# Editorial

# Blood Pressure Genome-Wide Association Studies, Missing Heritability, and Omnigenics

Brian J. Morris, DSc, PhD



The past 4 decades have seen enormous progress toward understanding the genetic basis of complex polygenic traits. In the hypertension field, the search began with a simple association study of a restriction fragment length polymorphism in the human renin gene in 1988.[1] Although the result was negative, it was followed by a plethora of studies comparing allele frequencies of polymorphisms in potential candidate genes in subjects with essential hypertension with allele frequencies in subjects with normal blood pressure. There were also linkage studies involving pedigrees affected by rare monogenic forms of hypertension. Soon, positive results began to emerge for essential hypertension[2,3] and in identification of single gene mutations responsible for various forms of familial hypertension.[4] In addition, linkage analyses were used to identify hypertension loci in rat models of hypertension[5] and in essential hypertension. Positive findings from the human studies emerged from sib-pair analysis of restriction fragment length polymorphisms at the angiotensinogen gene locus on chromosome 1,[3] in a sib-pair study of microsatellite markers spanning all of chromosome 1,[6] and by genome-wide linkage analysis.[7–9] Next came large-scale genome-wide association studies (GWAS) involving single nucleotide polymorphisms (SNPs) aimed at identification of all genetic loci for blood pressure and essential hypertension.[10] These became, and remain, the tour de force for studies of complex polygenic diseases. More recently, transcriptome-wide studies have been undertaken.[11] These identify, but do not discriminate between, genes that are either directly or indirectly involved in hypertension. The first human study, my my Lab, resulted in the comprehensive identification of numerous messenger RNAs and microRNAs differentially expressed in the kidney in essential hypertension.[11] The study further identified a potential causative mechanism whereby microRNA-mediated repression of the renin gene is lost in hypertensives, so increasing renal renin and thus blood pressure. The size of GWAS has grown enormously during the years, both in number of subjects, number of SNPs with minor allele frequency ≥1%, and number of authors involved in each study, leading to an ever-growing number of loci showing genome-wide significance ($P<5\times10^{-8}$) for blood pressure and hypertension (see reference lists in Kraja et al[12] and Wain et al[13] for the key GWAS studies).

## See Article by Kraja et al

In the current issue of *Circulation: Cardiovascular Genetics*, a collaborative study involving consortia of workers in the United States, United Kingdom, and Europe report findings from one of the largest ever GWAS (n=475 000 subjects) undertaken to identify loci for blood pressure.[12] In all, 21 SNPs were significant at the genome-wide level, 4 of these being at novel loci. The authors discuss how these could affect blood pressure via effects on ion transport in the kidney, vascular endothelial function, and cardiac function. Although some variants were nonsynonymous, resulting in an amino acid change affecting structure and possibly function of the encoded variants, most were synonymous. The SNPs either affect gene expression themselves or serve as markers for a functional variant with which they are in linkage disequilibrium. Functional experiments will be required to confirm whether or not variation in a gene does indeed affect blood pressure and, if so, to identify the causative mechanism. Such studies may involve human cells or animal models of hypertension.

I would now like to place the new findings in the broader context of new ideas on GWAS data analysis and molecular biological concepts relevant to the longstanding issue of missing heritability in GWAS findings.

The estimated variance in a phenotype—in this case blood pressure—explained by SNPs discovered by GWAS has been only a fraction of the heritability estimated from family and twin studies.[14] Recently, issues that could influence the accuracy of SNP-based heritability (the fraction of phenotypic variance explained by additive contributions from SNPs), definitions, assumptions, and interpretation of the models used, as well as pitfalls of misusing the methods and misinterpreting the models and results, has provided a roadmap that may help in understanding the genetic architecture of complex traits and assist in design of experiments to dissect, in full, genetic variation and genetic correlations.[15,16] The estimation of SNP heritability requires assumptions about the distribution of heritability across the genome. Missing heritability could arise from the presence of a large number of yet-to-be discovered common variants of small effect size, rare variants of large effect size that are not tagged by common SNPs on genotyping arrays, and inflation of heritability as a result of shared environmental effects, nonadditive genetic variation,

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Basic and Clinical Genomics Laboratory, Department of Physiology, School of Medical Sciences and Bosch Institute, University of Sydney, New South Wales, Australia.

