

# Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension

Xiaofeng Zhu,<sup>1,\*</sup> Tao Feng,<sup>1,2</sup> Bamidele O. Tayo,<sup>3</sup> Jingjing Liang,<sup>1</sup> J. Hunter Young,<sup>4</sup> Nora Franceschini,<sup>5</sup> Jennifer A. Smith,<sup>6</sup> Lisa R. Yanek,<sup>4</sup> Yan V. Sun,<sup>7</sup> Todd L. Edwards,<sup>8</sup> Wei Chen,<sup>9</sup> Mike Nalls,<sup>10</sup> Ervin Fox,<sup>11</sup> Michele Sale,<sup>12</sup> Erwin Bottinger,<sup>13</sup> Charles Rotimi,<sup>14</sup> The COGENT BP Consortium, Yongmei Liu,<sup>15</sup> Barbara McKnight,<sup>16</sup> Kiang Liu,<sup>17</sup> Donna K. Arnett,<sup>18</sup> Aravinda Chakravati,<sup>19</sup> Richard S. Cooper,<sup>3</sup> and Susan Redline<sup>20</sup>

Genome-wide association studies (GWASs) have identified many genetic variants underlying complex traits. Many detected genetic loci harbor variants that associate with multiple—even distinct—traits. Most current analysis approaches focus on single traits, even though the final results from multiple traits are evaluated together. Such approaches miss the opportunity to systemically integrate the phenotype-wide data available for genetic association analysis. In this study, we propose a general approach that can integrate association evidence from summary statistics of multiple traits, either correlated, independent, continuous, or binary traits, which might come from the same or different studies. We allow for trait heterogeneity effects. Population structure and cryptic relatedness can also be controlled. Our simulations suggest that the proposed method has improved statistical power over single-trait analysis in most of the cases we studied. We applied our method to the Continental Origins and Genetic Epidemiology Network (COGENT) African ancestry samples for three blood pressure traits and identified four loci (*CHIC2*, *HOXA-EVX1*, *IGFBP1/IGFBP3*, and *CDH17*;  $p < 5.0 \times 10^{-8}$ ) associated with hypertension-related traits that were missed by a single-trait analysis in the original report. Six additional loci with suggestive association evidence ( $p < 5.0 \times 10^{-7}$ ) were also observed, including *CACNA1D* and *WNT3*. Our study strongly suggests that analyzing multiple phenotypes can improve statistical power and that such analysis can be executed with the summary statistics from GWASs. Our method also provides a way to study a cross phenotype (CP) association by using summary statistics from GWASs of multiple phenotypes.

## Introduction

Genome-wide association studies (GWASs) have been a major design to discover the genetic determinants of complex traits, with thousands of common genetic variants thus uncovered. The identification of genetic variants reveals important biological insights into the genetic architecture of complex traits such as hypertension (MIM 145500), human height (MIM 606255), and blood lipids.<sup>1–3</sup> It is well known that the effect sizes of identified common variants are often small and that a large sample size is necessary to ensure sufficient power to detect such variants. A common strategy is to perform a meta-analysis, combining the effect and variance estimates from as many independent studies as possible with the same or similar

phenotypes, which does not require one to pool the individual-level data, thereby reducing the logistical and regulatory burden associated with transferring individual data across centers,<sup>4</sup> as well as avoiding the need to explicitly model study design differences or manage confounding by genotyping batch effects and trait ascertainment that differentially affect allele frequency estimates across studies. Notably, GWASs are generally performed with single traits (at a univariate level), although multiple related phenotypes are often collected and studied and are expected to reflect common physiological processes. For example, a hypertension study often measures systolic blood pressure (SBP), diastolic blood pressure (DBP), and hypertension status (HTN), and these phenotypes are analyzed separately.<sup>1,5</sup> It has been suggested

<sup>1</sup>Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA; <sup>2</sup>College of Mathematical Science, Heilongjiang University, Harbin 150080, P.R. China; <sup>3</sup>Department of Public Health Science, Loyola University Chicago Stritch School of Medicine, Maywood, IL 60153, USA; <sup>4</sup>Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; <sup>5</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA; <sup>6</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; <sup>7</sup>Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA; <sup>8</sup>Center for Human Genetics Research, Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN 37212, USA; <sup>9</sup>Tulane Center for Cardiovascular Health, Tulane University, New Orleans, LA 70112, USA; <sup>10</sup>Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD 20892, USA; <sup>11</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39126, USA; <sup>12</sup>University of Virginia Center for Public Health Genomics, Charlottesville, VA 22908, USA; <sup>13</sup>The Charles Bronfman Institute for Personalized Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA; <sup>14</sup>Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, MD 20892, USA; <sup>15</sup>Department of Epidemiology & Prevention, Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; <sup>16</sup>Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; <sup>17</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA; <sup>18</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA; <sup>19</sup>Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; <sup>20</sup>Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA

\*Correspondence: [xiaofeng.zhu@case.edu](mailto:xiaofeng.zhu@case.edu)

<http://dx.doi.org/10.1016/j.ajhg.2014.11.011>. ©2015 by The American Society of Human Genetics. All rights reserved.

that univariate analysis could suffer low statistical power compared with multivariate analysis.<sup>4</sup> Systematic simultaneous analysis of multiple traits could improve the quality of inferences from analysis of outcomes that all relate to the biological construct of interest.

The joint analysis of multiple phenotypes within a cohort has recently become popular for improving statistical power to detect genetic linkage and association. Solovieff et al. provided a detailed summary of such kinds of approaches.<sup>4</sup> Most multivariate methods are based on a multivariate regression framework and require both genotypes and phenotypes at the individual level, with an assumption of approximately normally distributed phenotypes. Extensions to allow for nonnormally distributed phenotypes and categorical phenotypes have also been developed based on generalized estimating equations (GEEs), ordinal regression, and a Bayesian framework.<sup>6–10</sup> On the other hand, approaches have been developed based on a dimension reduction technique on the phenotypes, such as principal-components (PCs) analysis and canonical correlation analysis.<sup>11,12</sup> However, it has been reported that testing only the top PCs has lower power than the combined-PC approach.<sup>13</sup>

In contrast to multivariate analysis, a method for integrating the results from standard univariate analyses across various phenotypes in GWASs has recently been developed. The Fisher's combined p value method can be applied to independent studies but is not straightforward when aggregating p values of different but correlated phenotypes within the same cohort, which will result in inflated type I error. The cross-phenotype meta-analysis (CPMA) was developed for testing whether there is association of a SNP to multiple phenotypes, rather than directly evaluating the aggregated association evidence between a SNP and multiple phenotypes.<sup>14</sup> Although the method can be applied to multiple independent studies, it does not allow overlapping or correlated samples among studies. Standard fixed and random effects meta-analysis methods are also used to combine association across multiple phenotypes but are not well suited to situations where a genetic variant has opposite effects on different phenotypes. As an example, psoriasis (MIM 309480) and Crohn disease (CD) have been reported to be positively correlated, with the prevalence of CD in individuals with psoriasis significantly higher than in controls.<sup>15</sup> However, the G allele of SNP rs12720356 in *TYK2* (MIM 176941) increases risk for Crohn disease and decreases risk for psoriasis.<sup>16,17</sup> For HTN-related traits, SBP rises linearly with age, in the absence of treatment, whereas DBP has an inverted "U" pattern with a zenith around age 50. Although SBP and DBP are positively correlated, a genetic variant might have opposite effect for the two traits. An extension of fixed effects meta-analysis is the subset-based meta-analysis,<sup>18</sup> which allows opposite effects and is able to test association to a subset of phenotypes. This method exhaustively searches all possible phenotype subsets and identifies the subset of

traits with the strongest association, but with the cost of exponentially increased multiple tests. In addition, the method does not allow for heterogeneity across cohorts for the same phenotype. Several methods have also been developed based on a linear combination of the univariate test statistics.<sup>19,20</sup> These methods have been further used to test for association between correlated traits and genetic markers.<sup>21,22</sup> However, the authors focus on only a single study with multiple traits measured in the same individuals. In addition, individual-level genotype and phenotype data are also required for the method by Yang et al.<sup>22</sup> The trait-based association test that uses an extended simes (TATES) procedure combines p values obtained in univariate analysis of traits measured in the same individuals while correcting for correlations among phenotypes.<sup>23</sup> This approach can be challenging when combining association evidence across multiple independent studies because the phenotype correlation matrix can change from cohort to cohort. Another approach, the pleiotropy regional identification method (PRIME),<sup>24</sup> evaluates pleiotropic loci in a genomic region with multiple phenotypes based on results of GWASs. This method calculates a pleiotropic index defined by the number of traits with low association p values in a genomic region. The flipping sign test uses p values obtained from individual trait analysis to combine association from multiple correlated traits but requires computationally intensive simulations to obtain combined p values at the genome-wide significance level.<sup>5</sup>

In this study, we propose a general approach that can integrate association evidence from multiple correlated continuous and binary traits from one or multiple studies. We allow for heterogeneity of effects for the same trait in different studies that might result from different populations, environmental exposures, or designs. We also allow heterogeneity effects for different phenotypes, which is not unexpected in practice. In addition, population structure and cryptic relatedness can be controlled. For cryptic relatedness, we also allow for overlapping or related subjects between the different cohorts studied. Although the proposed method is not specifically designed for identifying subsets of associated traits, we will offer insight into how to detect such subsets of traits.

## Material and Methods

Assume we have summary statistical results of GWASs from  $J$  cohorts with  $K$  phenotypic traits. In each cohort, single SNP-trait association was analyzed for each trait separately. Let  $T_{ijk}$  be a summary statistic for the  $i^{\text{th}}$  SNP,  $j^{\text{th}}$  cohort, and  $k^{\text{th}}$  trait. Let  $T_i = (T_{i1}, \dots, T_{i1}, \dots, T_{i1K}, \dots, T_{iJK})^T$ . For simplification, we omit the SNP index. Thus,  $T = (T_1, \dots, T_1, \dots, T_{1K}, \dots, T_{JK})^T$  represents a vector of test statistics for testing the association of a SNP with  $K$  traits. Let  $\beta = (\beta_{11}, \dots, \beta_{11}, \dots, \beta_{1K}, \dots, \beta_{JK})^T$  be the effect sizes of the SNP. The null hypothesis is  $H_0: \beta = \mathbf{0}$  and the alternative hypothesis  $H_1$  is that at least one of the elements of  $\beta$  is not equal to zero. We use a Wald test statistic  $T_{jk} = (\hat{\beta}_{jk} / \hat{s}_{jk})$ , where  $\hat{\beta}_{jk}$  and  $\hat{s}_{jk}$  are the

estimated coefficient and corresponding standard error for the  $k^{\text{th}}$  trait in the  $j^{\text{th}}$  cohort. It is reasonable to assume that  $T$  follows a multivariate normal distribution with mean  $\mathbf{0}$  and correlation matrix  $R$  under the null hypothesis. In practice,  $R$  needs to be estimated and we will address that later. A standard method to test  $\beta = \mathbf{0}$  is the test statistic  $S_{JK} = T^T R^{-1} T$ , which has asymptotically a  $\chi^2$  distribution with  $J \times K$  degrees of freedom. This test is omnibus with respect to the alternative hypothesis. When heterogeneous effects exist, in particular if a variant contributes to only a subset of traits, this test is less powerful because of the large number of degrees of freedom. When the effect is homogeneous, i.e., the effect sizes are all the same regardless of traits or cohorts, the most powerful test statistic is

$$S_{Hom} = \frac{e^T (RW)^{-1} T \left( e^T (RW)^{-1} T \right)^T}{e^T (WRW)^{-1} e}, \quad (\text{Equation 1})$$

which follows a  $\chi^2$  distribution with one degree of freedom, where  $e^T = (1, \dots, 1)$  has length  $J \times K$  and  $W$  is a diagonal matrix of weights for the individual test statistics.<sup>19,20</sup> When the statistics in  $T$  are mutually independent and  $W$  is diagonal with inverses of variances as elements,  $S_{Hom}$  is equivalent to an inverse variance weighted meta-analysis. A similar method to  $S_{Hom}$  has previously been proposed but it is not specifically for combining the effects across multiple traits.<sup>24</sup> Equivalently, we can take the sample sizes for the weights, i.e.,  $w_{jk} = \sqrt{n_j}$  for the sample size  $n_j$  of the  $j^{\text{th}}$  cohort. Here we assigned more weight to a large study because a large study carries more information than a small study. The advantage of using sample size over inverse variance is that then we do not need to worry about the traits being on different scales, because a variance is dependent on the scale of measurement. In this study we take  $w_{jk} = \sqrt{n_j}$ .

