Supplementary Section I

Modeling Framework and Likelihood Estimation

Consider the 3-way table of G, E and D:

Frequency	G	E	D
$\mathbf{n_1}$	0	0	0
$\mathbf{n_2}$	1	0	0
$\mathbf{n_3}$	0	1	0
$\mathbf{n_4}$	1	1	0
$\mathbf{n_5}$	0	0	1
$\mathbf{n_6}$	1	0	1
$\mathbf{n_7}$	0	1	1
$\mathbf{n_8}$	1	1	1

Denoting the vector of all cell counts as $\mathbf{n} = [n_1 \dots n_8]^T$, we assume a Poisson distribution for $\mathbf{n}$ given model $\mathcal{M}_i$

$$\mathbf{n}|\boldsymbol{\mu}, \mathcal{M}_i \sim \text{Poisson}(\boldsymbol{\mu})$$

and use a natural log link to model the Poisson parameter given model $\mathcal{M}_i$ and design matrix $\mathbf{X}_i$,

$$\log(\boldsymbol{\mu}|\mathcal{M}_i) = \mathbf{X}_i\beta_i$$

$$\boldsymbol{\mu}|\mathcal{M}_i = e^{\mathbf{X}_i\beta_i}$$

$$\beta_i|\mathcal{M}_i \sim N(\mathbf{0}, \sigma^2\mathbf{V}_i)$$

The marginal likelihood of model $\mathcal{M}_i$ is

$$\Pr(\mathbf{n}|\mathcal{M}_i) = \int \Pr(\mathbf{n}|\boldsymbol{\theta}_i, \mathcal{M}_i) \Pr(\boldsymbol{\theta}_i|\mathcal{M}_i)\, d\boldsymbol{\theta}_i, \quad \boldsymbol{\theta}_i = [\boldsymbol{\beta}_i, \sigma^2]^T$$

where a closed-form solution for $\Pr(\mathbf{n}|\mathcal{M}_i)$ is not analytically attainable due to the lack of conjugacy between the Gaussian prior and the Poisson likelihood. Hence, we use the GLIB (A.E.

Raftery & Richardson, 1996) routine within the BMA R package which utilizes Laplace estimation to estimate this likelihood. We implement GLIB with prior covariance matrix

$$\boldsymbol{V} = \sigma^2 \begin{bmatrix} \phi^2 \left(\frac{1}{n}\boldsymbol{X}_1^T \boldsymbol{X}_1\right)^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \phi^2 \left(\frac{1}{n}\boldsymbol{X}_p^T \boldsymbol{X}_p\right)^{-1} \end{bmatrix}.$$

where the hyperparameter ϕ refers to a user-specified hyperparameter which is used in the prior variance calculation of effect estimates (Adrian E. Raftery, Madigan, & Hoeting, 1997), with larger values of ϕ resulting in a preference for simpler models (A. E. Raftery, Madigan, D.M. and Hoeting, J., 1993). In our simulations and analyses $\phi = 1$ is chosen based on suggestions by Raftery (1993).

Supplementary Section II

Simulation Specifications

We simulated an underlying population using the following sampling distributions and logistic regression equations in the following order:

$$\mathrm{E} \sim \mathrm{Bernoulli}(\mathrm{p_E}) \quad (1)$$

$$\mathrm{logit}(\Pr(\mathrm{G} = 1|\mathrm{E})) = \mathrm{logit}(\mathrm{q_A}) + \alpha_{\mathrm{cc_{ge}}}(\mathrm{E} - \overline{\mathrm{E}}) \quad (2)$$

$$\mathrm{logit}(\Pr(\mathrm{Y} = 1|\mathrm{E}, \mathrm{G})) = \mathrm{logit}(\mathrm{p}_Y) + \beta_{\mathrm{cc_E}}(\mathrm{E} - \overline{\mathrm{E}}) + \beta_{\mathrm{cc_G}}(\mathrm{G} - \overline{\mathrm{G}}) + \beta_{\mathrm{cc_{G\times E}}}(\mathrm{E} - \overline{\mathrm{E}})(\mathrm{G} - \overline{\mathrm{G}}) \quad (3)$$

From this population, we sampled equal numbers of cases and controls for all simulation scenarios for both single-marker and genome-wide simulations. When fitting the 1 and 2-degree-of-freedom

BMA models, we used the GLIB function in the BMA R package based on the Laplace approximation to the marginal likelihood (Adrian E Raftery, 1996). Prior means for all model parameters were set to $\mathbf{0} = [\,0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\,]^T$. Prior model weights were set according to prior specified CC:CO odds for the models with 1:1 odds $\Rightarrow \Pr(\mathcal{M}_{cc}) = \Pr(\mathcal{M}_{co}) = 0.5$ and 100:1 odds $\Rightarrow \Pr(\mathcal{M}_{cc}) = 0.990099$ and $\Pr(\mathcal{M}_{co}) = 0.00990099$.

