npg

ARTICLE

Replication of the Wellcome Trust genome-wide association study of essential hypertension: the Family Blood Pressure Program

Georg B Ehret[1], Alanna C Morrison[2,3], Ashley A O'Connor[1], Megan L Grove[2,3], Lisa Baird[4], Karen Schwander[5], Alan Weder[6], Richard S Cooper[7], DC Rao[5], Steven C Hunt[4], Eric Boerwinkle[2,3] and Aravinda Chakravarti*,[1]

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [2]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; [3]Division of Epidemiology, University of Texas Health Science Center at Houston, Houston, TX, USA; [4]Department of Medicine, Cardiovascular Genetics Division, University of Utah, Salt Lake City, UT, USA; [5]Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; [6]Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical School, Ann Arbor, MI, USA; [7]Department of Preventive Medicine and Epidemiology, Loyola University Stritch School of Medicine, Maywood, IL, USA

Essential hypertension is a principal cardiovascular risk factor whose origin remains unknown. Classical genetic studies have shown that blood pressure is at least partially heritable, opening a window to understanding the pathophysiology of essential hypertension in the human using modern genetic tools. The Wellcome Trust Case Control Consortium has recently published the results of screening the genomes of 2000 essential hypertension cases and 3000 controls using 500 000 genome-wide single nucleotide polymorphisms (SNPs). None of the variants proved to be genome-wide significant after correction for multiple tests but the most significantly associated SNPs ($P < 10^{-5}$) constitute a priority list that warrant follow-up in other studies. We describe here replication studies of the top six SNPs in subjects from the US National Heart, Lung, and Blood Institute funded Family Blood Pressure Program comprising 11 433 individuals recruited by hypertensive families. The results suggest that only one of the six SNPs might be associated with essential hypertension in Americans of European origin. This SNP shows a significant but opposite effect in Americans of Hispanic origin and no association in African Americans. The significance of the opposing effect estimates is unclear. No replication could be shown for hypertension status, but there are differences in study design. This attempted replication highlights that essential hypertension studies will require more comprehensive and larger genetic screens.

*Correspondence: Professor A Chakravarti, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Broadway Research Building, Suite 579, 733 N. Broadway, Baltimore, MD 21205, USA. Tel: +1 410 502 7525; Fax: +1 410 502 7544;
E-mail: aravinda@jhmi.edu

Introduction

Blood pressure (BP) is a classical heritable quantitative trait in humans.[1] When elevated and the primary cause is unknown, it is termed essential hypertension (EH) and is a leading susceptibility factor for stroke, end-stage renal disease, and coronary artery disease. Hypertension affects 33.6% of Americans (2004 prevalence estimate)[2] and a similar

proportion of the entire world population. Worldwide, it is the third leading risk factor for all morbidity and mortality explaining 4.4% of the global burden of disease.[3] Consequently, EH entails huge morbidity, mortality, and cost when untreated. Despite progress on a number of fronts, the pathophysiology of hypertension remains largely obscure.

Ever since the classic work of Pickering and co-workers,[1] hypertension has been considered a multifactorial trait. It is a common assumption that EH is in substantial part due to the genetic effects of many genes and interaction of these genes with the environment, each of these factors imparting a small contribution to risk. The elucidation of the molecular details of rare monogenic forms of hypertension by Lifton *et al.*[4] have described pathways for BP regulation but the genes identified have not been shown to explain a substantial fraction of EH risk in the general population. EH genetic studies have identified innumerable genetic variants in one or several populations, however, few variants can be consistently detected in multiple studies. The Family Blood Pressure Program (FBPP) is a National Heart, Lung, and Blood Institute (NHLBI)-funded multicenter study with coordinated clinical protocols and a pooled data resource that is publicly accessible.[5] The four FBPP networks have recruited a total of 17 129 individuals including four major United States ethnic groups: African Americans (AA), European Americans (EA), Hispanic Americans (HA), and Asian Americans.

