HHS Public Access

Author manuscript

Curr Hypertens Rep. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Curr Hypertens Rep. 2017 March ; 19(3): 23. doi:10.1007/s11906-017-0718-1.

A review of the genetics of hypertension with a focus on gene-environment interactions

RJ Waken[1,a], **Lisa de las Fuentes**[1,2], and **D.C. Rao**[1]

[1] Division of Biostatistics, Washington University in St. Louis, School of Medicine, 660 S. Euclid Ave, Campus Box 8067, St. Louis, MO, 63110, USA

[2] Division of Cardiology, Department of Medicine, 660 S. Euclid Ave, Campus Box 8086, St. Louis, MO, 63110, USA

Abstract

Purpose of review—Here, we discuss the interpretation and modeling of gene-environment interactions in hypertension related phenotypes, with a focus on the necessary assumptions and possible challenges.

Recent findings—Recently, small cohort studies have discovered several novel genetic variants associated with hypertension-related phenotypes through modeling gene-environment interactions. Several consortia-based meta-analytic efforts have uncovered many novel genetic variants in hypertension without modeling interaction terms, giving promise to future meta-analytic efforts that incorporate gene-environment interactions.

Summary—Heritability studies and genome-wide association studies have established that hypertension, a prevalent cardiovascular disease, has a genetic component that may be modulated by the environment (such as lifestyle factors). This review includes a discussion of known genetic associations for hypertension/blood pressure, including those resulting from the incorporation of gene-environmental interaction modeling.

1 Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide, and the burden is expected to increase in the coming years [1]. Further, CVD is considered to be largely preventable, and while prevention and mitigation efforts are leading to declines in CVD mortality in high income countries, the incidence of and mortality from CVD continues to increase in low- to middle-income countries [2]. In part, increasing rates of CVD mortality in low- to middle-income countries is a function of longer life expectancy resulting from decreasing prevalence of infectious disease [3]. This, combined with high smoking rates [4], increasing obesity rates, and relatively poor access to treatment [3], combine to increase CVD mortality in low- to middle-income countries. Hypertension (HTN) is characterized by high systolic blood pressure (SBP) and/or diastolic blood pressure (DBP), and is a major causal factor in heart failure, stroke, renal disease, and cardiovascular death [5]. A seminal study in BP research characterized the relationships between blood pressure and risk of

[a]Corresponding Author: Attn: RJ Waken, rjwaken@wustl.edu, Phone number: 314-273-1222.

CVD mortality and stroke as continuous, consistent, and independent of other risk factors; between the ages 40–69 years, CVD risk doubles for every 20 mmHg increase in SBP or 10 mmHg increase in DBP above 115/76 mmHg [6]. For 2015, the American Heart Association reported that high blood pressure (BP) was the leading factor in 40.6% of CVD and 34.7% ischemic heart disease attributable deaths in the US [7]. HTN incidence has been shown to differ among ancestral groups, as evidenced by comparisons between Caucasian and Hispanic American as well as Caucasian and African American groups [8, 9]. Between 2003 and 2013, the overall death rate attributable to HTN in the United States increased 14.4% and 1.7% in non-Hispanic Caucasians and Hispanics respectively, and declined 9.1%, and non-Hispanic African Americans [7].

CVD, as most human diseases, arises from a complex interaction between the environmental and genetic factors. Genome-wide association studies (GWAS) have identified a large number of genetic factors in a variety of diseases, including HTN, as reviewed by Zheng et al. [10]. In BP, family and twin studies have yielded heritability estimates in the ranges of 48% to 60% (SBP) and 34% to 67% (DBP) [11, 12, 13]. As of 2011, identified genetic loci from GWAS accounted for less than 2.5% of the phenotypic variance in BP studies [14], which has modestly increased to ≈3.5% with more recent GWAS [15, 16, 17]. Overestimation of heritability may occur in twin studies due to violations of shared environment assumptions, and poor phenotyping practices in control cohorts, failure to account for epistasis, GxE interactions, and other non-genetic sources of phenotype modulation may lead to underestimation of heritability in GWAS, all of which may contribute to the large difference in heritability estimates. Nonetheless, it is widely assumed that many of the causal loci are as of yet unidentified, a phenomenon commonly referred to as “missing heritability” [18, 19, 20]. Whereas GWAS have primarily focused on identifying genetic factors through modeling main effects of common SNPs, a focus on modeling gene-environment (GxE) and gene-gene (GxG) interaction may identify additional causal factors in human disease [18]. BP is known to be modulated by a variety of lifestyle factors, such as diet [21], exercise [22], and smoking [23], as well as other factors such as age [24, 25] and obesity [26]. In GWAS for HTN-related phenotypes, significant GxE interactions have been identified with alcohol consumption [27], body mass index (BMI) [28], smoking [29, 30], education levels [31], and sodium intake [46] which are all modifiable through lifestyle changes. These findings suggest that further investigation into GxE interaction may identify causal genetic loci that contribute to missing heritability [33].

The remainder of this review is organized as follows: Section 2 presents a discussion of interaction, including common definitions of interaction, rationales for modeling interaction, and limitations in modeling GxE interaction. Section 3 briefly reviews selected methodology employing interaction that has been used to identify genetic factors in HTN and HTN-related phenotypes. Section 4 summarizes the genetic factors in HTN and HTN-related phenotypes, and focuses further on associations identified using models that account for GxE interactions. Section 5 ends with a discussion of future direction for GxE studies and a conclusion.

2 Interaction and accounting for interaction through statistical modeling

Although the term “interaction” is prevalent in GWAS and the statistical literature, the definition and interpretation of interaction is contextual. Here, we discuss possible interpretations of interaction in the contexts of GWAS and statistics, then discuss how these interpretations coincide. A series of rationales is given for investigation into GxE interactions, and we follow with a commentary on the difficulties surrounding modeling GxE interaction terms in GWAS.

