G	15	72,788,283
	rs1456432	A/G	15	72,790,104
	rs4646421	T/C	15	72,803,245
	rs2606345	T/G	15	72,804,229
	rs7495708	G/A	15	72,806,896
	rs2472299	C/T	15	72,820,453
CYP1B1	rs162549	T/A	2	38,148,960
	rs1056837	T/C	2	38,151,654
	rs1056836	G/C	2	38,151,707
	rs162560	A/G	2	38,153,019
	rs10012	C/G	2	38,155,894
	rs2617266	C/T	2	38,156,048
GSTT2	rs2719	T/G	22	22,629,757
	rs140194	G/A	22	22,655,095
GSTT1	NA	Gene deletion	22	22,706,128

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE VII

Estimated SDAs and *P*-values from 1,000 parametric bootstraps at 15 candidate markers

Gene	SNP rs Num	|SDA|*	*P*-values from the parametric bootstrap	*P*-values from the Wald test
CYP1A1	rs2198843	0.58	0.760	0.842
	rs1456432	1.44	0.584	0.532
	rs4646421	5.08	0.061	0.087
	rs2606345	0.19	0.925	0.947
	rs7495708	0.75	0.841	0.830
	rs2472299	3.94	0.110	0.107
CYP1B1	rs162549	5.08	*0.032***	0.113
	rs1056837	4.67	0.148	0.065
	rs1056836	5.57	*0.047***	*0.022***
	rs162560	1.62	0.568	0.533
	rs10012	4.16	0.073	0.194
	rs2617266	3.73	0.068	0.058
GSTT2	rs2719	1.99	0.410	0.589
	rs140194	3.22	0.175	0.165
GSTT1	NA	10.91	*0.002***	*0.006***

*Absolute value.

**Significant at $P < 0.05$.