Recent developments in genotyping technology permit association and family-based studies using hundreds of thousands of markers in thousands of individuals. This is an exciting opportunity to assess a major part of the human genome in an unbiased fashion for genes associated with EH. The Wellcome Trust Case Control Consortium (WTCCC) has published the largest genome-wide association study on hypertension to this date with 2000 cases of EH and 3000 controls.[6] Although no single SNP examined reached genome-wide significance ($P < 5 \times 10^{-7}$), the six variants most significantly ($P < 10^{-5}$) associated with EH constitute a priority list for follow-up in other well-characterized hypertension samples. We have analyzed the six most significant SNPs in 11 433 individuals from the FBPP, using systolic blood pressure (SBP) and diastolic blood pressure (DBP), and derived phenotypes as quantitative phenotypes as well as hypertension status.

Methods

FBPP and phenotyping

Three of four FBPP networks participated in this replication study: GenNet, GENOA (Genetic Epidemiology of Arteriosclerosis Network), and HyperGEN (Hypertension Genetic Epidemiology Network). The respective institutional review boards approved all studies and all individuals included gave written informed consent. Information on population characteristics is listed in Table 1. Detailed information on procedures of inclusion and participant characteristics are published elsewhere.[5] BP measurements were carried out according to standard operating procedures in a sitting position after a resting period.[5] For GenNet, the average of two manual BP measurements was used. For GENOA and HyperGEN, the average of up to three Dinamap measurements were used where available and the average of up to two Omron measurements if no Dinamap measurement was recorded. For the current analysis, raw SBP and DBP values were used, without correction for treatment. Analysis of treatment corrected SBP and DBP values (adding 10 mm Hg to SBP and 5 mm Hg to DBP if treatment present) gave very similar results. Correction by addition has been shown to perform well in simulation studies, but all approaches to this problem are necessarily approximations.[7] Hypertension status was defined as SBP ≥ 140 mm Hg and/or DBP ≥ 90 mm Hg and/or antihypertensive treatment.

Table 1 Basic characteristics of study participants

	total number of individuals	% female	age: median; range [years]	BMI: median; range [kg/m²]	number of families	number of unrelated unaffected	SBP: median; range [mmHg]	DBP: median; range [mmHg]	% hypertensive	% with anti-hypertensive therapy	no. of individuals genotyped
GenNet											
AA	1,101	60.2	38; 12-81	28.5; 16.1-63.3	354	333	121; 77-202	75; 30-136	33.2%	21.6%	1,094
EA	1,497	54.6	43; 11-85	28.0; 14.5-65.2	429	449	121; 88-246	77; 41-112	30.1%	20.0%	1,478
HA	847	61.6	33; 11-83	28.1; 16.1-69.8	213	237	110; 70-220	69; 40-119	13.7%	8.6%	841
total	3,445	58.1	39; 11-85	281; 14.5-69.8	996	1,019	118; 70-246	74; 30-136	27.1%	17.7%	3,413
Genoa											
AA	1,857	69.1	58; 20-91	30.0; 14.4-61.4	642	289	129; 78-236	70; 44-116	72.3%	62.1%	1,761
EA	1,578	55.4	55; 24-100	29.5; 15.8-67.9	553	236	132; 81-232	76; 40-109	74.1%	64.2%	1,443
HA	1,799	58.9	55; 18-93	30.0; 13.6-63.8	438	331	129; 75-225	71; 43-115	44.3%	23.6%	1,674
total	5,234	61.5	57; 18-100	29.8; 13.6-67.9	1,633	856	130; 75-236	72; 40-116	63.2%	49.5%	4,878
HyperGen											
AA	2,010	64.0	49; 18-88	30.7; 13.8-73.7	948	281	128;72-228	74; 42-123	71.4%	62.3%	1,603
EA	1,904	53.3	60; 18-95	29.5; 15.1-68.7	872	262	119; 74-216	68; 40-115	66.8%	61.5%	1,539
total	3,914	58.8	55; 18-95	29.6; 13.8-73.7	1,820	543	123; 72-228	71; 40-123	69.1%	61.9%	3,142
joint dataset	**(all of the 3 above)**										
AA	4,968	65.1	50; 12-91	30.1; 13.8-73.7	1,944	903	126; 72-236	73; 30-136	63.1%	53.2%	4,458
EA	4,979	54.4	55; 11-100	28.7; 14.5-68.7	1,854	947	124; 74-246	73; 40-115	57.8%	49.9%	4,460
HA	2,646	59.8	50; 11-93	29.4; 13.6-69.8	651	568	121; 70-225	70; 40-119	34.0%	18.8%	2,515
total	12,593	59.7	52; 11-100	29.4; 13.6-73.7	4,449	2,418	124; 70-246	72; 30-136	54.9%	44.6%	11,433