In an oft-cited review of GxE interaction in GWAS, Thomas [34] presents five definitions of interaction, given in Table 1. Of the five, biological interactions are the assumed to account for a portion of the missing heritability. Using a more general definition, biological interaction occurs when the effect of one factor on the process driving a phenotypic response is affected by the presence or magnitude of another factor [24]. Qualitative and quantitative interactions are forms of statistical interaction, as mentioned by Thomas [34], and public health synergy is collapsible into statistical interaction using the definition of statistical interaction given by Cox [35]. In all cases where statistical interaction is modeled, it should be assumed that biological interaction exists, even if the biological mechanism is not understood. Ottman [36] gives five plausible models for environmental and genetic factors to cooperatively influence disease etiology, of which four could be appropriately modeled by the definition of statistical interaction given above. Un-modeled interactions, in the case of GxE, GxG, or gene-covariate interactions, may contribute to confounding if left unincorporated. In general, the statistical literature tends to focus on modeling interaction terms as a nuisance, while the genetic epidemiology literature models interaction to increase power in finding disease relevant genetic loci [33]. In the context of statistical analysis, erring on the side of overfitting rather than underfitting with respect to interaction terms is recommended when interaction is believed to exist [35]. In cases where GxE interaction is biologically feasible, modeling these interactions can aid in accounting for departures from independence between the genetic and environmental factors of interest.

Hunter [37] gives a series of rationales for studying GxE interactions in GWAS, which is presented in its entirety in Table 2. Thomas [34] also describes the goal of incorporating GxE interaction terms in GWAS is to *use* interactions – not discover interactions, but *use* interactions – to “discover novel genes that act synergistically with other factors” in modeling disease etiology, which intersects with the first of the rationales given by Hunter [37] in Table 2. The resulting findings from modeling GxE interactions in GWAS focus on presenting results in light of the remaining four rationales. The genetic associations discovered in GWAS are often used to drive bioinformatics research in finding the biological pathways mentioned in rationales three and four. In cases where the biological pathways for genes associated with GWAS identified genetic loci are understood, findings are presented in light of the processes underlying the biological pathways. Incorporating the interaction term into modeling can make finding a sufficiently explanatory relationship between the phenotypic response and model possible. Ultimately, the goal of GWAS is to drive personalized medicine, which is reflected in the fifth and final rationale presented.

Appropriately modeling interaction from a statistical standpoint is difficult, and modeling interaction in the context of GWAS presents additional complexity. Much of the difficulty in appropriately modeling interaction in GWAS comes in light of the generally small sample sizes relative to the large number of possible combinations of factors in observational studies [38]. Further exacerbating this problem is the possibility that minor allele frequencies are small, that the disease prevalence is low, or both [33, 38]. Confounding issues stemming from failing to model non-GxE interactions may arise, and will be discussed briefly in the conclusion. Outside of practical modeling concerns, some researchers believe that modeling interaction in the GxE or GxG context is inherently problematic. Some “environmental” factors, such as alcohol intake in studies concerning alcohol metabolism, may be driven by the causal gene in question and should not be used as covariates [39]. This suggests that additional factors related to the phenotype of interest, especially genetic factors that could be considered intrinsic, may be better incorporated through methodologies that model pleiotropy [40]. Furthermore, because the genotype frequency of the marker found in linkage disequilibrium and the true causal locus likely differ, the imputed genotype frequency at an identified marker is not an unbiased estimator of the frequency at the causal locus, leading to unknown misclassification probabilities [40]. Misclassification or measurement error for environmental and nuisance factors may also exist in the data. While it has been shown analytically [41] and through simulation [42] that misclassification in environmental factors biases the interaction term toward zero, these studies assume independence in the prevalence of genetic and environmental factors in the population, and results may not hold when this assumption is violated [40].

3 Common methodologies used to account for interaction in GWAS modeling

Although the GWAS literature that do not account for interaction is diverse, these methodologies are well reviewed and do not fit the scope of this writing. The literature concerning the modeling of interaction in GWAS is also diverse, but we focus on methods that appear to be successful in uncovering genome-wide significant (GWS) SNP associations in BP-related traits. More specifically, we compare the assumptions, perceived notions of power and type I error rates, and discuss how violations of these assumptions may affect type I error rates. For all tests discussed, researchers mitigate multiple testing issues and control type I error rates with a restrictive definition of GWS by setting the target type I error rate at $\alpha=5\times10^{-8}$., and studies often further control for false positives by accounting for population substructures through genomic control methods [43].

3.1 Tests involving a single cohort, or multiple cohorts with fully available data

Applications of the general linear model, or GLM, are appropriate to test for association between genomic markers and the phenotype of interest. In analyses where complete data for all individuals is readily available, a GLM can be fit to the full data. These tests can account for any confounding factor that may arise through modeling of nuisance factors. Unfortunately, these tests are limited by the complexities that arise from analyzing massive cohorts, and require full access to raw data from all cohorts, which is often unattainable. While a number of tests could arise from the use of GLMs in the GWAS GxE literature, we

focus on the joint 2 degree of freedom likelihood ratio test (referred to as J2DF) and the Wald type 1 degree of freedom interaction effect test (referred to as 1DF). In applying a GLM, assumptions are imposed on the error structure of the resulting model. Typically, population substructures within the errors resulting from a GLM fit will inflate type I error rates [43]. For a more formal review of the mechanics surrounding these hypothesis tests and test statistics, see Appendix A.1.

3.2 Meta-analytic tests involving multiple cohorts

Studies concerning the detection of causal genetic loci are demanding regardless of the inclusion of interaction effects, which further compounds the issue [44]. In addition, fitting GLMs to datasets large enough to allow for the discovery of significant interaction terms may be computationally difficult, and in many cases, researchers do not have access to full cohort data. Meta-analytic procedures allow researchers to incorporate results from study-specific analyses to implement more powerful tests than what could be achieved through the use of a single cohort analysis; this does not require access to the individual level data used in each of the contributing cohorts.