Correspondence to Brian J. Morris, DSc, PhD, FAHA, Basic & Clinical Genomics Laboratory, School of Medical Sciences, University of Sydney, Bldg F13, Sydney, New South Wales 2006, Australia. E-mail brian.morris@sydney.edu.au

**(*Circ Cardiovasc Genet*. 2017;10:e001943. DOI: 10.1161/CIRCGENETICS.117.001943.)**



***Circ Cardiovasc Genet* is available at http://circgenetics.ahajournals.org**

**DOI: 10.1161/CIRCGENETICS.117.001943**

and epigenetic influences. Thus, the proportion of the variance attributable to all common SNPs (those with minor allele frequency ≥1%, as used in GWAS) needs to be quantified. If heritability is mostly contributed by common SNPs, then a sufficiently large sample size for GWAS would eventually absorb the heritability that had seemed to be missing. For complex traits, the likely large number of common variants having effect sizes too small to survive the genome-wide significance threshold of $P<5\times10^{-8}$ when sample size is, say, 10 000 will emerge as sample size increases toward, let us say, 1 million.

Recently, Speed et al[16] produced an empirical model that more accurately ascertains approximate relationships between the expected heritability of an SNP and minor allele frequency, levels of linkage disequilibrium with other SNPs, and genotype certainty. Their linkage disequilibrium-adjusted kinships model consistently gave a better description of the relationship between heritability and the linkage disequilibrium specified. For complex traits, it implied that common SNPs explain considerably more phenotypic variance than previously reported, so leading to higher estimates of common SNP heritability. For data sets larger than the ones they tested, they predicted that inclusion of rare SNPs will increase SNP heritability, so assisting in the search for missing heritability.

In another recent article, by Yang et al,[15] the proportion of phenotypic variance explained by all SNPs used in GWAS was provided by their genomic-relatedness matrix restricted maximum-likelihood estimate. The latter estimates the upper limit of genome-wide significant heritability. SNP-based heritability will always be smaller than overall heritability. A sparse SNP array is less likely to discover trait-associated variants, even when sample size is large. Less phenotypic variance will be explained in a genomic-relatedness matrix restricted maximum-likelihood analysis than will be with a denser SNP array. If causal variants are located in genomic regions with a different linkage disequilibrium property compared with the rest of the genome, bias in SNP-based heritability can result. Further attention is required to resolve the relationship between local linkage disequilibrium, locus heterozygosity, and additive genetic variance for complex traits. It is likely that these may differ across the genome and between different traits.

Other advances will come from a better molecular understanding of complex polygenic traits. Many GWAS hits may have no relevance to disease, acting instead through complex molecular regulatory networks to influence a few core genes.[17] Boyle et al[17] proposed an omnigenic model whereby complex polygenic traits are caused by miniscule contributions from a vast number of sufficiently interconnected peripheral DNA variants that affect core disease-related genes in relevant tissues. As a result, most heritability can be explained by effects on genes outside core pathways.[17] Importantly, the recognition that complex traits are driven mainly by noncoding variants has led to the discovery that these are enriched in regions of active chromatin, such as promoters and enhancers in trait-relevant cell types. Boyle et al[17] argue that because enrichment of the signal in relevant genes is weak, current conceptual models for complex diseases are incomplete. Their analyses suggest that there may be as many as 100 000 SNPs whose tiny effect sizes contribute to complex traits and that these variants are uniformly distributed across the genome. Surprisingly, they found that although genetic contribution is concentrated in regions that are transcribed or marked by active chromatin in relevant tissues, genes broadly expressed in other tissues contribute more. Thus, although the variants with the largest effect are modestly enriched in specific genes or pathways likely to have direct roles in disease, variants that contribute to most of the heritability tend to be spread across the genome and are far from genes with disease-relevant functions. The highly interconnected cell regulatory networks are such that any expressed gene is likely to affect the regulation or function of core genes. The connections would be multiple and include transcriptional networks, post-translational modifications, protein–protein interactions, and intercellular signaling. Many kinds of networks contain structures consisting of distinct modules of connected nodes and multiple long-range connections. Thus, any gene expressed in a disease-relevant cell is likely to be only a step or so away from one or more core genes. As a result, any variant that affects expression of a peripheral gene will affect the regulation of core gene(s), so contributing to a small effect on disease risk. Because the totality of expressed genes may exceed core genes by, say, 100:1, the sum of the small effects of all peripheral genes would exceed the contribution of genetic variants affecting just the core genes. *Cis*-acting expression-quantitative trait loci may affect mRNA or protein levels of unlinked genes via the regulatory network, so also being *trans* acting expression-quantitative trait loci, and might have other functional effects, such as post-translational modification or alteration in subcellular localization. Many trans–expression-quantitative trait loci may act through protein networks, so may not be detectable from RNA data.