The homogeneous effect size assumption is unlikely to be true, especially when multiple phenotypes are analyzed together. We introduce a test statistic based on  $S_{Hom}$ , which is more powerful than  $S_{JK}$  for homogenous data, because it has one degree of freedom. Ideally, we would like to have a test statistic that includes only the cohorts and traits with a true contribution to the association of a genetic variant under the alternative hypothesis  $H_1$ . The truncated statistic methods for combining statistical evidence have been suggested for such an analysis.<sup>25,26</sup> We adopt a similar idea here. For a given  $\tau > 0$ , we let  $T(\tau)$  be the subvector of  $T$  satisfying  $|T_{jk}| > \tau$ . That is, only the statistics in the vector  $T$  with an absolute value larger than  $\tau$  will be kept. Similarly, we let  $R(\tau)$  be a submatrix of  $R$  representing the correlation matrix and  $W(\tau)$  to be the diagonal submatrix of  $W$ , corresponding to  $T(\tau)$ . To further allow for different effect directions of a variant for different traits in different cohorts, we let  $w_{jk} = \sqrt{n_j} \times \text{sign}(T_{jk})$ . Thus, the signed weights will lead to adding evidence to the association, whether or not the effects of a variant are in the same or different directions. Define

$$S(\tau) = \frac{e^T (R(\tau)W(\tau))^{-1} T(\tau) \left( e^T (R(\tau)W(\tau))^{-1} T(\tau) \right)^T}{e^T W(\tau)^{-1} R(\tau)^{-1} W(\tau)^{-1} e}. \quad (\text{Equation 2})$$

When  $\tau$  is large,  $S(\tau)$  can be undefined if  $|T_{jk}| \leq \tau$  for all  $j$  and  $k$ . In this case we define  $S(\tau) = 0$ . Our test statistic is then

$$S_{Het} = \max_{\tau > 0} S(\tau). \quad (\text{Equation 3})$$

The asymptotic distribution of  $S_{Het}$  does not follow a standard distribution but can be evaluated by simulation (see the section Evaluating the Distribution of  $S_{Het}$  under the Null Hypothesis).

We can rank the statistics  $|T_{jk}|$  for all  $j$  and  $k$  and evaluate  $S(\tau)$  at these values to obtain  $\max_{\tau > 0} S(\tau)$ . Because  $S_{Het}$  gives more weight to the large trait-cohort specific statistics, it can maintain statistical power when heterogeneity exists.

## Estimation of the Correlation Matrix $R$ among Test Statistics

We assume that the traits are quantitative but we can apply the same argument for qualitative traits. Let  $Y_1$  and  $Y_2$  be two trait vectors and  $G_1$  and  $G_2$  be two genotype vectors. Without loss of generality, we assume no covariates. If  $Y_1$  and  $Y_2$  come from the same study cohort,  $G_1$  and  $G_2$  are the same. We assume that linear regressions were applied to evaluate the association evidence between  $Y_1$  and  $G_1$  and between  $Y_2$  and  $G_2$ . That is, we have  $Y_1 = G_1 \beta_1 + \varepsilon_1$  and  $Y_2 = G_2 \beta_2 + \varepsilon_2$ , where  $\varepsilon_1 \sim N(0, I\sigma_1^2)$  and  $\varepsilon_2 \sim N(0, I\sigma_2^2)$ . Assume least-squares estimates were obtained and let  $T_1$  and  $T_2$  be the corresponding Wald test statistics for testing  $\beta_1 = 0$  and  $\beta_2 = 0$ , respectively. Then  $T_1 = G_1^T Y_1 / \sqrt{G_1^T G_1 \sigma_1^2}$  and  $T_2 = G_2^T Y_2 / \sqrt{G_2^T G_2 \sigma_2^2}$ . The correlation between  $T_1$  and  $T_2$  is

$$\text{corr}(T_1, T_2) = \frac{G_1^T \text{cov}(Y_1, Y_2) G_2}{\sqrt{G_1^T G_1 \sigma_1^2} \sqrt{G_2^T G_2 \sigma_2^2}}. \quad (\text{Equation 4})$$

If  $Y_1$  and  $Y_2$  are two traits from the same cohort, then we have  $G_1 = G_2$  and  $\text{cov}(Y_1, Y_2) = r_{12} I \sigma_1 \sigma_2$ , where  $r_{12}$  is the correlation between trait 1 and trait 2. We then have

$$\text{corr}(T_1, T_2) = r_{12}. \quad (\text{Equation 5})$$

If  $Y_1$  and  $Y_2$  are from different cohorts, then a correlation can be induced only by either overlapping samples or related subjects in the two cohorts. In either case, Equation 5 can capture the correlation. We note that Equation 5 does not depend on individual genotypes but only on the correlation between the two traits. Thus,  $\text{corr}(T_1, T_2)$  can be estimated from the summary statistics for all the independent SNPs in a genome-wide association study. That is,

$$\text{corr}(T_1, T_2) = \frac{\sum_i (T_{i1} - \mu_1)(T_{i2} - \mu_2)}{\sqrt{\sum_i (T_{i1} - \mu_1)^2} \sqrt{\sum_i (T_{i2} - \mu_2)^2}}, \quad (\text{Equation 6})$$

where  $T_1, T_2$  are the test statistics for the SNP for traits 1 and 2 in their corresponding cohorts, and  $\mu_1$  and  $\mu_2$  are their corresponding means. We suggest using independent SNPs across the genome in Equation 6 rather than all the SNPs regardless of the LD among them, which will lead to inflated correlation estimation.<sup>27</sup>

## Evaluating the Distribution of $S_{Het}$ under the Null Hypothesis

Although  $S_{Hom}$  follows a  $\chi^2$  distribution with one degree of freedom,  $S_{Het}$  does not have a closed form under the null hypothesis because of the data-adaptive approach we are using. However,  $S_{Het}$  can be viewed as the maximum of weighted sum of trait-specific test statistics satisfying different thresholds, which is closely related to a gamma distribution.<sup>28</sup> We assume under the null hypothesis that  $S_{Het}$  follows a gamma distribution with a mean shift and we use simulations to estimate this gamma distribution. Under the null, a correlation between two test statistics  $T_1, T_2$  is determined by the trait correlation in Equation 2 and is independent of the SNP. Instead of generating genotype and phenotype data, we can directly simulate the test statistic  $T$  from a multivariate normal

**Table 1. Comparison of Trait Correlations Used for Simulation and Estimated by Test Statistics from SNPs of GWASs**

	SBP	DBP	HTN
SBP	–	0.76	0.73
DBP	0.69	–	0.70
HTN	0.66	0.60	–

Values above diagonal are generating correlation coefficients, and under diagonal are estimated based on the test statistics of 100,000 SNPs.

distribution  $N(0, \hat{R})$ , where  $\hat{R}$  is the estimated correlation matrix given by Equation 6. We use the following steps to evaluate the distribution under the null hypothesis:

- (1) Estimate the correlation matrix  $\hat{R}$  by Equation 6 with the observed test statistics across the common genome-wide SNPs for all cohorts and traits.
- (2) Generate  $L$  random vectors from the multivariate normal distribution  $N(0, \hat{R})$ .
- (3) For the  $l^{\text{th}}$  realization in step 2, calculate the test statistic  $S_{Het,l}$  as in Equation 3. The weight matrix in Equation 2 is calculated with the cohort sizes.
- (4) Fit  $S_{Het,1}, S_{Het,2}, \dots, S_{Het,L}$  to a gamma distribution  $gamma(\alpha, \beta) + c$ , where  $\alpha, \beta$  are the shape and scale parameters and  $c$  is a constant, and we estimate parameters  $\alpha, \beta$ , and  $c$  by matching the first three moments to the data. The estimated parameters are represented by  $\hat{\alpha}, \hat{\beta}$ , and  $\hat{c}$ , respectively.
- (5) For any observed statistic  $S_{Het,o}$ , the p value is calculated by  $p = Prob(S_{Het,o} - \hat{c} > S_0)$ , where  $S_0$  is the fractile corresponding a significance level  $\alpha_0$  of the distribution  $gamma(\hat{\alpha}, \hat{\beta})$ .

The above procedure is not dependent on the particular SNP and this null distribution  $gamma(\hat{\alpha}, \hat{\beta})$  can be used for testing any SNP association, which drastically reduces the computation time. Notably, we can also directly use simulated  $S_{Het,1}, S_{Het,2}, \dots, S_{Het,L}$  as the distribution under the null hypothesis. However, the computation is still intensive for GWASs because of the need to evaluate test statistics at a significance level of  $5 \times 10^{-8}$ . We therefore suggest the use of the estimated gamma distribution.

### Simulations

We conducted simulations to evaluate the type I error and power of the proposed method. Five cohorts with 3,000 individuals each and three blood pressure traits were generated. We allowed a portion of the samples to overlap among the cohorts, which simulated cryptic relatedness between samples from different cohorts. The simulated data mimicked the blood pressure data from the COGENT BP consortium,<sup>5</sup> which included SBP, DBP, and HTN. The correlations between SBP and DBP was obtained from the CARE data. We first simulated a genotype  $g_i$  for the  $i^{\text{th}}$  individual with a minor allele frequency being sampled from a uniform distribution between 0.01 and 0.5, with the assumption of Hardy-Weinberg equilibrium.

To simulate the phenotype data, we first generated latent phenotypes by using a linear additive model

$$Y'_i = \mu + \beta g_i + \varepsilon_i, \quad (\text{Equation 7})$$

where  $Y'_i, \mu, \beta$ , and  $\varepsilon_i$  are column vectors of length 2, representing traits, intercepts, trait-specific effect sizes of a genotype, and

random errors. We used parameters estimated from actual African American data. In our simulation, we simulated  $Y'_i$  consisting of SBP and DBP, with population means 127 and 78, respectively, and  $\varepsilon_i \sim MVN\left(0, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$ , where we let  $\sigma_1^2 = 412$ ,  $\sigma_2^2 = 132$  and  $\rho$  is the correlation between SBP and DBP, which we varied. To simulate HTN, we assumed that individuals with either  $Y'_{i1} > 140$  or  $Y'_{i2} > 90$  have antihypertensive medications. The effects of antihypertensive medications for SBP and DBP followed the normal distributions  $N(10, \sigma_1^2/4)$  and  $N(5, \sigma_2^2/4)$ , respectively. These effects of medications mimic the current GWASs for which ten and five units are often added to SBP and DBP when an individual is on antihypertensive medications.<sup>5,29</sup> The observed phenotype values for SBP and DBP are

$$Y_i = Y'_i - \varepsilon'_i \text{ when the } i^{\text{th}} \text{ individual is on medication,} \quad (\text{Equation 8})$$