Figure 3 Specifications

Figure 3 was created using a simulation of 1,000 replicates of a sample with size N = 10,000 made up of 500 cases and 500 controls. We simulated 999,999 independent SNPs, and one designated 'causal' SNP with a non-zero interaction effect. Part (A) depicts a simulation without the marginal effects of E and G $\left(\beta_{cc_E} = \beta_{cc_G} = \text{Log}(1.0)\right)$ for the designated SNP. Part (B) depicts a simulation with constant marginal effects for all values of $\beta_{cc_{G\times E}}$, $\beta_{cc_E} = \beta_{cc_G} = \text{Log}(1.2)$, and part C is based on marginal effects induced through the increasing interaction effect based on values produced by Quanto (http://biostats.usc.edu/Quanto.html).

Figure 4 Specifications

ROC curves shown in Figure 4 (A-C) were produced using a simulation of 1,000 replicates of a sample sized N = 10,000 with 500 cases and 500 controls. We simulated 9980 independent SNPs part (A) and 9480 (parts B and C), and 20 designated 'causal' SNPs. Parts A and B depict effect sizes of $\beta_{cc_{G\times E}} = \text{Log}(1.3)$ and $\beta_{cc_G} = \text{Log}(1.2)$ and $\beta_{cc_E} = \text{Log}(1.0)$, with $\alpha_{cc_{GE}} = \text{Log}(1.0)$ and $\alpha_{cc_{GE}} = \text{Log}(1.2)$ respectively for (A) and (B). Part (C) depicts effect sizes of $\beta_{cc_{G\times E}} = \text{Log}(1.0), \beta_{cc_G} = \text{Log}(1.0)$, $\beta_{cc_E} = \text{Log}(1.0)$, and $\alpha_{cc_{GE}} = \text{Log}(1.0)$.

Supplementary Section III

Asthma Application Models

We conducted the $G \times PM_{2.5}$ analysis by specifying the following case-control log-linear equation:

$$\mathrm{Log}(n|G, E, Y, C_k) = \alpha_{cc_0} + \alpha_{cc_G}G + \alpha_{cc_E}E + \alpha_{GE}GE + \beta_{cc_0}Y + \beta_{cc_G}GY + \beta_{cc_E}EY + \beta_{cc_{G\times E}}GEY + \sum_{k}^{4} \alpha_{cc_{C_k}}C_k + Y\sum_{k}^{4} \beta_{cc_{C_k}}C_k$$

where

$C_1 =$ Sex (1: male, 0: female)

$C_2 =$ Native American Ancestry (1: 5% - 50%, 0: otherwise)

$C_3 =$ Native American Ancestry (1: >50%, 0: otherwise)

$C_4 =$ Hispanic White (1: Hispanic White, 0: Non-Hispanic White).

Likewise, the $G \times$ Hispanicity analysis used the following case-control log-linear equation:

$$\mathrm{Log}(n|G, E, Y, C_k) = \alpha_{cc_0} + \alpha_{cc_G}G + \alpha_{cc_E}E + \alpha_{GE}GE + \beta_{cc_0}Y + \beta_{cc_G}GY + \beta_{cc_E}EY + \beta_{cc_{G\times E}}GEY + \sum_{k}^{3} \alpha_{cc_{C_k}}C_k + Y\sum_{k}^{3} \beta_{cc_{C_k}}C_k$$

with the omission of C_4 as Hispanicity is captured here by E.