In meta-analysis concerning synthesis of uncorrelated regression slopes, inverse variance weighted least squares approaches are common. Ordinary and weighted least squares approaches are inappropriate to model correlated regression slopes and may not result in consistent estimates. Becker and Wu [45] present a robust generalized least squares meta-analytic approach to estimate possibly correlated population regression coefficients given summaries of study-specific analyses. This technique was later applied to meta-analysis of GWAS results [33]. As with all meta-analytic approaches, we must assume that the results from the k studies incorporated are representative of the true processes in the population of interest, and are not pruned to the most significant results via publication bias. This assumption is somewhat safe in the realm of genetic epidemiology, where the aim of the meta-analysis is to bolster power through the incorporation of results from multiple cohorts performing agnostic GWAS. This methodology also assumes that either a) each of the k studies considered accounted for the same nuisance parameters, or b) that although the k studies did not incorporate the same nuisance parameters, the resulting parameter estimates are unbiased representations of the true slopes of interest. The validity of the latter assumption is generally corroborated, so long as modeling techniques in the individual studies are the same; studies have found little to no evidence of bias in the synthesized slopes of interest as a result of the estimation of additional nuisance parameters in the considered studies [46, 47], while analyses synthesizing results from different models show that the type of model, as well as the difference in nuisance parameters, will affect the meta-analytic slope estimates [48]. We focus on the joint Wald type 2 degree of freedom test (referred to as JMA2DF), which tests the genetic main and GxE interaction effects, and the Wald type 1 degree of freedom interaction effect test (referred to as MA1DF). For a more formal review of the mechanics surrounding these hypothesis tests and test statistics, see Appendix A.2.

4 Significant findings in the HTN literature

Although the above is primarily focused on incorporation of GxE interaction terms into models evaluating the effect of allelic presence at genetic loci of interest, major findings concerning causal genetic loci have been accomplished without modeling GxE interaction terms. Online Resource 1 gives a summary of GWS findings from studies that did not incorporate GxE interaction, including the gene, cytogenetic or molecular location, associated phenotype, lead/tag SNP (a representative SNP in an area of the genome that exhibits high linkage disequilibrium), and the ancestral groups. A short summary of all HTN GWAS GWS findings is given in Table 3.

4.1 Findings relying on the incorporation of GxE interactions

A set of GWS findings from models incorporating interactions is given in Online Resource 2. Notably, this table is somewhat smaller, suggesting that this is a more recent direction of research which needs more effort. Also, some of the identified loci that were not novel may have been observed in studies that did not include interactions. It is notable that when findings of association were also detected in studies accounting for GxE interactions, these models lend great power to uncovering possibly causal genomic variants. An example of increased power from the inclusion of GxE terms in GWAS can be seen through the analysis of 6,889 participants from the Framingham Heart Study cohort by Sung et al. [29]. In Table 4, the findings overlap at the MECOM gene are displayed as well as cohort information for the studies involved. Although these findings have shortcomings, recall that our purpose is not to detect interaction, but rather detect novel loci by accounting for interaction, similar to the arguments made by Cox [35] and Thomas [34]. In this context, the interaction can be thought of more as a nuisance parameter, much like the covariates modeled.

5 Conclusion and future direction

We have reviewed the definition, rationale, and interpretation of GxE interaction in the context of GWAS. The discussion includes limitations in select methodology, and relevant GWS findings within the HTN GWAS literature. Accounting for GxE interaction in GWAS can aid in the detection of new GWS genetic loci [34], which is shown through the examples in Online Resource 2 and Table 3. Additionally, although the mechanistic processes underlying GxE interactions is unknown, their existence is well accepted and studied [36], which gives credibility to the incorporation of GxE interaction terms in GWAS.

It is commonly known that even in the absence of a true effect, increases in the study sample size n for a given study leads to more significant p-values, which may cause a researcher to declare spurious results statistically significant. Through simulation, VanderWeele et al. [79] show that this phenomenon is more pronounced at smaller sample sizes as the effect size of an uncontrolled confounder increases. Keller [80] attributes this type I error rate inflation to either uncontrolled confounding involving environmental and/or un-modeled population substructures in GxE GWAS studies or improper model specification, adding that, of 45 reviewed studies with GWS findings in the GxE schizophrenia literature, only 12 were thus far replicated. Further, in situations where measurement error or misclassification exists in the genetic, phenotype, and/or environmental covariates, effect estimates may be

inconsistent, resulting in lower power, or bias, which may increase type I error rates [81]. In light of the previous discussion regarding the assumption of a true biological interaction between the genetic factor and environmental covariates of interest, note that we implicitly assume independence of all possible un-modeled covariates and factors, interactions included. As meta-analytic approaches become more prevalent, increasing the sample size and power of GxE GWAS, this assumption will need to be given more attention to avoid inflating type I error rates. While misclassification of environmental factors has been shown to bias the interaction term toward zero, thus not contributing to type I error rate inflation [41, 42] these analyses assume independence of the genetic and environmental factors in the population conditional on the phenotype of interest, and these results may not hold when this independence assumption is violated [40]. Although research has not specifically been conducted to study type I error rates as a function of minor allele frequencies in GxE studies, inclusion of genetic variants with small minor allele frequencies is uncommon due to error rate inflation concerns [82]. While these concerns are echoed throughout the literature, it has been shown through simulation that inclusion of genetic loci with small minor allele frequencies does not increase type I error rates above expected levels [83].

Incorporation of GxE interaction terms into GWAS may enhance statistical power without increasing sample sizes by reducing error variance, which may facilitate personalized medicine approaches in healthcare and may help in the development of pharmaceutical treatment for complex diseases. Although arguments against interaction studies exist, replication in additional cohorts and biological explanations of gene expression can mitigate spurious findings. Investigation of pleiotropy and epistasis can yield further insights into the genetic underpinnings of complex diseases, but require much larger sample sizes. As the cost of sequencing decreases, these approaches may shed light on genetic variants that either work in parallel to influence disease, or contribute to multiple diseases.