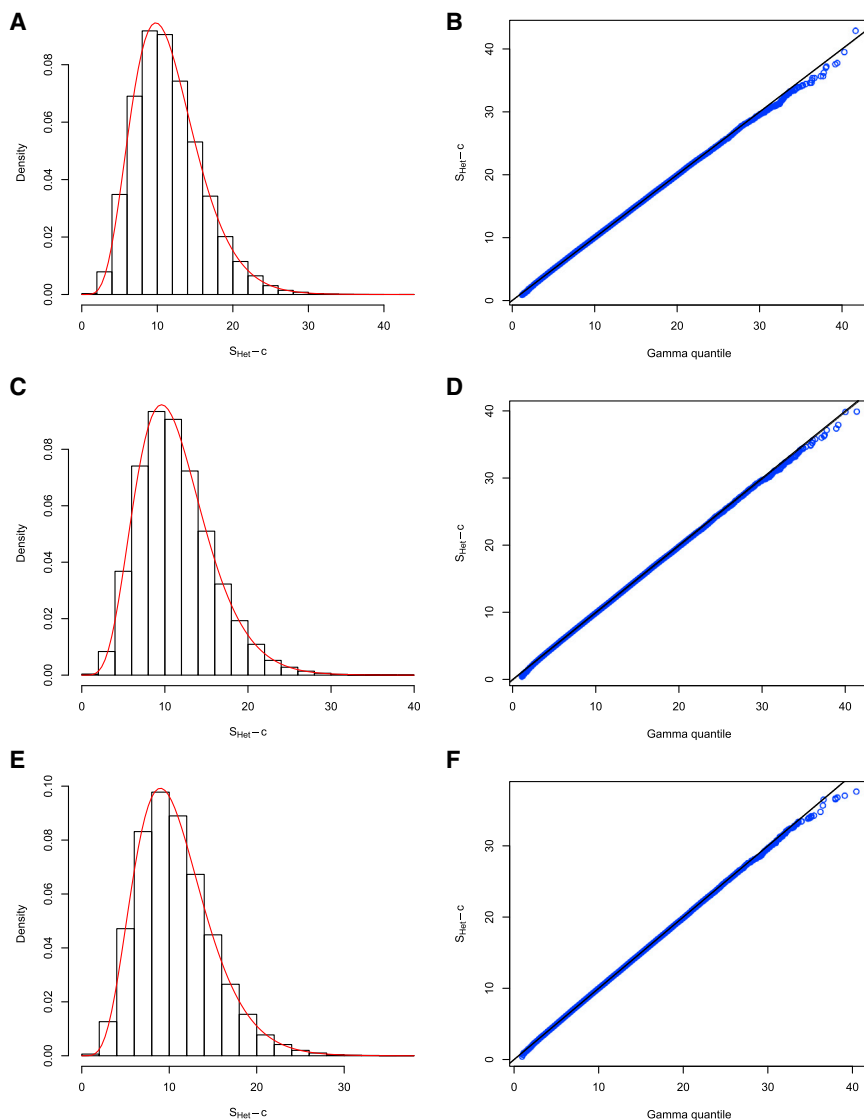
$$Y_i = Y'_i \text{ when the } i^{\text{th}} \text{ individual is not on any medication,}$$

where  $\varepsilon'_i \sim MVN\left(\begin{bmatrix} 10 \\ 5 \end{bmatrix}, \begin{bmatrix} \sigma_1^2/4 & 0 \\ 0 & \sigma_2^2/4 \end{bmatrix}\right)$ . The hypertensive case is defined as either observed SBP > 140 (equivalently  $Y_{i1} > 140$ ), observed DBP > 90 (equivalently  $Y_{i2} > 90$ ), or on antihypertensive medication. Otherwise, a control is defined. The above simulation process was used to generate the data under both the null and alternative hypotheses for five cohorts. For data under the null hypothesis, we simply let  $\beta = 0$  in Equation 7. For data under the alternative hypothesis, we set the phenotypic variance explained by a variant as 0.3% of total variance and calculated the  $\beta$  value according to the simulated causal variant allele frequency assuming an additive model of inheritance. In the case of heterogeneity across cohorts, we set  $\beta$  with the calculated value in one cohort and 0 in the rest of the cohorts.

### Results

We first examined whether the correlation among phenotypes can be well estimated by the corresponding genome-wide test statistics, as represented in Equations 5 and 6. We simulated three correlated traits (SBP, DBP, and HTN) via the method above but with genetic effects  $\beta = 0$  and  $\rho = 0.7$  for 3,000 subjects. We also simulated 100,000 independent SNPs for each individual with minor allele frequencies being sampled from a uniform distribution between 0.01 and 0.5, with the assumption of Hardy-Weinberg equilibrium. We performed association tests for all the SNPs with the software Plink.<sup>30</sup> Table 1 shows the estimated correlation among the three traits using trait values and the estimated correlation matrix based on the test statistics of 100,000 SNPs. We observed that the trait correlations can be approximated through the calculation of correlations of test statistics of SNPs corresponding to traits (Table 1).

We next examined the type I error of statistics  $S_{Hom}$  and  $S_{Het}$  under three scenarios: (1)  $\rho = 0$ , (2)  $\rho = 0.25$ , and (3)  $\rho = 0.5$ . For each scenario, we simulated 5 independent cohorts each with 3,000 subjects, 100,000 SNPs, and three traits (SBP, DBP, and HTN) under the null hypothesis of



**Figure 1.  $S_{Het}$  Distribution**

Distribution of the test statistic  $S_{Het}$  under three scenarios: trait correlation is 0 (A and B), trait correlation is 0.25 (C and D), and trait correlation is 0.5 (E and F). We generated 5 cohorts, each with sample size 3,000, with no overlapping samples between cohorts. Left panel is the histogram of  $S_{Het}$  based on 100,000 replications and the red curve represents the theoretical distribution  $gamma(\alpha, \beta)$ , where  $\alpha, \beta$  are the shape and scale parameters that were estimated by matching the first two moments. Right panel is a QQ plot of  $S_{Het}$ .

statistics. We calculated p values of  $S_{Hom}$  statistics by  $\chi_1^2$  and  $S_{Het}$  statistics by using the shift gamma distribution with previous estimated corresponding parameters, respectively. We then calculated the type I error rates by evaluating the proportion of the p values less than a significance level. Table 2 presents the type I error rates at different significance levels. We observed that the type I error rates were well controlled for  $S_{Hom}$ . For  $S_{Het}$ , type I error rates were slightly inflated when significance levels were less than  $10^{-5}$ . Thus, the proposed  $S_{Het}$  test statistics can be fitted reasonably well by an estimated shift gamma distribution under the null hypothesis. Because the distribution of  $S_{Het}$  under the null hypothesis depends only on the correlation matrix among traits and number of cohorts and does not

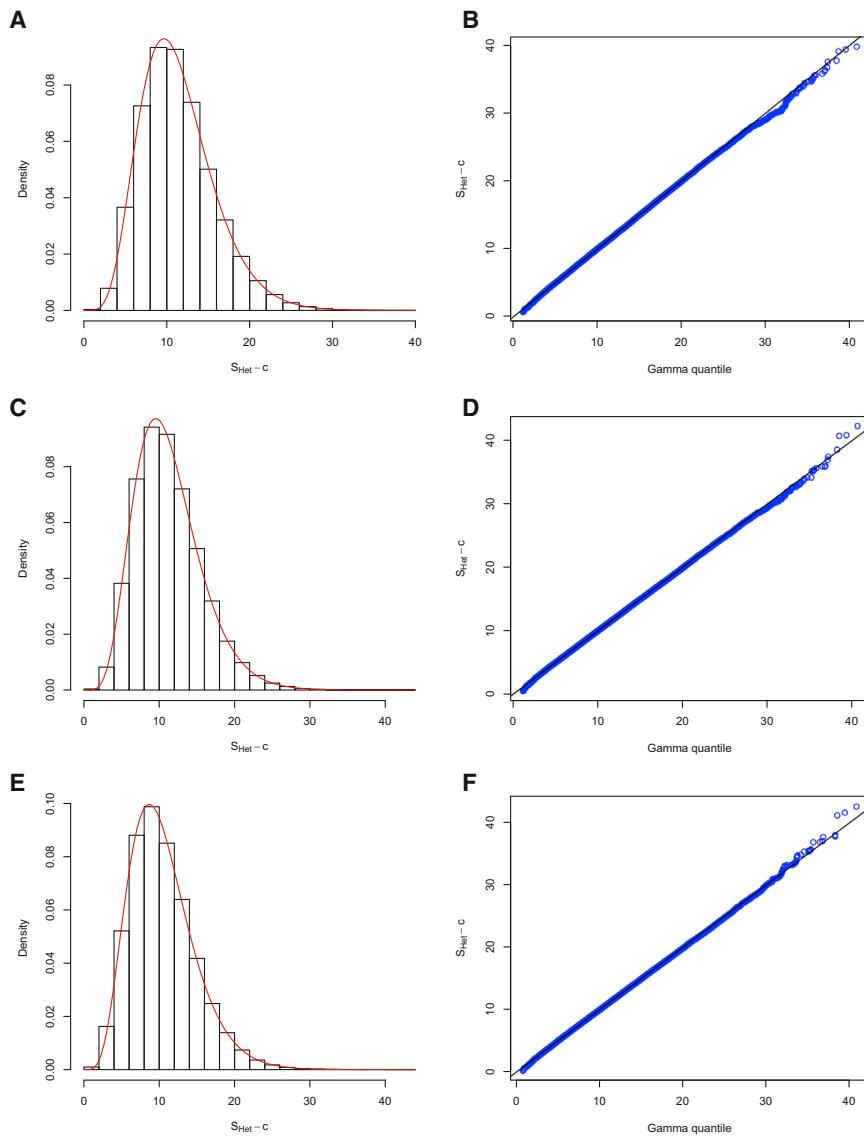
depend on a SNP, we can estimate the  $gamma(\alpha, \beta) + c$  that can be used for testing any SNP. This method is computationally feasible. We also observed that the type I error can be well controlled for  $S_{Hom}$ , which follows  $\chi_1^2$  under the null hypothesis. The well-controlled type I error rates for both  $S_{Hom}$  and  $S_{Het}$  regardless of nonoverlapping or overlapping samples among cohorts indicates that correlations estimated by Equation 6 are well enough for approximating the induced correlations among the summary statistics by overlapped samples.

no genetic contribution to any of the three traits. We performed the association tests for all the SNPs with the software Plink<sup>30</sup> to obtain the test statistics in each cohort separately. We next calculated  $S_{Het}$  for all the 100,000 SNPs and estimated the parameters  $\alpha$ ,  $\beta$ , and  $c$  for the shifted gamma distribution by matching the first three moments. Figure 1 depicts the empirical distributions of  $S_{Het} - \hat{c}$  under the three scenarios when no overlapping subjects were generated between the five cohorts. We observed that the estimated  $gamma(\hat{\alpha}, \hat{\beta})$  fits the empirical distribution of  $S_{Het} - \hat{c}$  well (Figure 1). In particular, we did not observe a departure in the tail of the gamma distribution. We observed similar results when there were 500 overlapping subjects among the five simulated cohorts (Figure 2). We next used these estimated gamma distribution parameters to evaluate the type I error rates correspondingly.

Under each of six scenarios ( $\rho = (0, 0.25, 0.5) \times$  (nonoverlapping, overlapping sample)), we repeated the above process 100 times, which led to  $10^7$   $S_{Hom}$  and  $S_{Het}$

### Power

We evaluated the power of the statistics  $S_{Hom}$  and  $S_{Het}$  by simulating three traits, SBP, DBP, and HTN. To simulate the three traits, a SNP was simulated and its genetic effect was added correspondingly. For illustration, we simulated two scenarios:  $\rho = 0$  and  $\rho = 0.5$ . In both scenarios, HTN is always correlated with SBP and DBP because of the way the data were simulated. In each scenario, we generated 5 cohorts, each with a sample size of 3,000 subjects. We



**Figure 2.  $S_{Het}$  Distribution when Cohorts Have Overlapping Subjects**

Distribution of the test statistic  $S_{Het}$  under three scenarios as in Figure 1. We generated 5 cohorts, each with sample size 3,000; 500 subjects were overlapping between cohorts. Left and right panels are as in Figure 1.

We first examined the power when SBP and DBP were simulated independently. Figure 3 shows the power when a genetic variant contributes to only one of the five cohorts, which represents heterogeneity across cohorts for the same trait. When a genetic variant contributes only to SBP,  $S_{Het}$  has the best power, followed by  $S_{Hom-SBP}$ ,  $S_{Hom}$ , and  $S_{Hom-HT}$  (Figure 3A).  $S_{Hom-DBP}$  did not have power because there was no genetic contribution to DBP. This result suggests that the proposed statistic  $S_{Het}$  is able to capture association evidence even when the data include noise because of heterogeneity among traits and cohorts.  $S_{Hom-SBP}$  has more power than  $S_{Hom}$ , which is not surprising because  $S_{Hom}$  included DBP, which had no genetic contribution. The overlap of subjects across the five cohorts is equivalent to reducing the number of subjects who did not have a genetic contribution from SBP. It is thus not surprising that  $S_{Hom-SBP}$  had the most power in this situation (Figure 3B).