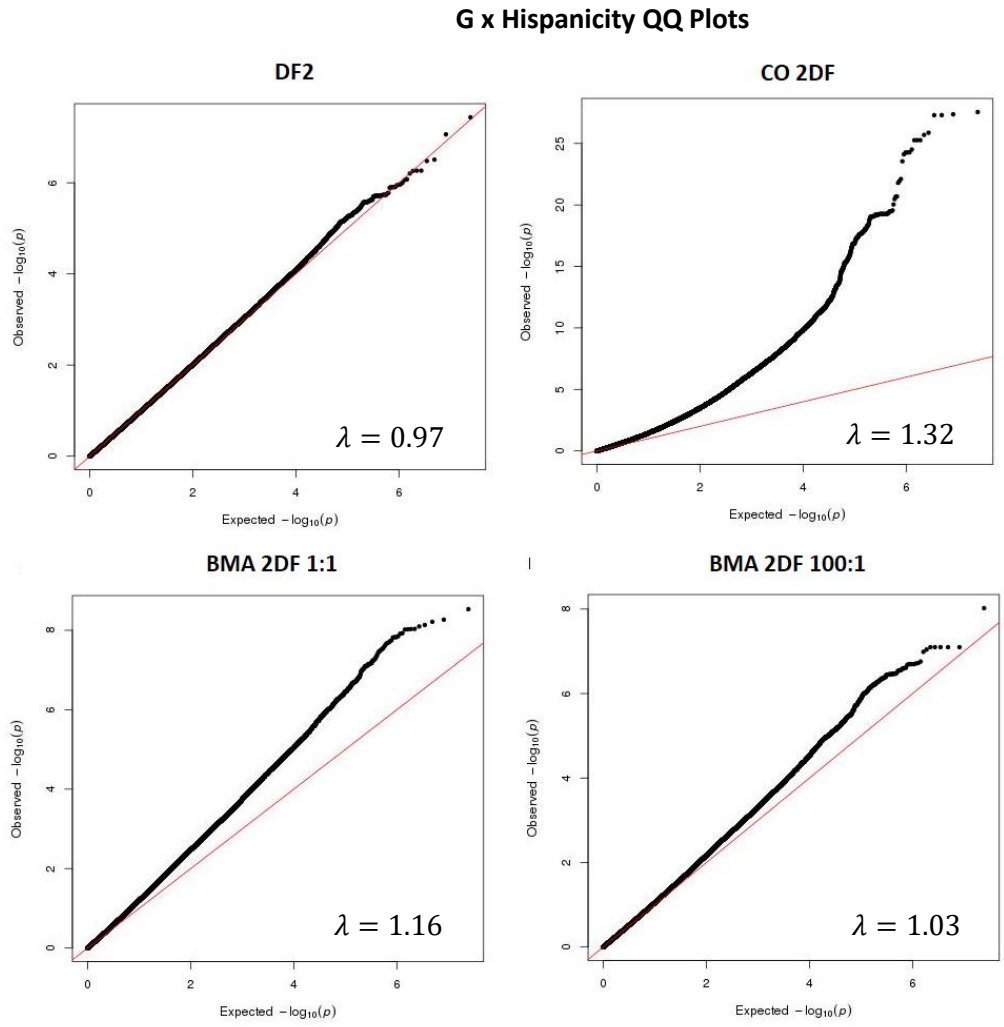

Figure 1 QQ-plots from asthma analysis in the CHS for Hispanicity x G analysis by method (clockwise) DF2, CO 2DF, and BMA 2DF. BMA 2DF analysis was conducted using a CC:CO prior odds of 1:1 and 100:1 favoring the CC model over the CO model. Markers which result in zero contingency table cells have been removed.

Supplementary Section IV

Software

Simulations and analysis used the GLIB function in the BMA R-package (Adrian Raftery, 2015; A.E. Raftery & Richardson, 1996). Software used to carry out the BMA 2DF approach specifically is available for download as the "bma.gxe" R-package through GitHub at https://github.com/LilithMoss/bma.gxe.git.

References

Adrian Raftery, J. H., Chris Volinsky, Ian Painter and Ka Yee Yeung. (2015). BMA: Bayesian Model Averaging (Version 3.18.6). Retrieved from http://CRAN.R-project.org/package=BMA

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models *Biometrika, 83*(2), 251-266.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian Model Averaging for Linear Regression Models *Journal of the American Statistical Association, 92*(437), 179-191.

Raftery, A. E., Madigan, D.M. and Hoeting, J. (1993). Model selection and accounting for model uncertainty in linear regression models (U. o. W. Department of Statistics, Trans.) *Technical Report*.

Raftery, A. E., & Richardson, S. (1996). Model selection for generalized linear models via GLIB, with application to epidemiology. In D. A. B. a. D. K. Stangl (Ed.), *Bayesian Biostatistics* (pp. 321-354). New York: Marcel Dekker.