Added in proof: As of this revision, Warren et al. [84] have identified an additional 32 novel blood pressure loci, and further validated 75 loci discovered by Hoffman et al. [55].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012; 380(9859):2095–2128. http://dx.doi.org/10.1016/S0140-6736(12)61728-0.
2. Mendis, S., Puska, P., Norrving, B. Global Atlas on Cardiovascular Disease Prevention and Control. World Health Organization; Geneva: 2011.
3. Gaziano TA, Bitton A, Anand S, et al. Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries. Current problems in cardiology. 2010; 35(2):72–115. DOI: 10.1016/j.cpcardiol.2009.10.002 [PubMed: 20109979]
4. Tobacco killing in low-income and middle-income countries. The Lancet. 2012; 379(9822):1172. http://dx.doi.org/10.1016/S0140-6736(12)60492-9.
5. James PA, Oparil S, Carter BL, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA. 2014; 311:507–20. [PubMed: 24352797]

6. Lewington S, Clarke R, Qizilbash N, et al. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002; 360:1903–13. [PubMed: 12493255]

7. Mozaffarian D, Benjamin EJ, Go AS, et al. American Heart Association Statistics C, Stroke Statistics S. Heart disease and stroke statistics-2016 update: A report from the American Heart Association. Circulation. 2016; 133:e38–e360. [PubMed: 26673558]

8. Haffner SM, Mitchell BD, Valdez RA, et al. Eight-Year Incidence of Hypertension in Mexican-Americans and Non-Hispanic Whites: The San Antonio Heart Study. Am J Hypertension. 1992; 5:147–153.

9. Snieder H, Harshfield GA, Treiber FA. Heritability of Blood Pressure and Hemodynamics in African and European American Youth. Hypertension. 2003; 41:1196–1201. [PubMed: 12719445]

10. Zheng J, Rao DC, Shi G. An update on genome-wide association studies of hypertension. Appl Inform. 2015; 2:10.doi: 10.1186/s40535-015-0013-7

11. Hottenga JJ, Boomsma DI, Kupper N, et al. Heritability and stability of resting blood pressure. Twin Res Hum Genet. 2005; 8:499–508. [PubMed: 16212839]

12. Levy D, Larson MG, Benjamin EJ, et al. Framingham Heart Study 100k Project: genome-wide associations for blood pressure and arterial stiffness. BMC Med Genetics. 2007; 8:S3. [PubMed: 17903302]

13. Miall WE, Oldham PD. The Hereditary factor in arterial blood pressure. BMJ. 1963; 1:75–80. [PubMed: 13935402]

14. Fox ER, Young JH, Li Y, et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. Hum Mol Genet. 2011; 20(11):2273–2284. DOI: 10.1093/hmg/ddr092 [PubMed: 21378095]

15*. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342415 individuals. Nature Genetics. 2016; 48(10):1171–1184. http://dx.doi.org/10.1038/ng.3667This paper is one of four recent genome wide association studies to uncover a large number of genetic loci associated with BP measurements through meta-analysis, contributing 66 newly discovered and validated genetic variants. [PubMed: 27618452]

16*. Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nature Genetics. 2016; 48(10):1151–1161. http://dx.doi.org/10.1038/ng.3654This paper is one of four recent genome wide association studies to uncover a large number of genetic loci associated with BP measurements through meta-analysis, contributing 49 newly discovered and validated genetic variants. [PubMed: 27618447]

17*. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nature Genetics. 2016; 48(10):1162–1170. http://dx.doi.org/10.1038/ng.3660This paper is one of four recent genome wide association studies to uncover a large number of genetic loci associated with BP measurements through meta-analysis, contributing 21 newly discovered and validated genetic variants. [PubMed: 27618448]

18**. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–753. This paper discusses the missing heritability issue, and gives strategies (one of which being the investigation of gene-environment interaction) for addressing this issue. DOI: 10.1038/nature08494 [PubMed: 19812666]

19. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008; 456:18–21. [PubMed: 18987709]

20. Zhang K, Weder AB, Eskin E, O'Connor DT. Genome-wide case/control studies in hypertension: only the 'tip of the iceberg'. J Hypertens. 2010; 28:1115–23. [PubMed: 20216088]

21. Zhao D, Qi Y, Zheng Z, et al. Dietary factors associated with hypertension. Nat Rev Cardiol. 2011 Jul 5; 8(8):456–65. DOI: 10.1038/nrcardio.2011.75 [PubMed: 21727918]

22. Boutcher YN, Boutcher SH. Exercise intensity and hypertension: what's new? Journal of Human Hypertension. 2016; doi: 10.1038/jhh.2016.62