Abbreviations: AA, African Americans; BMI, body mass index; DBP, diastolic blood pressure; EA, European Americans; GENOA, Genetic Epidemiology of Arteriosclerosis Network; HA, Hispanic Americans; HyperGEN, Hypertension Genetic Epidemiology Network; SBP, systolic blood pressure.

Genotyping

We used the Taqman technology with fluorescent probes for genotyping (Applied Biosystems, Foster City, CA, USA). Genotypes were called manually using the manufacturer's software. Each network performed genotyping separately. Extensive DNA quality control has been carried out previously on all samples. GenNet typed 1% repeat samples and 1% HapMap samples with known genotypes for the six SNPs. HyperGEN typed 7% HapMap samples. GENOA typed 3.5% repeats. No discrepancy between repeats or between the HapMap sample genotype and the reference was detected, except for the GENOA repeats for rs2398162, which showed a concordance rate of 99.4%.

Power calculations

To estimate power in our study we used Purcell's genetic power calculator[8] (http://pngu.mgh.harvard.edu/~purcell/gpc/), using a case–control sampling for threshold-selected quantitative traits. The following parameters were used: type I error = 5%; percent variance explained = 0.5% or 1%, additive genetic effects model; number of cases = 2500 for AA and EA, 1250 for HA; cases assumed to be ascertained from >1 SD of the phenotype distribution and controls from the entire population; control:case ratio = 1.

Analysis methods

Data file preparation and descriptive statistics were carried out using custom code written in R version 2.6.0 (The R Foundation for Statistical Computing) and by Perl scripts. For pedigree verification and Hardy–Weinberg tests, the Pedstats software package was used.[9] For regression of the phenotype on age, sex, and BMI, the residuals of the following model were used for further analysis: $y = \alpha + \beta1 \times \text{age(years)} + \beta2 \times \text{sex}(1$ for male subject, 2 for female subject$) + \beta3 \times \text{BMI(kg/m}^2) + \varepsilon$. For regression-based tests of association, the Merlin software package (assoc option)[10] and the XTGEE procedure (GEE in the following) of STATA (StataCorp LP, College Station, TX, USA) were used for continuous phenotypes, the Lamp software package[11] for dichotomous phenotypes. Merlin assumes an additive model and Lamp was set to an additive model. For TDT-based tests, the FBAT/PBAT software package (hbat–p option)[12] and the QTDT software package[13] were used. P-values were not adjusted for multiple testing. The transcription factor binding site prediction was carried out using AliBaba 2.1 software package using the TRANSFAC database (Biological Databases, Wolfenbüttel, Germany).

Results

The FBPP data set

The combined FBPP data (three networks: GenNet, GENOA, and HyperGEN) comprises a large data set with phenotype data available on 12 593 individuals. It is composed of 39% AA, 40% EA, and 21% HA participants (Table 1). DNA was genotyped on 11 433 (91%) of participants on which DNA was available. The most salient difference among the three networks is the younger age of participants in GenNet, their lower BMI, family ascertainment based on elevated BP rather than hypertension, and a lower frequency of antihypertensive treatment (for details see Online Data Supplement).