When a genetic variant contributes to both SBP and DBP in one cohort with the effect sizes in the same direction, the combined trait analyses by  $S_{Hom}$  and  $S_{Het}$  had much improved power although  $S_{Het}$  still had the greatest power (Figure 3C). The power for  $S_{Hom}$  was further improved when there were overlapping samples between studies (Figure 3D), which could be attributed to reduced heterogeneity across cohorts. When a genetic variant contributed to both SBP and DBP in one cohort but with the effects in opposite directions,  $S_{Het}$  still maintained power, whereas  $S_{Hom}$  had almost no power because of the cancellation of the SBP and DBP contributions whether or not there were overlapping samples among cohorts (Figures 3E and 3F).

Figure 4 shows the power when a genetic variant contributes to all five cohorts, which represents no heterogeneity across cohorts for the same trait. When a genetic variant contributed only to SBP,  $S_{Hom-SBP}$  had the most power, regardless of whether there were overlapping samples or not between cohorts (Figures 4A and 4B). The power of

allowed different genetic contributions to the traits: a genetic variant contributes to a trait in only one of five cohorts (heterogeneity across cohorts within the same trait) or in all five cohorts (no heterogeneity within a trait), and a genetic variant affects SBP only (heterogeneity between traits) or affects both SBP and DBP. After both genotypes and phenotypes were simulated, we perform association tests for all SNPs with the software Plink and calculated  $S_{Hom}$  and  $S_{Het}$ . We calculated the p values for  $S_{Hom}$  by a  $\chi^2_1$  and with the previous estimated shift gamma distribution for  $S_{Het}$ , respectively. Power was defined as the proportion of test statistics with p values less than a corresponding significance level. We also examined the power when there were 500 overlapping subjects between cohorts. We analyzed the power of  $S_{Hom}$  for SBP, DBP, and HTN, separately as well as combined, and this is denoted as  $S_{Hom-SBP}$ ,  $S_{Hom-DBP}$ ,  $S_{Hom-HT}$ , and  $S_{Hom}$ , respectively. Power analysis was calculated based on 1,000 replications.

**Table 2. Type I Error Rates for  $S_{Hom}$  and  $S_{Het}$** 

$\rho$	No. of Overlapped Samples	Significance Level						
		0.05	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$10^{-6}$	$10^{-7}$
<b><math>S_{Hom}</math></b>								
0.00	0	$5.03 \times 10^{-2}$	$9.99 \times 10^{-3}$	$9.96 \times 10^{-4}$	$9.74 \times 10^{-5}$	$1.00 \times 10^{-5}$	$8.00 \times 10^{-7}$	$3.00 \times 10^{-7}$
	500	$4.76 \times 10^{-2}$	$9.18 \times 10^{-3}$	$9.02 \times 10^{-4}$	$9.05 \times 10^{-5}$	$6.50 \times 10^{-6}$	$8.00 \times 10^{-7}$	$2.00 \times 10^{-7}$
0.25	0	$5.04 \times 10^{-2}$	$1.02 \times 10^{-2}$	$1.03 \times 10^{-3}$	$1.07 \times 10^{-4}$	$1.12 \times 10^{-5}$	$7.00 \times 10^{-7}$	$0.00 \times 10^0$
	500	$4.94 \times 10^{-2}$	$9.81 \times 10^{-3}$	$9.52 \times 10^{-4}$	$9.57 \times 10^{-5}$	$9.50 \times 10^{-6}$	$7.00 \times 10^{-7}$	$1.00 \times 10^{-7}$
0.50	0	$5.06 \times 10^{-2}$	$1.02 \times 10^{-2}$	$1.03 \times 10^{-3}$	$1.06 \times 10^{-4}$	$1.23 \times 10^{-5}$	$1.10 \times 10^{-6}$	$0.00 \times 10^0$
	500	$4.74 \times 10^{-2}$	$9.21 \times 10^{-3}$	$8.65 \times 10^{-4}$	$7.97 \times 10^{-5}$	$7.30 \times 10^{-6}$	$7.00 \times 10^{-7}$	$2.00 \times 10^{-7}$
<b><math>S_{Het}</math></b>								
0.00	0	$4.99 \times 10^{-2}$	$1.00 \times 10^{-2}$	$1.03 \times 10^{-3}$	$1.10 \times 10^{-4}$	$1.17 \times 10^{-5}$	$2.00 \times 10^{-6}$	$1.00 \times 10^{-7}$
	500	$4.99 \times 10^{-2}$	$1.00 \times 10^{-2}$	$1.01 \times 10^{-3}$	$1.06 \times 10^{-4}$	$1.30 \times 10^{-5}$	$1.70 \times 10^{-6}$	$1.00 \times 10^{-7}$
0.25	0	$4.98 \times 10^{-2}$	$9.98 \times 10^{-3}$	$1.04 \times 10^{-3}$	$1.11 \times 10^{-4}$	$1.25 \times 10^{-5}$	$1.10 \times 10^{-6}$	$2.00 \times 10^{-7}$
	500	$4.98 \times 10^{-2}$	$1.00 \times 10^{-2}$	$1.02 \times 10^{-3}$	$1.11 \times 10^{-4}$	$1.28 \times 10^{-5}$	$8.00 \times 10^{-7}$	$0.00 \times 10^0$
0.50	0	$4.99 \times 10^{-2}$	$1.01 \times 10^{-2}$	$1.05 \times 10^{-3}$	$1.10 \times 10^{-4}$	$1.23 \times 10^{-5}$	$1.30 \times 10^{-6}$	$1.00 \times 10^{-7}$
	500	$4.99 \times 10^{-2}$	$1.01 \times 10^{-2}$	$1.05 \times 10^{-3}$	$1.15 \times 10^{-4}$	$1.26 \times 10^{-5}$	$1.30 \times 10^{-6}$	$2.00 \times 10^{-7}$

Type I error rate was calculated from the asymptotic  $\chi^2$  distribution with 1 d.f. for  $S_{Hom}$ , and the shift gamma distribution parameters were estimated by matching the first three moments for  $S_{Het}$ . We simulated ten million replications for each scenario.

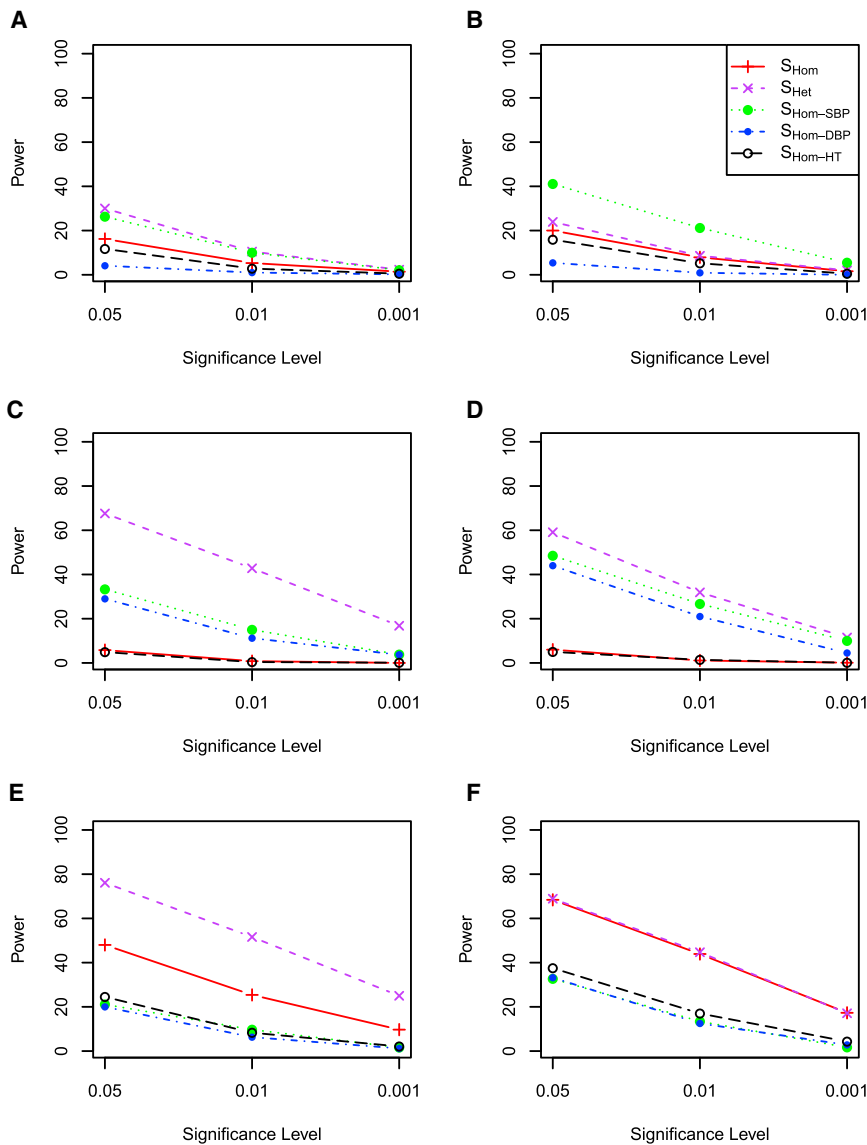
$S_{Hom}$  and  $S_{Het}$  were comparable, with  $S_{Het}$  outperforming  $S_{Hom}$  for no overlapping samples between cohorts. When a genetic variant contributed to both SBP and DBP and the effects were in the same direction (no trait or cohort heterogeneity),  $S_{Het}$  performed similarly to  $S_{Hom}$  (Figures 4C and 4D). However,  $S_{Hom}$  had no power although  $S_{Het}$  maintained power when the genetic effects were in opposite directions (Figures 4E and 4F).

The results were similar when SBP and DBP were highly correlated (correlation coefficient = 0.5). That is,  $S_{Het}$  usually outperformed or performed equivalently well as the other test statistics when heterogeneity due to cohorts or traits were present in our simulated data (Figures 5 and 6). Interestingly,  $S_{Het}$  improved power substantially when a genetic variant contributed to SBP and DBP in opposite directions but SBP and DBP were positively correlated, compared with no correlation (Figures 3E and 3F versus Figures 5E and 5F). Intuitively, if two traits are highly positively correlated, we are less likely to observe the estimated effects for a variant in opposite directions under the null hypothesis. The same is true if two traits are highly negatively correlated, wherein we will less likely observe the estimated effects for a variant in the same directions under the null hypothesis. The test statistic  $S_{Het}$  apparently captures this information whereas  $S_{Hom}$  does not.