23. Doonan RJ, Hausvater A, Scallan C, et al. The effect of smoking on arterial stiffness. Hypertens Res. 2010; 33(5):398–410. DOI: 10.1038/hr.2010.25 [PubMed: 20379189]
24. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. Int J Epidemiol. 1981; 10:383–387. [PubMed: 7327838]
25. Simino J, Shi G, Bis JC, et al. Gene– age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. Am J Hum Genet. 2014; 95(1):24–38. DOI: 10.1016/j.ajhg.2014.05.010 [PubMed: 24954895]
26. Pausova Z. From big fat cells to high blood pressure: a pathway to obesity-associated hypertension. Curr Opin Nephrol Hypertens. 2006; 15(2):173–178. [PubMed: 16481885]
27. Simino J, Sung YJ, Kume R, et al. Gene-alcohol interactions identify several novel blood pressure loci including a promising locus nearSLC16A9. Frontiers in Genetics. 2013; 4:277.doi: 10.3389/fgene.2013.00277 [PubMed: 24376456]
28. Kim YK, Kim Y, Hwang MY, et al. Identification of a genetic variant at 2q12.1 associated with blood pressure in East-Asians by genome-wide scan including gene-environment interactions. BMC Medical Genetics. 2014; 15:65.doi: 10.1186/1471-2350-15-65 [PubMed: 24903457]
29. Sung YJ, de Las Fuentes L, Schwander KL, et al. Gene–Smoking interactions identify several novel blood pressure loci in the Framingham heart study. Am J Hypertens. 2014; 28(3):343–354. DOI: 10.1093/ajh/hpu149 [PubMed: 25189868]
30. Basson J, Sung YJ, de las Fuentes L, et al. Influence of smoking status and intensity on discovery of blood pressure loci through gene-smoking interactions. Genetic epidemiology. 2015; 39(6):480–488. DOI: 10.1002/gepi.21904 [PubMed: 25940791]
31. Basson J, Sung YJ, Schwander K, et al. Gene–Education Interactions Identify Novel Blood Pressure Loci in the Framingham Heart Study. American Journal of Hypertension. 2014; 27(3): 431–444. DOI: 10.1093/ajh/hpt283 [PubMed: 24473254]
32**. Li C, He J, Chen J, et al. Genome-Wide Gene-Sodium Interaction Analyses on Blood Pressure: The Genetic Epidemiology Network of Salt-Sensitivity Study. Hypertension. 2016 Aug; 68(2): 348–55. This paper identifies three genome-wide significant loci (and five loci with promising effect sizes) in a small cohort investigating gene-environment interactions associated with blood pressure. DOI: 10.1161/HYPERTENSIONAHA.115.06765 [PubMed: 27271309]
33*. Manning AK, LaValley M, Liu C-T, et al. Meta-analysis of Gene-Environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. Genetic Epidemiology. 2011; 35(1):11–18. This paper lays the foundation for meta-analytic procedures for genome-wide association studies to investigate gene-environment interactions with a working application in type 2 diabetes research. DOI: 10.1002/gepi.20546 [PubMed: 21181894]
34. Thomas D. Gene-Environment-Wide Association Studies: Emerging Approaches. Nature reviews Genetics. 2010; 11(4):259–272. DOI: 10.1038/nrg2764
35. Cox D. Interaction. International Statistical Review/Revue Internationale De Statistique. 1984; 52(1):1–24. http://www.jstor.org/stable/1403235.
36. Ottman R. An Epidemiologic Approach to Gene-Environment Interaction. Genetic epidemiology. 1990; 7(3):177.doi: 10.1002/gepi.1370070302 [PubMed: 2369997]
37. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6:287–98. DOI: 10.1038/nrg1578 [PubMed: 15803198]
38. Engelman CD, Baurley JW, Chiu Y-F, et al. Detecting Gene-Environment Interactions in Genome-Wide Association Data. Genetic Epidemiology. 2009; 33(Suppl 1):68–73. DOI: 10.1002/gepi.20475
39. Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. Nature Reviews Genetics. 2005; 6:521–532.
40. Dudbridge F, Fletcher O. Gene-Environment Dependence Creates Spurious Gene-Environment Interaction. American Journal of Human Genetics. 2014; 95(3):301–307. DOI: 10.1016/j.ajhg.2014.07.014 [PubMed: 25152454]
41. García-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene-environment interactions in case-control studies. Am J Epidemiol. 1998; 1475:426–433.
42. Hein R, Beckmann L, Chang-Claude J. Sample size requirements for indirect association studies of gene-environment interactions (G × E). Genet Epidemiol. 2008; 32:235–245. [PubMed: 18163529]

43. Devlin B, Roeder K. Genomic Control for Association Studies. Biometrics. 1999; 55(4):997–1004. DOI: 10.1111/j.0006-341X.1999.00997.x [PubMed: 11315092]
44. Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. Am J Epidemiol. 2006; 164:609–614. [PubMed: 16893921]
45. Becker BJ, Wu M-J. The Synthesis of Regression Slopes in Meta-Analysis. Statistical Science. 2007; 22(3):414–429. http://projecteuclid.org/euclid.ss/1199285041. DOI: 10.1214/07-STS243
46. Ashenfelter O, Harmon C, Oosterbeek H. A review of estimates of the schooling/earnings relationship, with tests for publication bias. National Bureau of Economic Research Working Paper Series. 2000; :7457.doi: 10.3386/w7457
47. Peterson RA, Brown SP. On the use of beta coefficients in meta-analysis. J Appl Psychol. 2005; 90:175–181. [PubMed: 15641898]
48. Doucouliagos H, Paldam M. Aid Effectiveness on Accumulation: A Meta Study. Kyklos. 2006; 59(2):227–254. DOI: 10.1111/j.1467-6435.2006.00326.x
49. Adeyemo A, Gerry N, Chen G, et al. A genome-wide association study of hypertension and blood pressure in African Americans. PLoS Genet. 2009; 5(7):e1000564.doi: 10.1371/journal.pgen.1000564 [PubMed: 19609347]
50. Cho YS, Go MJ, Kim YJ, et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet. 2009; 41(5):527–34. DOI: 10.1038/ng.357 [PubMed: 19396169]
51. Franceschini N, Fox E, Zhang Z, et al. 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. 2013; 93(3):545–554. DOI: 10.1016/j.ajhg.2013.07.010 [PubMed: 23972371]
52. Ganesh SK, Tragante V, Guo W, et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum Mol Genet. 2013; 22(8):1663–1678. DOI: 10.1093/hmg/dds555 [PubMed: 23303523]
53. Ganesh SK, Chasman DI, Larson MG, et al. Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. Am J Hum Genet. 2014; 95(1):49–65. DOI: 10.1016/j.ajhg.2014.06.002 [PubMed: 24975945]
54. Ho JE, Levy D, Rose L, et al. Discovery and replication of novel blood pressure genetic loci in the Women's Genome Health Study. J Hypertens. 2011; 29(1):62–69. DOI: 10.1097/HJH.0b013e3283406927 [PubMed: 21045733]
55. Hoffman TJ, Ehret GB, Nandakuma P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet. 2017; 47(1):54–64. This paper is one of four recent genome wide association studies to uncover a large number of genetic loci associated with BP measurements through meta-analysis, contributing 316 newly discovered loci. DOI: 10.1038/ng.3715
56. Hong KW, Jin HS, Lim JE, et al. Recapitulation of two genomewide association studies on blood pressure and essential hypertension in the Korean population. J Hum Genet. 2010; 55(6):336–341. DOI: 10.1038/jhg.2010.31 [PubMed: 20414254]
57. Hong KW, Min H, Heo BM, et al. Recapitulation of genome-wide association studies on pulse pressure and mean arterial pressure in the Korean population. J Hum Genet. 2012; 57(6):391–393. DOI: 10.1038/jhg.2012.31 [PubMed: 22475680]
58. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011; 478(7367):103–109. DOI: 10.1038/nature10405 [PubMed: 21909115]
59. Johnson AD, Newton-Cheh C, Chasman DI, et al. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. Hypertension. 2011a; 57(5):903–910. DOI: 10.1161/HYPERTENSIONAHA.110.158667 [PubMed: 21444836]
60. Johnson T, Gaunt TR, Newhouse SJ, et al. Blood pressure loci identified with a gene-centric array. Am J Hum Genet. 2011b; 89(6):688–700. DOI: 10.1016/j.ajhg.2011.10.013 [PubMed: 22100073]