Many complex traits and diseases, such as blood pressure and hypertension, are mediated through multiple cell types in divergent tissues. Although SNPs from GWAS are commonly enriched in active chromatin, to date, few specific expression-quantitative trait loci have been identified. It is probably the case that risk variants operate in only specific cells in one or more tissue types and then only under the influence of certain environmental conditions, such as—in the case of essential hypertension—a diet high in sodium chloride. The networks would vary between cell types, and the effect of a particular variant would be high in some cells and low or nonexistent in others.

Thus, the omnigenic model of complex disease proposes that (1) any gene with regulatory variants in any tissue that contributes to disease pathogenesis may increase risk for that disease and (2) because core genes are greatly outnumbered by peripheral genes, peripheral genes not directly involved in the disease nevertheless make a large contribution to that disease or trait, that is, have a relatively large effect size. Moreover, Boyle et al suggested that a single variant may affect multiple traits if the different traits involve the same cell type and its cell networks. They refer to this as network pleiotropy. Traits that share core genes or involve genes that are close in the network would tend to have effects that are correlated. On the contrary, traits mediated through the same tissue, but with no overlap of core genes, would not exhibit a correlation of effects, even though they share many causal

variants. The most strongly associated SNPs from GWAS point to core genes. Those variants with the largest effect size are more likely to affect protein-coding sequences, so they suggested that large-scale sequencing, starting with exomes, may be a promising next step.[17] At the same time, large-scale genotyping will be needed for modeling the flow of regulatory information through cellular networks. If the omnigenic model is correct, then a much better understanding of cell-specific regulatory networks will be needed to fully understand the biology of complex traits, such as blood pressure, and diseases, such as hypertension.

I would like to suggest that physical interactions between genes might be a contributory factor in the omnigenic model. Chromatin architecture is organized by CTCF (CCCTC-binding factor zinc finger protein)—a transcription factor. CTCF binds to tens of thousands of sites across the genome by using different combinations of its 11 zinc finger domains to bind different DNA target sequences and proteins. *Cis*-regulatory elements are brought together into coregulated islands and multiple islands are brought together into a functional neighborhood, or archipelago, by chromatin looping. These chromatin islands, referred to as topologically associating domains, can be detected through crosslinking experiments and are of the order of several hundred kilo-base pairs, whereas archipelago connections can be on the order of 3 to 5 Mb. Natural selection acts on the genome to maintain combinations of genes and their regulatory elements that are essential to fundamental biological processes. Through CTCF binding and chromatin looping, genes can interact with each other. We recently published findings that showed how the gene *FOXO3* (forkhead/winged helix box gene, group O, type 3), which encodes the key transcription factor FoxO3, uses this mechanism to interact with genes in a 7.3-Mb, 46-gene neighborhood surrounded by gene deserts at chromosome 6q21.[18] In a recent editorial in the present journal, I discussed methods that can be used to identify and study local gene–gene interactions.[19] FoxO3 is expressed ubiquitously and regulates a vast repertoire of genes genome wide whose functions enhance cell resilience and thus healthy aging. *FOXO3* is located at the hub of an early replicating region of chromosome 6q21 and is at the center of the 46-gene chromatin domain. Like *FOXO3*, these genes encode proteins with functions relevant to cell resilience. *FOXO3* SNPs are associated with protection against mortality from coronary artery disease and with survival to age ≥95 years. Cells with protective genotypes exhibit 3-fold higher expression of *FOXO3* mRNA.[18] In response to cell stress, we showed by fluorescence in situ hybridization that *FOXO3* moves from a quiescent location in the cell nucleus to one that promotes expression and that in so doing, it transits toward both nearby and distant genes in the 46-gene neighborhood as the gene complex formed is drawn into the transcription apparatus.[18] The effect was stronger for a *FOXO3* genotype known to be associated with greater cell resilience, lower coronary death, and longevity. The existence and operation of gene factories that amplify the effect of a trait-associated genotype is worth exploring in the case of the various genes that have been identified by GWAS. Perhaps such gene–gene interactions could explain some of the missing heritability.

Recently, Joung et al[20] found that there are long noncoding RNAs that act locally to regulate the expression of genes in their neighborhood. This work involved a genome-wide CRISPR-Cas9 screen targeting >10 000 long noncoding RNA-transcription start sites across the genome. Together with other experiments, they found that transcription of both long noncoding RNA and mRNA can regulate one another in a positive feedback mechanism that then activates a broader gene neighborhood. They provided an example of coordinated regulation of 4 nearby genes. This kind of approach should help elucidate the complex roles of the noncoding genome in complex traits and diseases. Perhaps there might be a role for long noncoding RNAs and other noncoding RNAs in explaining missing heritability.