#### Application to the BP GWAS Data from the Continental Origins and Genetic Epidemiology Network

We applied  $S_{Het}$  and  $S_{Hom}$  to the BP data from COGENT BP, which consists of 19 African ancestry cohorts, as

detailed in Franceschini et al.<sup>5</sup> In brief, the COGENT study includes 29,378 subjects, 20 years or older, from 18 U.S. African American cohorts and one cohort from Yoruba, Nigeria. The African American cohorts include Biological bank of Vanderbilt University (BioVU), Atherosclerosis Risk In Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Cleveland Family Study (CFS), Jackson Heart Study (JHS), Multi-Ethnic Study of Atherosclerosis (MESA), Cardiovascular Health Study (CHS), Genetic Study of Atherosclerosis Risk (GeneSTAR), Genetic Epidemiology Network of Arteriopathy (GENOA), Healthy Aging in Neighborhoods of Diversity Across the Life Span Study (HANDLS), Health, Aging, and Body Composition (Health ABC) Study, the Hypertension Genetic Epidemiology Network (HyperGEN), the Mount Sinai study (New York City, NY, USA), the Women's Health Initiative SNP Health Association Resource (WHI-SHARE), the Howard University Family Study (HUFS), the Bogalusa Heart Study (Bogalusa), the Sea Islands Genetic NETWORK (SIGNET) and REGARDS, and the Loyola Maywood study (Maywood). Each study received IRB approval of its consent procedures, examination and surveillance components, data security measures, and DNA collection and its use for genetic research. All participants in each study gave written informed consent for participation in the study and to conduct genetic research. Genotyping for the 19 cohorts was performed with either Affymetrix or the Illumina whole-genome SNP genotyping arrays. Quality control of genotyping data and imputation were performed in each cohort separately. Uniform protocols for analysis were conducted by each study. The summary statistics from the



**Figure 3. Power Comparison of  $S_{Hom}$  and  $S_{Het}$  when One Cohort Has Genetic Contribution**

SBP and DBP were simulated independently. HTN was simulated according to SBP and DBP and simulated medication status. Five cohorts were simulated, but only one of the five cohorts has a genetic contribution. Left: No overlapping samples among the five cohorts. Right: 500 samples were the same in each cohort and a genetic variant contributes phenotypic variation for the same samples.

(A and B) A genetic variant affects only SBP. (C and D) A genetic variant affects both SBP and DBP but with opposite effect directions.

(E and F) A genetic variant affects both SBP and DBP with the same effect direction.

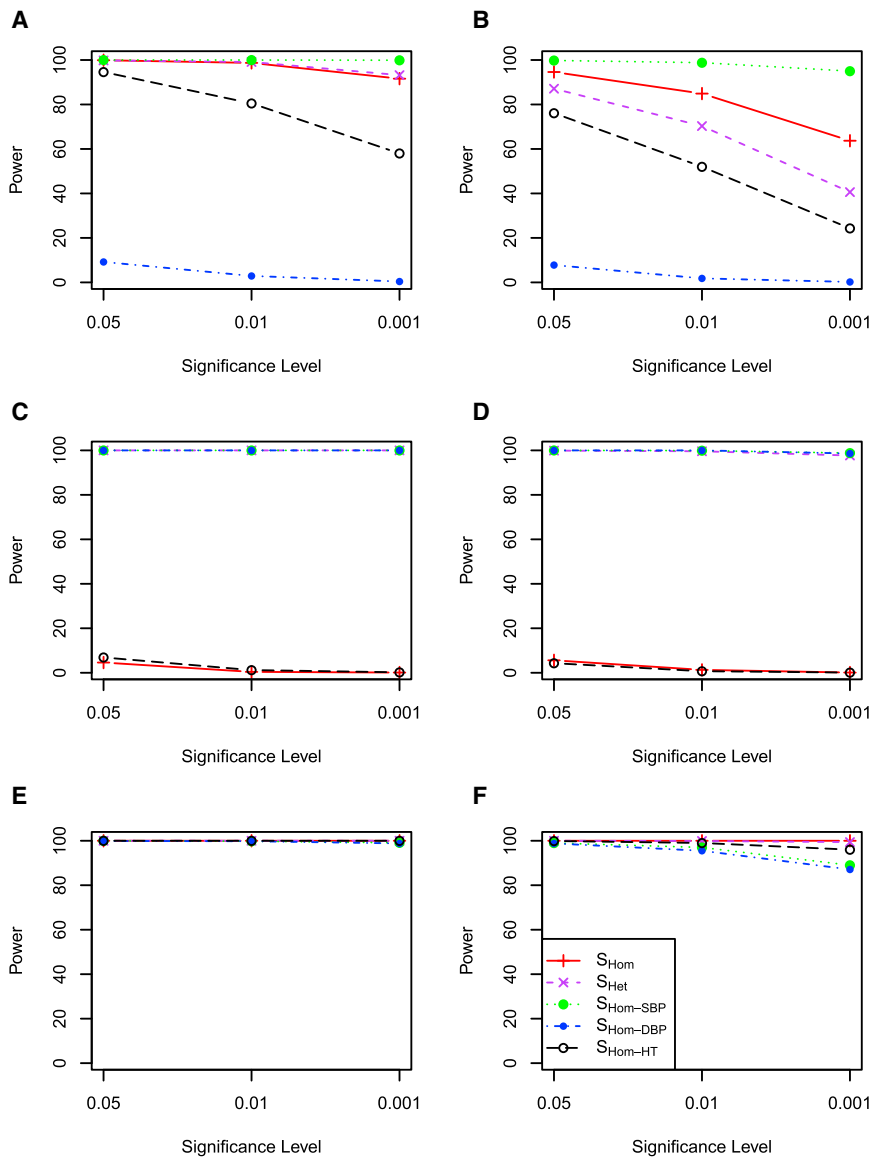
GWASs, including the SNP estimated effect sizes and their corresponding standard errors for SBP, DBP, and HTN, were collected for meta-analysis.<sup>5</sup> In this analysis, we applied  $S_{Het}$  and  $S_{Hom}$  to these summary statistics. Because we observed many unexpected large estimated effect sizes for the HTN analysis in SIGNET, we excluded the HTN results of SIGNET. Thus, our analysis included 56 trait-specific results from GWASs.

We obtained the inverse variance weighted meta-analysis results with the software METAL<sup>31</sup> for SBP, DBP, and HTN from the original report,<sup>5</sup> where SNP rs11041530 at *CYB5R2* (MIM 608342) is the only variant reaching genome-wide significance (SBP,  $p = 4.0 \times 10^{-8}$ ). We calculated the proposed statistics  $S_{Het}$  and  $S_{Hom}$  for each SNP and the corresponding p values with a gamma distribution for combining SBP, DBP, and HTN. Figure 7 presents the QQ plots and Manhattan plots of  $S_{Het}$  and  $S_{Hom}$ . The genomic control (GC) inflation factors are 1.08 and 1.05 for  $S_{Het}$  and  $S_{Hom}$ , respectively. A possible

reason for the slightly inflated GC factors could be that the hypertension traits are polygenic with a large number of genetic variants of small effect sizes contributing to the phenotypic variation. Combining SBP, DBP, and HTN would further aggregate these variants.  $S_{Hom}$  detected the *HOXA-EVX1* (MIM 142996) locus on chromosome 7 at a genome-wide significance level (Table 3 and Figure 7,  $p = 2.35 \times 10^{-9}$ ). This locus was also reported in the original study and was replicated in Asian and European populations, although SBP, DBP, or HTN trait-specific meta-analyses did not reach genome-wide significance in the discovery phase.<sup>5</sup> In comparison,  $S_{Het}$  was able to detect four

loci at genome-wide significance level ( $p < 5.0 \times 10^{-8}$ ), including *HOXA-EVX1* on chromosome 7, *CHIC2* (MIM 604332) on chromosome 4, *IGFBP1-IGFBP3* (MIM 146730, 146732) on chromosome 7, and *CDH17* (MIM 603017) on chromosome 8 (Table 3 and Figure 7). The regional plots for these four loci are presented in Figure 8. Figure S1 (available online) shows forest plots of the cohort-specific effect sizes of SBP, DBP, and HTN for these four loci. Note that these loci have opposite directions in SBP and DBP meta-analysis except *HOXA-EVX1* (Table 3), suggesting that the same genetic variant increases SBP but decreases DBP level (i.e., increases pulse pressure) or vice versa. Interestingly, *CHIC2* has been reported to be associated with pulse pressure by large GWASs in a European population.<sup>32</sup> The most significant variant, rs11725861 in *CHIC2*, is located 3 kb away from the sentinel SNP rs871606 reported in Wain et al.,<sup>32</sup> and these two SNPs are in strong linkage disequilibrium in HapMap CEU data ( $r^2 = 0.35$ ,  $D' = 1$ ). Because pulse





**Figure 4. Power Comparison of  $S_{Hom}$  and  $S_{Het}$  when Five Cohorts Have Genetic Contribution**

Five cohorts were simulated and the genetic variant has contribution in all five cohorts. Details as in Figure 3.

pressure, although linkage evidence was reported for this locus to essential hypertension in European population.<sup>36</sup>

Six additional loci were also identified by  $S_{Het}$  with suggestive evidence ( $p < 5 \times 10^{-7}$ , Table 3). The mutations in *CACNA1D* (MIM 114206), encoding Cav1.3, were reported to regulate  $Na^+$ ,  $K^+$ , and  $Ca^{2+}$  and to underlie a common subtype of adrenal hypertension.<sup>37</sup> The SNP rs6886515 on chromosome 5 is located in an intergenic region. SNP rs9401512 is in *HSF2/PKIB* (MIM 140581, 606914) on chromosome 6 and has not been reported to be associated with hypertension-related traits. However, this region has been reported to harbor BP variants in admixture mapping analyses.<sup>38,39</sup> SNP rs11041530 in *CYB5R2* was reported in the original study but was not significant in a replication analysis.<sup>5</sup> We have not seen any report of *PLXNC1* (MIM 604259) and hypertension. SNP rs430685 locates in *WNT3* (MIM 165330) on chromosome 17 and this locus is close to the locus *GOSR2* (MIM 604027) reported in Ehret

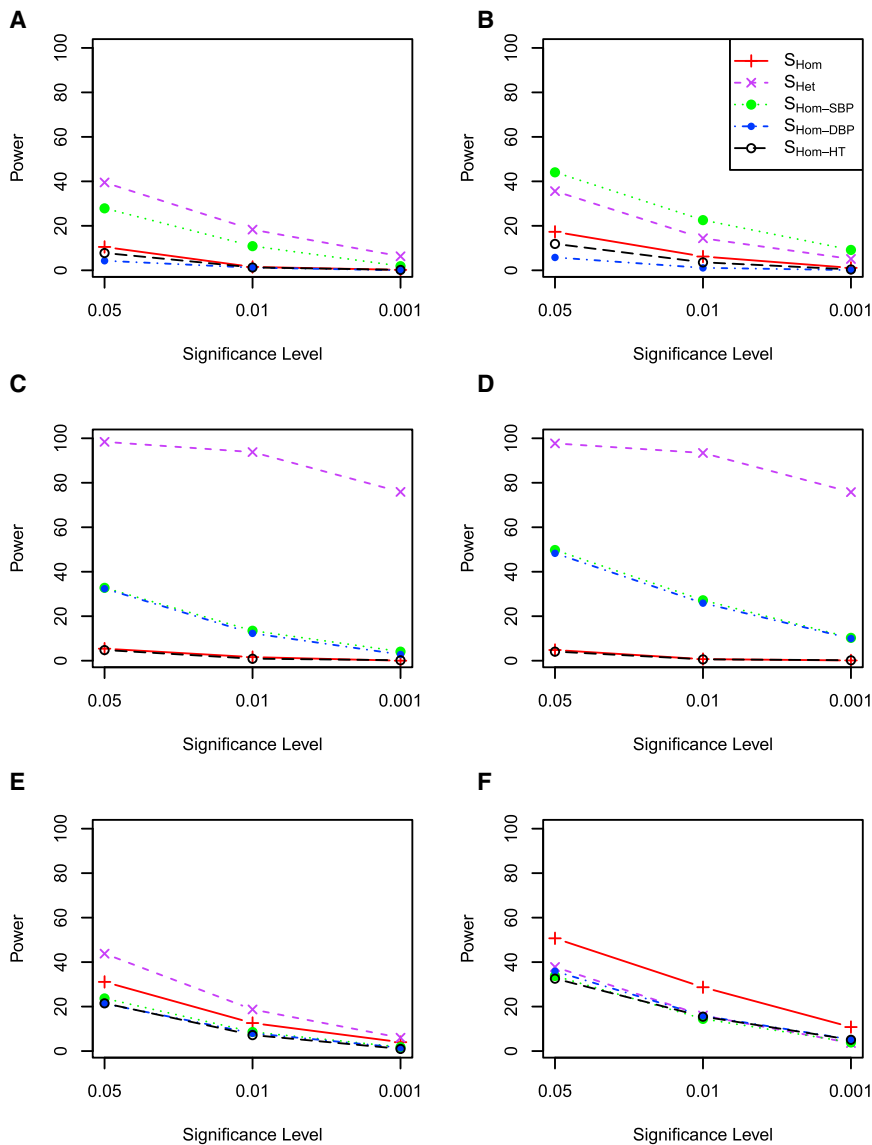
et al.<sup>1</sup> SNP rs430685 is 153 kb away from the sentinel SNP rs17608766 reported in Ehret et al.<sup>1</sup> with  $D'$  of 1.0 but  $r^2$  value of 0.003 between these two SNPs in HapMap CEU data, indicating that these two signals might be different. *WNT3* belongs to the Wnt/frizzled receptor signaling pathway, which has been suggested to have an important functional role in cardiovascular and renal disorders.<sup>40</sup>