61. Kato N, Takeuchi F, Tabara Y, et al. Meta- analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat Genet. 2011; 43(6): 531–538. DOI: 10.1038/ng.834 [PubMed: 21572416]
62. Kelly TN, Takeuchi F, Tabara Y, et al. Genome-wide association study meta-analysis reveals transethnic replication of mean arterial and pulse pressure loci. Hypertension. 2013; 62(5):853–859. DOI: 10.1161/HYPERTENSIONAHA.113.01148 [PubMed: 24001895]
63. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009; 41(6):677–687. DOI: 10.1038/ng.384 [PubMed: 19430479]
64. Lin Y, Lai X, Chen B, et al. Genetic variations in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in She ethnic minority of China. Atherosclerosis. 2011; 219(2):709–714. DOI: 10.1016/j.atherosclerosis.2011.09.006 [PubMed: 21963141]
65. Lu X, Wang L, Lin X, et al. Genome- wide association study in Chinese identifies novel loci for blood pressure and hypertension. Hum Mol Genet. 2014; 24(3):865–874. DOI: 10.1093/hmg/ddu478 [PubMed: 25249183]
66. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet. 2009; 41(6):666–676. DOI: 10.1038/ng.361 [PubMed: 19430483]
67. Org E, Eyheramendy S, Juhanson P, et al. Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. Hum Mol Genet. 2009; 18(12):2288–2296. DOI: 10.1093/hmg/ddp135 [PubMed: 19304780]
68. Padmanabhan S, Melander O, Johnson T, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. PLoS Genet. 2010; 6(10):e1001177.doi: 10.1371/journal.pgen.1001177 [PubMed: 21082022]
69. Qi Y, Zhao H, Wang Y, et al. Replication of the top 10 most significant polymorphisms from a large blood pressure genome-wide association study of northeastern Han Chinese East Asians. Hypertens Res. 2014; 37(2):134–138. DOI: 10.1038/hr.2013.132 [PubMed: 24196197]
70. Salvi E, Kutalik Z, Glorioso N, et al. Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. Hypertension. 2012; 59(2): 248–255. DOI: 10.1161/HYPERTENSIONAHA.111.181990 [PubMed: 22184326]
71. Tabara Y, Kohara K, Kita Y, et al. Common variants in the ATP2B1 gene are associated with susceptibility to hypertension: the Japanese Millennium Genome Project. Hypertension. 2010; 56(5):973–980. DOI: 10.1161/HYPERTENSIONAHA.110.153429 [PubMed: 20921432]
72. Takeuchi F, Isono M, Katsuya T, et al. Blood pressure and hypertension are associated with 7 loci in the Japanese population. Circulation. 2010; 121(21):2302–2309. DOI: 10.1161/CIRCULATIONAHA.109.904664 [PubMed: 20479155]
73. Tomaszewski M, Debiec R, Braund PS, et al. Genetic architecture of ambulatory blood pressure in the general population: insights from cardiovascular gene-centric array. Hypertension. 2010; 56(6): 1069–1076. DOI: 10.1161/HYPERTENSIONAHA.110.155721 [PubMed: 21060006]
74. Tragante V, Barnes MR, Ganesh SK, et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood- pressure-related loci. Am J Hum Genet. 2014; 94(3): 349–360. DOI: 10.1016/j.ajhg.2013.12.016 [PubMed: 24560520]
75. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nat Genet. 2011; 43(10):1005–1011. DOI: 10.1038/ng.922 [PubMed: 21909110]
76. Wang Y, O'Connell JR, McArdle PF, et al. Whole-genome association study identifies STK39 as a hypertension susceptibility gene. Proc Natl Acad Sci USA. 2009; 106(1):226–231. DOI: 10.1073/pnas.0808358106 [PubMed: 19114657]
77. Wang Y, Zhang Y, Li Y, et al. Common variants in the ATP2B1 gene are associated with hypertension and arterial stiffness in Chinese population. Mol Biol Rep. 2013; 40(2):1867–1873. DOI: 10.1007/s11033-012-2242-3 [PubMed: 23079715]

78. Zhu X, Young JH, Fox E, et al. Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. Hum Mol Genet. 2011; 20(11):2285–2295. DOI: 10.1093/hmg/ddr113 [PubMed: 21422096]

79. VanderWeele TJ, Ko Y-A, Mukherjee B. Environmental Confounding in Gene-Environment Interaction Studies. American Journal of Epidemiology. 2013; 178(1):144–152. DOI: 10.1093/aje/kws439 [PubMed: 23821317]

80. Keller MC. Gene-by-environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. Biological psychiatry. 2014; 75(1)doi: 10.1016/j.biopsych.2013.09.006

81. Gordon D, Finch SJ. Factors affecting statistical power in the detection of genetic association. Journal of Clinical Investigation. 2005; 115(6):1408–1418. DOI: 10.1172/JCI24756 [PubMed: 15931375]

82. Lam AC, Schouten M, Aulchenko YS, et al. Rapid and robust association mapping of expression quantitative trait loci. BMC Proceedings. 2007; 1(Suppl 1):S144. [PubMed: 18466488]

83. Tabangin ME, Woo JG, Martin LJ. The effect of minor allele frequency on the likelihood of obtaining false positives. BMC Proceedings. 2009; 3(Suppl 7):S41.doi: 10.1186/1753-6561-3-S7-S41 [PubMed: 20018033]

84. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017; doi: 10.1038/ng.3768

85. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics. 2010; 11:134.doi: 10.1186/1471-2105-11-134 [PubMed: 20233392]