I trust that the new ideas on statistical analysis of GWAS data and concepts revolving around the omnigenic model of polygenic heritability will assist in understanding and solving, at the molecular level, what it is that explains the genetic basis of blood pressure variance, as well as of other traits, and of complex polygenic diseases, such as hypertension.

## Disclosures

None.

## References


1. Morris BJ, Griffiths LR. Frequency in hypertensives of alleles for a RFLP associated with the renin gene. *Biochem Biophys Res Commun*. 1988;150:219–224.
2. Ying LH, Zee RY, Griffiths LR, Morris BJ. Association of a RFLP for the insulin receptor gene, but not insulin, with essential hypertension. *Biochem Biophys Res Commun*. 1991;181:486–492.
3. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, et al. Molecular basis of human hypertension: role of angiotensinogen. *Cell*. 1992;71:169–180.
4. Lifton RP, Dluhy RG, Powers M, Rich GM, Cook S, Ulick S, et al. A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature*. 1992;355:262–265. doi: 10.1038/355262a0.
5. Hilbert P, Lindpaintner K, Beckmann JS, Serikawa T, Soubrier F, Dubay C, et al. Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature*. 1991;353:521–529. doi: 10.1038/353521a0.
6. Glenn CL, Wang WY, Benjafield AV, Morris BJ. Linkage and association of tumor necrosis factor receptor 2 locus with hypertension, hypercholesterolemia and plasma shed receptor. *Hum Mol Genet*. 2000;9:1943–1949.
7. Xu X, Yang J, Rogus J, Chen C, Schork N, Xu X. Mapping of a blood pressure quantitative trait locus to chromosome 15q in a Chinese population. *Hum Mol Genet*. 1999;8:2551–2555.
8. Sharma P, Fatibene J, Ferraro F, Jia H, Monteith S, Brown C, et al. A genome-wide search for susceptibility loci to human essential hypertension. *Hypertension*. 2000;35:1291–1296.
9. Benjafield AV, Wang WY, Speirs HJ, Morris BJ. Genome-wide scan for hypertension in Sydney sibships: the GENIHUSS study. *Am J Hypertens*. 2005;18:828–832. doi: 10.1016/j.amjhyper.2004.12.010.
10. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang SJ, et al. Framingham heart study 100K project: genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet*. 2007;8(suppl 1):S3. doi: 10.1186/1471-2350-8-S1-S3.
11. Marques FZ, Campain AE, Tomaszewski M, Zukowska-Szczechowska E, Yang YH, Charchar FJ, et al. Gene expression profiling reveals renin mRNA overexpression in human hypertensive kidneys and a role for microRNAs. *Hypertension*. 2011;58:1093–1098. doi: 10.1161/HYPERTENSIONAHA.111.180729.
12. Kraja AT, Cook JP, Warren HR, Surendran P, Liu C, Evangelou E, et al. New blood pressure associated loci identified in meta-analyses of 475 000 individuals. *Circ Cardiovasc Genet*. 2017;10:e001778. doi: 10.1161/CIRCGENETICS.117.001778.

13. Wain LV, Vaez A, Jansen R, Joehanes R, van der Most PJ, Erzurumluoglu AM, et al. Novel blood pressure locus and gene discovery using genome-wide association study and expression data sets from blood and the kidney [published online ahead of print July 24, 2017]. *Hypertension*. doi: 10.1161/HYPERTENSIONAHA.117.09438. http://hyper.ahajournals.org/content/70/3/e4. Accessed September 22, 2017.
14. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753. doi: 10.1038/nature08494.
15. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet*. 2017;49:1304–1310. doi: 10.1038/ng.3941.
16. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ; UCLEB Consortium. Reevaluation of SNP heritability in complex human traits. *Nat Genet*. 2017;49:986–992. doi: 10.1038/ng.3865.
17. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–1186. doi: 10.1016/j.cell.2017.05.038.
18. Donlon TA, Morris BJ, Chen R, Masaki KH, Allsopp RC, Willcox DC, et al. *FOXO3* longevity interactome on chromosome 6. *Aging Cell*. 2017;16:1016–1025. doi: 10.1111/acel.12625.
19. Morris BJ. Gene team in bood pressure genetics. *Circ Cardiovasc Genet*. 2017;10:e001776. doi: 10.1161/CIRCGENETICS.117.001776.
20. Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, Aguet F, et al. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature*. 2017;548:343–346. doi: 10.1038/nature23451.

KEY WORDS: Editorials ■ blood pressure ■ genetics ■ genome, human ■ hypertension

# Blood Pressure Genome-Wide Association Studies, Missing Heritability, and Omnigenics

Brian J. Morris

*Circ Cardiovasc Genet.* 2017;10:
doi: 10.1161/CIRCGENETICS.117.001943
*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268