pressure is defined as the difference between SBP and DBP, the opposite effect sizes of SBP and DBP in this study are thus consistent with the reported association evidence with pulse pressure in European population. The IGF system is implicated in the development of cardiovascular disease. Low circulating levels of *IGFBP1* have been reported to be associated with the presence of macrovascular disease and hypertension in type 2 diabetes,<sup>33,34</sup> although there is no direct report of association evidence between the variants at *IGFBP1* and blood pressure. However, a recent meta-analysis of European cohorts identified *IGFBP3* as being associated with long-term averaging of pulse blood pressure.<sup>35</sup> The most significant variant, rs11977526 near *IGFBP3*, is located 43 kb away from the sentinel SNP rs2949837 reported in Ganesh et al.<sup>35</sup> and these two SNPs are in strong linkage disequilibrium in HapMap CEU data ( $r^2 = 0.66$ ,  $D' = 0.94$ ). There has been no association reported between *CDH17* on chromosome 8q21 and blood

et al.<sup>1</sup> SNP rs430685 is 153 kb away from the sentinel SNP rs17608766 reported in Ehret et al.<sup>1</sup> with  $D'$  of 1.0 but  $r^2$  value of 0.003 between these two SNPs in HapMap CEU data, indicating that these two signals might be different. *WNT3* belongs to the Wnt/frizzled receptor signaling pathway, which has been suggested to have an important functional role in cardiovascular and renal disorders.<sup>40</sup>

## Discussion

The observations from GWASs suggest that many genetic variants are associated with multiple correlated or even distinct traits, and such associations have been termed cross-phenotype (CP) associations,<sup>4</sup> which is relevant to pleiotropy in complex disease. We proposed a statistic  $S_{Het}$  and compared it with  $S_{Hom}$ , and both methods can integrate association evidence of multiple continuous



**Figure 5. Power Comparison of  $S_{Hom}$  and  $S_{Het}$  with Correlation 0.5 when One Cohort Has Genetic Contribution**  
 SBP and DBP were simulated with correlation 0.5. Five cohorts were simulated but only one of the five cohorts has a genetic contribution. Details as in Figure 3.

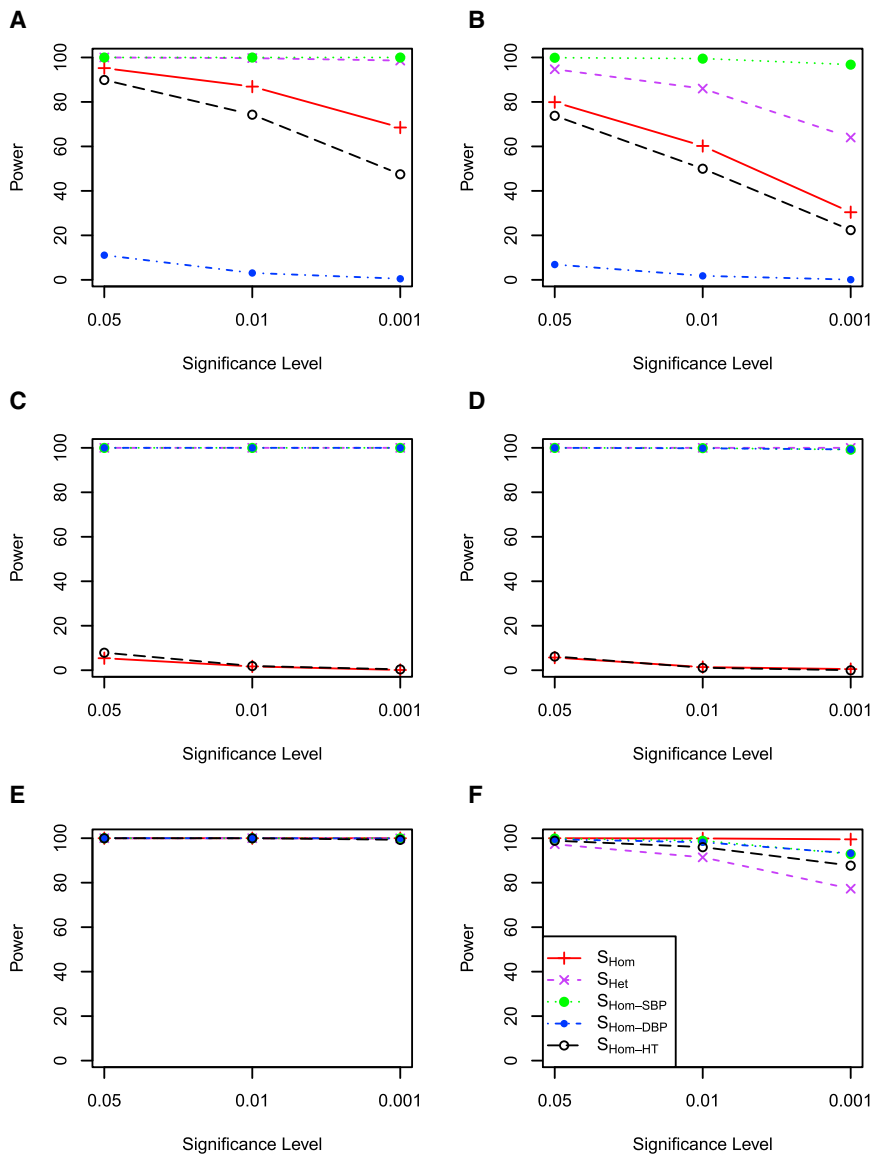
ures 1 and 2). The estimated gamma distributions well fitted the corresponding empirical distributions as observed from both histograms and the Q-Q plots of  $S_{Het}$  (Figures 1 and 2). When we applied both methods to the data from COGENT BP African ancestry GWASs for blood pressure traits, we did not observe any substantial inflation of the type I error rate (Figure 7), further indicating that both methods are valid.

Our simulations suggest that when heterogeneity is of less concern,  $S_{Hom}$  is more powerful than  $S_{Het}$ . In contrast, when heterogeneity is present,  $S_{Het}$  is more powerful than  $S_{Hom}$ . This property can also be observed from the application of both statistics to the data from COGENT BP African ancestry GWASs for blood pressure traits.  $S_{Hom}$  was able to detect the *HOXA-EVX1* locus ( $p = 2.35 \times 10^{-9}$ ) whereas  $S_{Het}$  identified four loci (*CHIC2*, *HOXA-EVX1*, *IGFBP1/IGFBP3*, and *CDH17*;  $p < 5.0 \times 10^{-8}$ ) at a genome-wide significance level. All four loci were missed by single-trait analysis at genome-wide significance level ( $p < 5.0 \times$

and binary traits from multiple GWASs and thus detect CP associations. Both methods need only the summary statistics obtained from GWASs.  $S_{Hom}$  is an extension of the linear combination of the univariate test statistics<sup>19,20</sup> but allows for sample size as weights.  $S_{Het}$  is a further extension of  $S_{Hom}$  allowing for heterogeneous effects of a trait from different studies, which could be due to different designs, environmental factors, or populations, as well as heterogeneity effects for different phenotypes, which are not unexpected in practice. Under the null hypothesis,  $S_{Hom}$  is asymptotically distributed as chi-square with 1 degree of freedom, although the asymptotical distribution of  $S_{Het}$  is less clear but related to a gamma distribution.<sup>28</sup> Our simulations indeed suggest that  $S_{Het}$  can be well approximated by a shift gamma distribution with parameters that can be estimated from the data.

The simulations indicated that the type I error rate is reasonably controlled for both methods (Table 2 and Fig-

$10^{-8}$ ) and only the *HOXA-EVX1* locus was identified by the flipping sign test in the original report,<sup>5</sup> suggesting that  $S_{Het}$  is more powerful than combined p values methods when heterogeneity is present. The flipping sign test is similar to Fisher's method for combining the p values for different traits but requires a large number of simulations to estimate the null distribution to account for the trait correlations. Thus the flipping sign test is similar to  $S_{Hom}$  but  $S_{Hom}$  is more flexible and requires substantially less computational time because of using the estimated asymptotic distribution. The HTN-, SBP-, and DBP-specific meta-analysis of the *HOXA-EVX1* locus did not show any heterogeneity (Table 3), which is the reason that this locus was detected by  $S_{Hom}$ . As suggested by our simulations,  $S_{Het}$  is more applicable to heterogeneous data and was able to identify three additional loci, *CHIC2*, *IGFBP1/IGFBP3*, and *CDH17*. This is also consistent with the fact that the effect sizes estimated in the SBP- and DBP-specific meta-analyses were in opposite directions



**Figure 6. Power Comparison of  $S_{Hom}$  and  $S_{Het}$  with Correlation 0.5 when Five Cohorts Have Genetic Contribution**  
 SBP and DBP were simulated with correlation 0.5. Five cohorts were simulated and the genetic variant has a contribution in all five cohorts. Details as in Figure 3.

tions of test statistics among traits or cohorts,  $S_{Hom}$  and  $S_{Het}$  are able to control the effect of cryptic relatedness occurring among cohorts, as observed in the simulations when overlapping samples were simulated among cohorts (Figures 3, 4, 5, and 6).  $S_{Hom}$  and  $S_{Het}$  are in principle able to control the effect of population structure occurring between cohorts, although this property of the methods has not been formally evaluated in our simulation studies.

In contrast, a fixed or random effect meta-analysis is able to combine results from multiple studies but is limited to the same traits and no overlapping or related subjects across studies. The cross-phenotype meta-analysis<sup>14</sup> is able to test CP association but cannot be applied when there are overlapping subjects. The linear combination of univariate test statistics applies only to a single study with multiple traits and requires individual-level genotype and phenotype data.<sup>17,18</sup> The TATES requires only p values, not the individual-level genotype and phenotype data, but cannot be applied to multi-