Appendix: Formulation of relevant test statistics

A.1: Tests involving individual level cohort data

Let $\boldsymbol{y}$ denote a quantitative vector containing measurements representing the phenotype of interest. We specify the GLM

$$f(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta})=\boldsymbol{X}_C\boldsymbol{\beta}_C+\boldsymbol{x}_G\beta_G+\boldsymbol{x}_E\beta_E+\boldsymbol{x}_G\odot\boldsymbol{x}_E\beta_{G\times E}+\boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{X}_C$ and $\boldsymbol{\beta}_C$ are the design matrix and regression coefficients for the nuisance factors, $\boldsymbol{x}_G$ and β_G is the covariate vector and regression coefficient for the genetic locus of interest, $\boldsymbol{x}_E$ and β_E is the covariate vector and regression coefficient for the environmental variable of interest, $\boldsymbol{x}_G\odot\boldsymbol{x}_E$ is the elementwise product of the genetic and environmental covariate vectors, $\beta_{G\times E}$ is the regression coefficient for the GxE interaction term, and $\boldsymbol{\varepsilon}$ is a multivariate normal random vector with mean 0 and covariance matrix Σ. Coding schemes for $\boldsymbol{x}_E$ depends on the study in question, as environmental factors may be discrete or continuous. $\boldsymbol{x}_G$ may be binary in the case of dominant alleles, or coded based on the number of copies of a reference allele at a locus of interest.

In analyses where all data for all individuals is readily available, (1) can be fit to the full data, usually with a set of assumptions on the distribution of $\boldsymbol{\varepsilon}$. In particular, researchers will assume independence amongst subjects, restricting the covariance matrix Σ to the diagonal case, or choose to only model within family correlation, giving Σ a block diagonal structure.

In some cases, researchers choose to isolate the genetic and environmental factors of interest portrayed in (1) through a two step modeling process. Let

$$f(\boldsymbol{y}|\boldsymbol{X}_C,\beta_C)=\boldsymbol{X}_C\beta_C+\boldsymbol{z},$$

where $\boldsymbol{X}_C$ and $\boldsymbol{\beta}_C$ are the design matrix and regression coefficients for the nuisance factors. Then, the model

$$f(\boldsymbol{z}|\boldsymbol{x}_G,\boldsymbol{x}_E,\beta_G,\beta_E,\beta_{G\times E})=\boldsymbol{x}_G\beta_G+\boldsymbol{x}_E\beta_E+\boldsymbol{x}_G\odot\boldsymbol{x}_E\beta_{G\times E}+\boldsymbol{\varepsilon} \quad (2)$$

is fit, where $\boldsymbol{x}_G$, $\boldsymbol{x}_E$, $\boldsymbol{\beta}_G$, $\boldsymbol{\beta}_E$, $\boldsymbol{\beta}_{G\times E}$ and $\boldsymbol{\varepsilon}$ are defined as in (1). While the model in (2) assumes that $\boldsymbol{x}_G$, $\boldsymbol{x}_E$, $\boldsymbol{\beta}_G$, $\boldsymbol{\beta}_E$, $\boldsymbol{\beta}_{G\times E}|\boldsymbol{z}$ is independent of $\boldsymbol{y}$, which is likely unrealistic, (2) fits the coefficients of interest much more quickly than one through only fitting $\boldsymbol{\beta}_C$ once.

In both (1) and (2), parameter estimates $\hat{\boldsymbol{\beta}}_G$, $\hat{\boldsymbol{\beta}}_E$, and $\hat{\boldsymbol{\beta}}_{G\times E}$, along with parameter variance estimates $var(\hat{\boldsymbol{\beta}}_G)$, $var(\hat{\boldsymbol{\beta}}_E)$ and $var(\hat{\boldsymbol{\beta}}_{G\times E})$ are obtained through maximum likelihood methods. For the Wald type hypothesis test

$$H_0:\beta_{G\times E}=0$$
$$H_A:\beta_{G\times E}\neq 0$$

the test statistic

$$T_{G\times E}=\frac{\left(\hat{\beta}_{G\times E}\right)^2}{var(\hat{\beta}_{G\times E})}$$

asymptotically follows a $\chi^2_1(0)$ distribution under the null hypothesis. This is commonly referred to as the one degree of freedom interaction test, shortened to 1DF here. The ProbABEL software [85] is commonly used to carry out this test.

Of course, given the discussion above regarding accounting for interaction rather than modeling it, the researcher may be more interested in tests that incorporate interaction into the whole model. Let $L(\boldsymbol{\beta}_G, \boldsymbol{\beta}_{G\times E}|Data)$ be the profile likelihood for the model proposed in (1) or (2). To test the hypothesis

$$H_0:\beta_G,\beta_{G\times E}=0$$
$$H_A:\beta_G \text{ or } \beta_{G\times E}\neq 0,$$

we perform a likelihood ratio test, whose test statistic

$$LRT = -2\log\left(\frac{L\left(\beta_G = \hat{\beta}_G, \beta_{G\times E} = \hat{\beta}_{G\times E} | Data\right)}{L(\beta_G = 0, \beta_{G\times E} = 0 | Data)}\right)$$

asymptotically follows a $\chi^2_2(0)$ distribution under the null hypothesis. This is commonly referred to as the joint two degree of freedom test, shortened to J2DF here. The ProbABEL software [85] may be used to carry this test out as well.