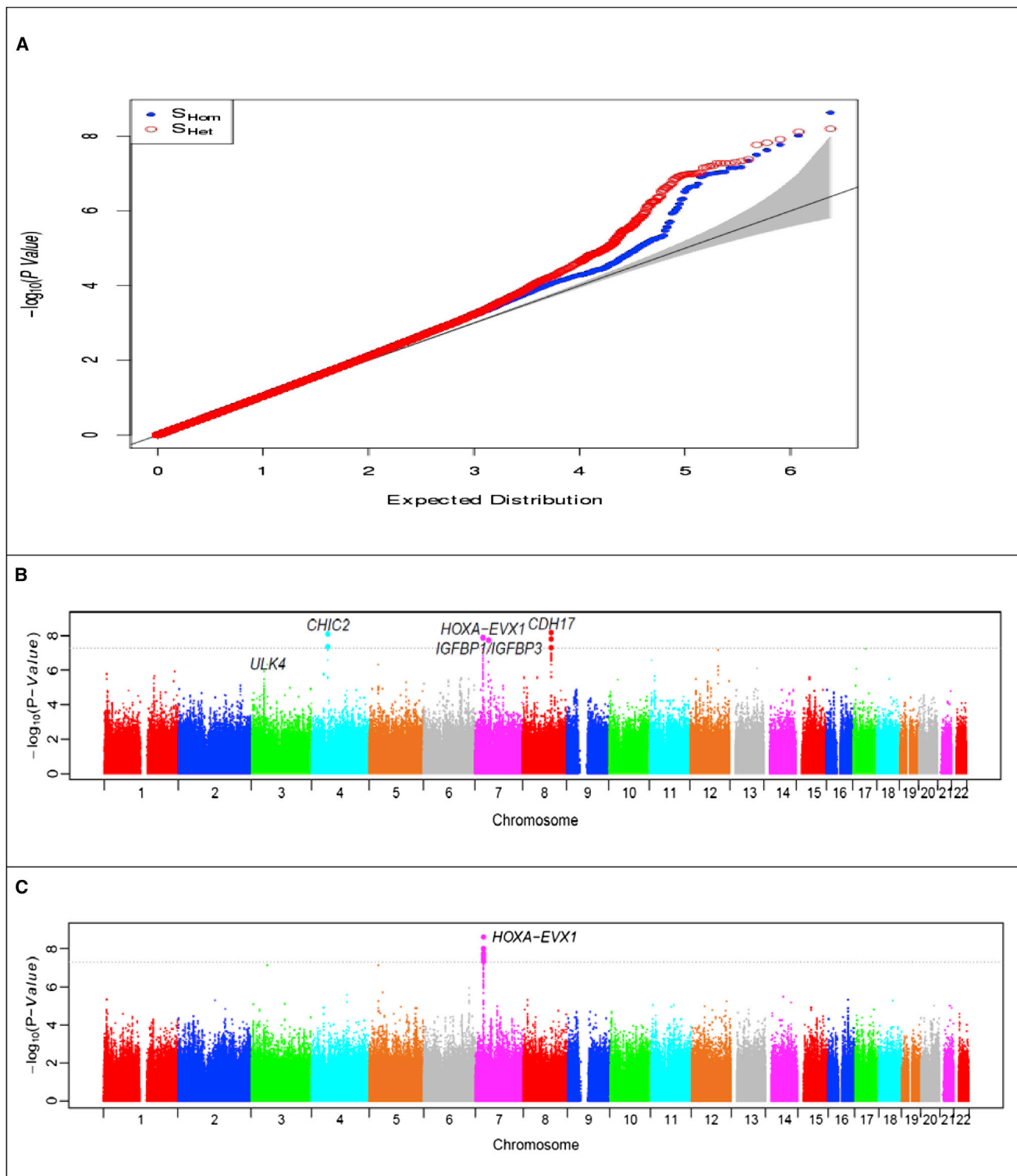
ple studies.<sup>23</sup> The scaled marginal model proposed by Schifano et al.<sup>41</sup> requires individual-level genotype and phenotype data. Thus,  $S_{Hom}$  and  $S_{Het}$  are quite general and can be applied to a wide range of data and study designs, and they require only that estimates exist for a given SNP.

for all three of these loci (Table 3). In addition,  $S_{Het}$  was able to identify six loci with suggestive association evidence for hypertension-related traits. Among the identified loci, *CHIC2*, *HOXA-EVX1*, *IGFBP1/IGFBP3*, *CACNA1D*, and *GOSR2/WNT3* have been confirmed to be associated with hypertension-related traits,<sup>5,32–35</sup> suggesting that our methods powerfully identify true signals.

Compared with existing methods,  $S_{Hom}$  and  $S_{Het}$  have multiple advantages for identifying cross-phenotype (CP) associations. Both methods are able to combine traits measured on different scales, including continuous and binary traits. Further,  $S_{Het}$  allows for heterogeneous effects. Both methods are able to accommodate overlapping or related subjects within and among different studies or cohorts, as our simulations suggested. We assumed that the trait-specific summary statistics have already well accounted for the confounding effect caused by either population structure or cryptic relatedness within a cohort. Since our approach accounts for correla-

tion of test statistics among traits or cohorts,  $S_{Hom}$  and  $S_{Het}$  are able to control the effect of cryptic relatedness occurring among cohorts, as observed in the simulations when overlapping samples were simulated among cohorts (Figures 3, 4, 5, and 6).  $S_{Hom}$  and  $S_{Het}$  are in principle able to control the effect of population structure occurring between cohorts, although this property of the methods has not been formally evaluated in our simulation studies.

In contrast to the subset-based meta-analysis,<sup>18</sup>  $S_{Hom}$  and  $S_{Het}$  do not specifically identify a subset of associated traits. However, trait-specific meta-analysis results can be examined after a SNP is identified by  $S_{Het}$ , which will allow one to see whether the association evidence from  $S_{Het}$  is contributed by only a subset of the traits. When maximizing the statistic in Equation 3, we can record which trait or cohort contributed to the final statistic. The current  $S_{Hom}$  and  $S_{Het}$  test CP associations for only one SNP. Huang et al.<sup>24</sup> developed an interesting method, PRIME, which can test a pleiotropic effect for multiple variants in a genomic region. It should not be difficult to extend  $S_{Hom}$  and  $S_{Het}$  to test for multiple



**Figure 7.** QQ Plots and Manhattan Plots after Combining SBP, DBP, and HTN via  $S_{Hom}$  and  $S_{Het}$  for the COGENT BP GWAS Data Shown are QQ plots (A), Manhattan plot of  $S_{Het}$  (B), and Manhattan plot of  $S_{Hom}$  (C).

variants in a genomic region by using a similar idea as the PRIME uses, although further studies are warranted to investigate this.

Our proposed statistics  $S_{Hom}$  and  $S_{Het}$  use the summary statistics to estimate the correlation coefficients among

traits and cohorts. If trait correlations are known,  $S_{Hom}$  and  $S_{Het}$  can also be applied by supplying the correlation matrix in Equation 3 without using genome-wide summary statistics. There are advantages to using summary statistics instead of individual-level data. First, as pointed out

**Table 3. The Top Genetic Variants Identified by  $S_{Het}$  and  $S_{Hom}$  in the COGENT BP Consortium by First Performing Meta-analysis and then  $S_{Het}$  and  $S_{Hom}$**

Chr	SNP	POS	Genes	Effect Allele	Freq	Meta-analysis HTN			Meta-analysis SBP			Meta-analysis DBP			$S_{Hom}$			$S_{Het}$		
						beta	SE	P	beta	SE	P	beta	SE	P	beta	SE	P	beta	SE	P
<b>Loci Reaching Genome-wide Significance</b>																				
4	rs11725861	54497062	CHIC2	A	0.84	0.04	0.03	$2.16 \times 10^{-1}$	0.79	0.22	$2.89 \times 10^{-4}$	-0.18	0.13	$1.63 \times 10^{-1}$	$2.56 \times 10^{-1}$	$8.45 \times 10^{-9}$				
7	rs11564022	27303571	HOXA-EVX1	T	0.23	-0.12	0.02	$2.16 \times 10^{-6}$	-0.89	0.19	$1.83 \times 10^{-6}$	-0.60	0.11	$7.66 \times 10^{-8}$	$2.35 \times 10^{-9}$	$1.34 \times 10^{-8}$				
7	rs11977526	45974635	IGFBP1, IGFBP3	A	0.32	-0.01	0.02	$5.30 \times 10^{-1}$	-0.37	0.18	$3.73 \times 10^{-2}$	0.30	0.11	$4.62 \times 10^{-3}$	$8.21 \times 10^{-1}$	$1.87 \times 10^{-8}$				
8	rs2446849	95172673	CDH17	T	0.80	-0.06	0.03	$3.00 \times 10^{-2}$	-0.63	0.20	$2.00 \times 10^{-3}$	0.22	0.12	$6.75 \times 10^{-2}$	$2.11 \times 10^{-1}$	$7.01 \times 10^{-9}$				
<b>Loci Reaching Suggestive Significance</b>																				
3	rs10049492	53571572	CACNA1D	A	0.74	0.12	0.03	$1.42 \times 10^{-5}$	0.85	0.19	$1.74 \times 10^{-5}$	0.57	0.12	$1.77 \times 10^{-6}$	$7.11 \times 10^{-8}$	$4.60 \times 10^{-7}$				
5	rs6886515	30702383	none	A	0.61	0.11	0.02	$4.61 \times 10^{-7}$	0.71	0.17	$2.46 \times 10^{-5}$	0.40	0.10	$6.02 \times 10^{-5}$	$7.15 \times 10^{-8}$	$4.63 \times 10^{-7}$				
6	rs9401512	122707473	HSF2/PKIB	T	0.30	0.00	0.02	$9.17 \times 10^{-1}$	0.44	0.17	$1.04 \times 10^{-2}$	-0.23	0.10	$2.75 \times 10^{-2}$	$9.43 \times 10^{-1}$	$5.75 \times 10^{-8}$				
11	rs11041530	7658079	CYB5R2	C	0.12	-0.09	0.03	$9.83 \times 10^{-3}$	-1.35	0.25	$4.04 \times 10^{-8}$	-0.54	0.15	$2.65 \times 10^{-4}$	$9.08 \times 10^{-6}$	$2.55 \times 10^{-7}$				
12	rs11837544	93192534	PLXNC1	A	0.20	-0.01	0.03	$5.89 \times 10^{-1}$	-0.50	0.20	$1.36 \times 10^{-2}$	0.27	0.12	$2.28 \times 10^{-2}$	$9.37 \times 10^{-1}$	$6.89 \times 10^{-8}$				
17	rs430685	42214309	GOSR2/WNT3	T	0.02	-0.42	0.14	$2.65 \times 10^{-3}$	1.26	0.84	$1.33 \times 10^{-1}$	1.51	0.50	$2.64 \times 10^{-3}$	$8.07 \times 10^{-1}$	$5.77 \times 10^{-8}$				

by Lin and Zeng,<sup>42</sup> there is no asymptotic efficiency gain by analyzing individual-level data compared with meta-analysis when the parameter of interest has a common value across studies, although this aspect of performance is less clear when heterogeneity is present. Second, in practice it is easier and more feasible to obtain summary statistics than individual-level data. Finally, study-specific analysis is better to control the confounding within different study designs by environmental factors and between study designs by batch effects across experiments than pooling all data.

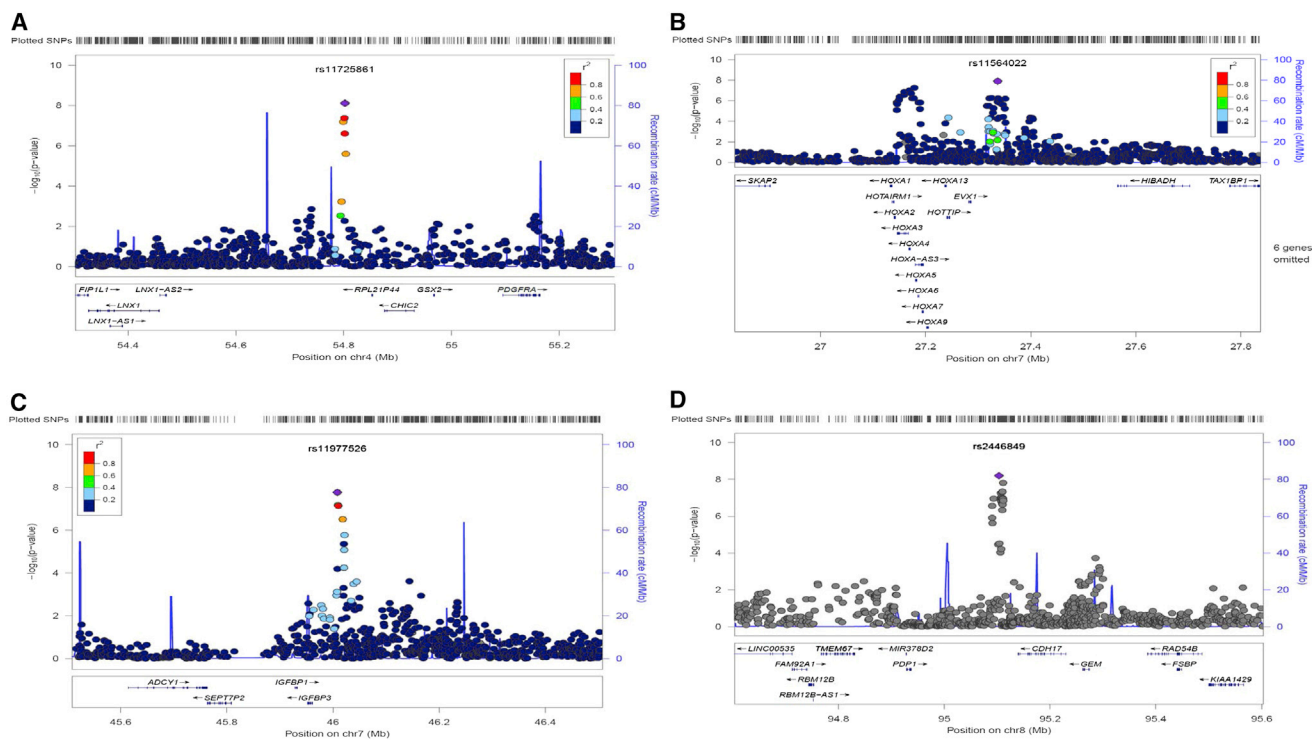
In summary, the proposed general statistics  $S_{Hom}$  and  $S_{Het}$  are useful for detecting CP associations. In particular,  $S_{Het}$  is better for analyzing multiple different phenotypes because heterogeneity occurs frequently. These methods could easily be deployed in existing consortia collections of association study metadata to improve the chances of novel discoveries and provide more return from those investments, as we have demonstrated in the COGENT consortium. The software of  $S_{Hom}$  and  $S_{Het}$  can be freely downloaded from the author's website.