A.2: Meta-analytic approaches

Let $\hat{\beta}_{G,i}$ and $\hat{\beta}_{G\times E,i}$ denote the regression coefficients associated with the genetic locus and gene by interaction effects estimated in the ith study in consideration. Then, for studies $i = 1, \ldots, k$,

$$\boldsymbol{b} = \begin{bmatrix} \hat{\beta}_{G,1} \\ \hat{\beta}_{G\times E,1} \\ \vdots \\ \hat{\beta}_{G,k} \\ \hat{\beta}_{G\times E,k} \end{bmatrix}, \hat{\sum} = \begin{bmatrix} \hat{\sum}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \hat{\sum}_k \end{bmatrix},$$

where $\hat{\boldsymbol{\Sigma}}_i = cov(\hat{\beta}_{G,i}, \hat{\beta}_{G\times E,i})$ is the covariance matrix for the regression coefficients from the ith study. We write the vector $\boldsymbol{b}$ as a linear combination of the parameters of interest β^I_G and $\beta^I_{G\times E}$

$$b = \boldsymbol{W}\boldsymbol{\beta}^I + \boldsymbol{e} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta^I_G \\ \beta^I_{G\times E} \end{bmatrix} + \boldsymbol{e},$$

where $\boldsymbol{e}$ has a multivariate normal distribution with zero mean vector and covariance matrix $\hat{\boldsymbol{\Sigma}}$. Using the generalized estimating equations as presented by Becker and Wu [45] and the imposed inverse variance weighting scheme from our definition of $\hat{\boldsymbol{\Sigma}}$, we estimate $\boldsymbol{\beta}^I$ as

$$\widehat{\boldsymbol{\beta}^I} = (\boldsymbol{W}^T \hat{\sum}^{-1} \boldsymbol{W})^{-1} \boldsymbol{W}^T \hat{\sum}^{-1} \boldsymbol{b},$$

and the covariance matrix associated with the $\boldsymbol{\beta}^I$ estimate vector as

$$Cov(\widehat{\boldsymbol{\beta}^I}) = (\boldsymbol{W}^T \hat{\sum}^{-1} \boldsymbol{W})^{-1}.$$

The Wald type hypothesis test

$$H_0: \boldsymbol{\beta}^I = \mathbf{0},$$
$$H_A: \boldsymbol{\beta}^I \neq \mathbf{0},$$

has test statistic

$$T_{\text{JMA2DF}} = \widehat{\boldsymbol{\beta}^I}^T Cov(\widehat{\boldsymbol{\beta}^I})^{-1} \widehat{\boldsymbol{\beta}^I},$$

which asymptotically follows a $\chi^2_2(0)$ distribution under the null hypothesis. Because $var(\hat{\beta}^I_{G\times E}) = Cov(\widehat{\boldsymbol{\beta}^I})_{2,2}$, we can also test for a significant interaction term under the Wald type hypothesis test

$$H_0: \beta^I_{G\times E} = 0,$$
$$H_A: \beta^I_{G\times E} \neq 0,$$

which has test statistic

$$T_{\text{MA1DF}} = \frac{\left(\hat{\beta}^I_{G\times E}\right)^2}{var(\hat{\beta}^I_{G\times E})}$$

which asymptotically follows a $\chi^2_1(0)$ distribution under the null hypothesis.

Table 1

Definitions of GxE interaction, as given by Thomas [34]

Biological interaction	An established biological relationship in which some environmental factor triggers a response in people that have a susceptible genotype.
Quantitative interaction	An interaction where effects of one factor go in the same direction at different levels of the other, but differ in magnitude
Qualitative interaction	An interaction in which the presence of factor A's effect only occurs in the presence/absence of factor B
Public health synergy	When phenotype measurements are attributable to a number of risk factors, but phenotype realization is not a function of the sum of the risk factors
Statistical interaction	Any departure from $y = v_1(x_1) + v_2(x_2)$ when modeling response y using factors and [9]

Table 2

Hunter's Rationales for studying GxE interactions in GWAS [37]

Obtain a better estimate of the population-attributable risk for genetic and environmental risk factors by accounting for their joint interactions
Strengthen the associations between environmental factors and diseases by examining these factors in genetically susceptible individuals
Help to dissect disease mechanisms in humans by using information on susceptibility (and resistance) genes to focus on the biological pathways that are most relevant to that disease, and the environmental factors that are most relevant to the pathways
Use the information on biological pathways to design new preventative and therapeutic strategies
Offer tailored preventative advice that is based on the knowledge that an individual carries susceptibility or resistance alleles

Table 3

GWAS studies identifying HTN loci

Paper	Loci Identified (known and novel, not unique)
Adeyemo et al., 2009 [49]	3
Cho et al., 2009 [50]	1
Ehret et al., 2016 [15]	66
Fox et al., 2011 [14]	3
Franceschini et al., 2013 [51]	5
Ganesh et al., 2013 [52]	12
Ganesh et al., 2014 [53]	19
Ho et al., 2011 [54]	10
Hoffman et al. 2017 [55]	316
Hong et al., 2010 [56]	3
Hong et al., 2012 [57]	4
ICBP, 2011 [58]	28
Johnson et al., 2011a [59]	1
Johnson et al., 2011b [60]	8
Kato et al., 2011 [61]	15
Kelly et al., 2013 [62]	4
Levy et al., 2009 [63]	10
Lin et al., 2011 [64]	3
Liu et al., 2016 [17]	21
Lu et al., 2014 [65]	17
Newton-Cheh et al., 2009 [66]	7
Org et al., 2009 [67]	1
Padmanabhan et al, 2010 [68]	1
Qi et al., 2014 [69]	3
Salvi et al., 2012 [70]	1
Simino et al., 2014 [25]	20
Sung et al., 2014 [29]	1
Surendan et al., 2016 [16]	49
Tabara et al., 2010 [71]	4
Takeuchi et al., 2010 [72]	6
Tomazewski et al., 2010 [73]	1
Tragante et al., 2014 [74]	55
Wain et al., 2011 [75]	7
Wang et al., 2009 [76]	1
Wang et al., 2013 [77]	1
Zhu et al., 2011 [78]	1
Total (Non-interaction)	**708**

Paper	Loci Identified (known and novel, not unique)
Basson et al., 2014 [31]	5
Basson et al., 2015 [30]	13
Li et al., 2016 [32]	3
Simino et al., 2013 [27]	1
Simino et al., 2014 [25]	9
Sung et al., 2014 [29]	7
Total (Interaction)	**38**
Grand Total	**746**

Table 4

Studies with Statistically Significant Findings at the MECOM Gene for HTN Related Phenotypes

Study (Interaction?)	Sample Size
ICBP, 2011 [58] (No)	200,000
Tragante et al., 2014 [74] (No)	87,736
Sung et al., 2014 [29] (Yes)	6,889
Ehret et al., 2016 [15] (No)	342,415
Surendan et al., 2016 [16] (No)	192,763