### Supplemental Data

Supplemental Data include one figure and supplemental consortia information and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2014.11.011>.

### Consortia

The members of the COGENT BP Consortium are Nora Franceschini, Ervin Fox, Zhaogong Zhang, Todd L. Edwards, Michael A. Nalls, Yun Ju Sung, Bamidele O. Tayo, Yan V. Sun, Omri Gottesman, Adebawole Adeyemo, Andrew D. Johnson, J. Hunter Young, Ken Rice, Qing Duan, Fang Chen, Yun Li, Hua Tang, Myriam Fornage, Keith L. Keene, Jeanette S. Andrews, Jennifer A. Smith, Jessica D. Faul, Zhang Guangfa, Wei Guo, Yu Liu, Sarah S. Murray, Solomon K. Musani, Sathanur Srinivasan, Digna R. Velez Edwards, Heming Wang, Lewis C. Becker, Pascal Bovet, Murielle Bochud, Ulrich Broeckel, Michel Burnier, Cara Carty, Wei-Min Chen, Guanjie Chen, Wei Chen, Jingzhong Ding, Albert W. Dreisbach, Michele K. Evans, Xiuqing Guo, Melissa E. Garcia, Rich Jensen, Margaux F. Keller, Guillaume Lettre, Vaneet Lotay, Lisa W. Martin, Alanna C. Morrison, Thomas H. Mosley, Adesola Ogunniyi, Walter Palmas, George Papanicolaou, Alan Penman, Joseph F. Polak, Paul M. Ridker, Babatunde Salako, Andrew B. Singleton, Daniel Shriner, Kent D. Taylor, Ramachandran Vasani, Kerri Wiggins, Scott M. Williams, Lisa R. Yanek, Wei Zhao, Alan B. Zonderman, Diane M. Becker, Gerald Berenson, Eric Boerwinkle, Erwin Bottinger, Mary Cushman, Charles Eaton, Gerardo Heiss, Joel N. Hirschhorn, Virginia J. Howard, Matthew B. Lanktree, Kiang Liu, Yongmei Liu, Ruth Loos, Karen Margolis, Bruce M. Psaty, Nicholas J. Schork, David R. Weir, Charles N. Rotimi, Michele M. Sale, Tamara Harris, Sharon L.R. Kardia, Steven C. Hunt, Donna Arnett, Susan Redline, Richard S. Cooper, Neil Risch, D.C. Rao, Jerome I. Rotter, Aravinda Chakravarti, Alex P. Reiner, Daniel Levy, Brendan J. Keating, and Xiaofeng Zhu.



**Figure 8. Regional Association Plots**

Regional association plots of the four SNPs reaching genome-wide significance ( $p < 5 \times 10^{-8}$ ) by  $S_{Het}$  for the COGENT BP GWAS data. The most significant SNP at each locus is shown in purple. The fine-scale recombination rate is shown as a blue vertical line. Gene positions are shown at the bottom.

## Acknowledgments

We are gratefully indebted to Robert C. Elston for his valuable discussions and suggestions that greatly improved the manuscript. The work was supported by the NIH grants HG003054 from the National Human Genome Research Institute and HL086718, HL053353, HL113338, and HL123677 from the National Heart, Lung, and Blood Institute. Funding information for the COGENT BP Consortium is provided in the [Supplemental Data](#).

Received: August 21, 2014

Accepted: November 17, 2014

Published: December 11, 2014

## Web Resources

The URLs for data presented herein are as follows:

CPASSOC software, <http://hal.case.edu/zhu-web/>  
 GWAS Catalog, <http://www.genome.gov/gwastudies/>  
 Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>

## References

- Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J., et al.; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM

- consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; CHARGE-HF consortium (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109.
- Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713.
- Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838.
- Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., and Smoller, J.W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14**, 483–495.
- Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y.V., Gottesman, O., Adeyemo, A., et al.; Asian Genetic Epidemiology Network Consortium (2013). Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am. J. Hum. Genet.* **93**, 545–554.
- Zeger, S.L., and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.
- Lange, C., Silverman, E.K., Xu, X., Weiss, S.T., and Laird, N.M. (2003). A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* **4**, 195–206.

8. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11, 407–409.
9. Zhang, H., Liu, C.T., and Wang, X. (2010). An association test for multiple traits based on the generalized Kendall's tau. *J. Am. Stat. Assoc.* 105, 473–481.
10. O'Reilly, P.F., Hoggart, C.J., Pomyen, Y., Calboli, F.C., Elliott, P., Jarvelin, M.R., and Coin, L.J. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* 7, e34861.
11. Ott, J., and Rabinowitz, D. (1999). A principal-components approach based on heritability for combining phenotype information. *Hum. Hered.* 49, 106–111.
12. Klei, L., Luca, D., Devlin, B., and Roeder, K. (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genet. Epidemiol.* 32, 9–19.
13. Aschard, H., Vilhjálmsson, B.J., Grelliche, N., Morange, P.E., Trégouët, D.A., and Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am. J. Hum. Genet.* 94, 662–676.
14. Cotsapas, C., Voight, B.F., Rossin, E., Lage, K., Neale, B.M., Wallace, C., Abecasis, G.R., Barrett, J.C., Behrens, T., Cho, J., et al.; FOCiS Network of Consortia (2011). Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* 7, e1002254.
15. Skroza, N., Proietti, I., Pampena, R., La Viola, G., Bernardini, N., Nicolucci, E., Tolino, E., Zuber, S., Soccodato, V., and Potenza, C. (2013). Correlations between psoriasis and inflammatory bowel diseases. *Biomed Res Int* 2013, 983902.
16. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* 42, 1118–1125.
17. Strange, A., Capon, F., Spencer, C.C., Knight, J., Weale, M.E., Allen, M.H., Barton, A., Band, G., Bellenguez, C., Bergboer, J.G., et al.; Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* 42, 985–990.
18. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., Yeager, M., Chung, C.C., Chanock, S.J., and Chatterjee, N.; GliomaScan Consortium (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am. J. Hum. Genet.* 90, 821–835.
19. Wei, L.J., and Johnson, W.E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* 72, 359–364.
20. O'Brien, P.C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* 40, 1079–1087.
21. Xu, X., Tian, L., and Wei, L.J. (2003). Combining dependent tests for linkage or association across multiple phenotypic traits. *Biostatistics* 4, 223–229.
22. Yang, Q., Wu, H., Guo, C.Y., and Fox, C.S. (2010). Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genet. Epidemiol.* 34, 444–454.
23. van der Sluis, S., Posthuma, D., and Dolan, C.V. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* 9, e1003235.
24. Huang, J., Johnson, A.D., and O'Donnell, C.J. (2011). PRIME: a method for characterization and evaluation of pleiotropic regions from multiple genome-wide association studies. *Bioinformatics* 27, 1201–1206.
25. Li, Y., Feng, T., and Zhu, X. (2011). Detecting association with rare variants for common diseases using haplotype-based methods. *Stat. Interface* 4, 273–284.
26. Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H., and Weir, B.S. (2002). Truncated product method for combining P-values. *Genet. Epidemiol.* 22, 170–185.
27. Zhu, X., Feng, T., and Elston, R.C. (2012). Linkage-disequilibrium-based binning misleads the interpretation of genome-wide association studies. *Am. J. Hum. Genet.* 91, 965–968, author reply 969–970.
28. Alam, K., and Wallenius, K.T. (1979). Distribution of a sum of order statistics. *Scand. J. Stat.* 6, 123–126.
29. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 41, 677–687.
30. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
31. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.
32. Wain, L.V., Verwoert, G.C., O'Reilly, P.F., Shi, G., Johnson, T., Johnson, A.D., Bochud, M., Rice, K.M., Henneman, P., Smith, A.V., et al.; LifeLines Cohort Study; EchoGen consortium; AortaGen Consortium; CHARGE Consortium Heart Failure Working Group; KidneyGen consortium; CKDGen consortium; Cardiogenics consortium; CardioGram (2011). Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* 43, 1005–1011.
33. Heald, A.H., Siddals, K.W., Fraser, W., Taylor, W., Kaushal, K., Morris, J., Young, R.J., White, A., and Gibson, J.M. (2002). Low circulating levels of insulin-like growth factor binding protein-1 (IGFBP-1) are closely associated with the presence of macrovascular disease and hypertension in type 2 diabetes. *Diabetes* 51, 2629–2636.
34. Rajwani, A., Ezzat, V., Smith, J., Yuldasheva, N.Y., Duncan, E.R., Gage, M., Cubbon, R.M., Kahn, M.B., Imrie, H., Abbas, A., et al. (2012). Increasing circulating IGFBP1 levels improves insulin sensitivity, promotes nitric oxide production, lowers blood pressure, and protects against atherosclerosis. *Diabetes* 61, 915–924.
35. Ganesh, S.K., Chasman, D.I., Larson, M.G., Guo, X., Verwoert, G., Bis, J.C., Gu, X., Smith, A.V., Yang, M.L., Zhang, Y., et al.; Global Blood Pressure Genetics Consortium (2014). Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. *Am. J. Hum. Genet.* 95, 49–65.
36. Ciullo, M., Bellenguez, C., Colonna, V., Nutile, T., Calabria, A., Pacente, R., Iovino, G., Trimarco, B., Bourgain, C., and Persico, M.G. (2006). New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate. *Hum. Mol. Genet.* 15, 1735–1743.
37. Azizan, E.A., Poulsen, H., Tuluc, P., Zhou, J., Clausen, M.V., Lieb, A., Maniero, C., Garg, S., Bochukova, E.G., Zhao, W., et al. (2013). Somatic mutations in ATP1A1 and CACNA1D

- underlie a common subtype of adrenal hypertension. *Nat. Genet.* 45, 1055–1060.
38. Zhu, X., Luke, A., Cooper, R.S., Quertermous, T., Hanis, C., Mosley, T., Gu, C.C., Tang, H., Rao, D.C., Risch, N., and Weder, A. (2005). Admixture mapping for hypertension loci with genome-scan markers. *Nat. Genet.* 37, 177–181.
  39. Zhu, X., and Cooper, R.S. (2007). Admixture mapping provides evidence of association of the VNN1 gene with hypertension. *PLoS ONE* 2, e1244.
  40. Balakumar, P., and Jagadeesh, G. (2011). Potential cross-talk between (pro)renin receptors and Wnt/frizzled receptors in cardiovascular and renal disorders. *Hypertens. Res.* 34, 1161–1170.
  41. Schifano, E.D., Li, L., Christiani, D.C., and Lin, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *Am. J. Hum. Genet.* 92, 744–759.
  42. Lin, D.Y., and Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 97, 321–332.