Family-based association tests

To maximize power, we analyzed networks jointly, but stratified by ethnicity. SBP and DBP were used as quantitative traits and residuals after regression on age, sex, and BMI were also tested to control for the most important confounders. We used the regression-based Merlin software (see 'Methods') for the primary analysis on quantitative traits. We also tested one other regression-based analysis method (GEE; see 'Methods') and two TDT-based tests (FBAT/PBAT and QTDT). P-values calculated by Merlin were found to be closely correlated to the P-values calculated with GEE ($r^2 = 0.91$ 286 data points); P-values from the FBAT/PBAT analyses were found to be weakly correlated to the P-values calculated by QTDT ($r^2 = 0.27$, 140 data points) and uncorrelated to the regression-based tests. Given the close correlation between the regression-based tests and the greater power of these methods (all samples and not only the informative are considered) we used Merlin for primary testing. To estimate power we used the genetic power calculator of Purcell et al.[8] Power was close to 0.7 for AA and EA and close to 0.4 for HA, assuming a total QTL effect of 1% (rs2820037 and rs1937506) (see 'Methods' and Online Data Supplement for parameters and details).

Table 2 depicts the results of the analysis by Merlin for two systolic, two diastolic phenotypes, and hypertension status. In addition to SBP and DBP, their residuals after regression on age, sex, and BMI were analyzed to investigate effects due to these confounders. The tests carried out are not two independent tests because the parent variable and the residuals remain highly correlated (see phenotype r^2 in Table 2). Additionally, SBP is correlated with DBP ($r^2 = 0.45$, 0.38, and 0.42 for AA, EA, and HA, respectively). We obtained significant results with one SNP (rs1937506) for more than one phenotype tested. This significance reproduces for both systolic phenotypes in EA and HA (Table 2). The effect size due to this variant is large for systolic phenotypes in both EA and HA, but the effect is in opposite directions. The effect sizes and their direction are depicted in Figure 1 for the analysis of SBP residuals. Each additional G allele of rs1937506 is associated with a 24.9 mm Hg decrease of SBP in EA and with a 27.7 mm Hg increase in HA. Considering DBP residuals, each additional allele is associated with an 8.7 mm Hg decrease of DBP in EA, whereas the results are

Table 2 Analysis of gene effects by ethnicity (regression-based tests)

ethnicity	variant	reference allele	HTN status	SBP		phenotype r^2 <— —>	SBP - asb regressed		DBP		phenotype r^2 <— —>	DBP - asb regressed	
			P-value	P-value	eff. size [mmHg]		P-value	eff. size [mmHg]	P-value	eff. size [mmHg]		P-value	eff. size [mmHg]
AA						0.87					0.92		
	rs2398162	A	0.400	0.668	-8.3		0.477	-11.1	1.000	-0.7		1.000	-0.4
	rs2820037	A	0.180	0.337	10.1		0.631	4.0	0.668	-1.4		0.668	-1.2
	rs6997709	G	0.740	0.439	10.4		0.710	3.8	1.000	-0.6		1.000	-0.1
	rs7961152	C	0.340	0.520	6.9		0.439	6.8	0.240	4.1		0.283	3.1
	rs11110912	C	0.660	0.830	-3.0		0.477	-15.0	0.710	-1.8		0.830	-1.3
	rs1937506	G	0.550	0.830	-1.6		0.762	-3.3	0.830	0.9		0.762	1.0
EA						0.77					0.77		
	rs2398162	A	0.640	0.265	-13.2		0.218	-11.5	0.570	2.0		0.762	0.9
	rs2820037	A	0.830	0.057	27.3		0.108	17.2	0.020	9.8		0.086	6.2
	rs6997709	G	0.670	0.520	-6.3		0.406	-7.8	0.439	-2.6		0.337	-2.7
	rs7961152	C	0.440	0.198	-12.3		0.830	-1.6	0.044	-6.1		0.081	-4.4
	rs11110912	C	0.910	0.762	6.2		0.762	-3.2	0.477	3.0		0.668	1.4
	rs1937506	G	0.600	0.032	-26.8		0.009	-24.9	0.077	-6.2		0.004	-8.7
HA						0.72					0.85		
	rs2398162	A	0.068	0.159	22.7		0.211	-13.4	0.350	3.3		0.710	-1.2
	rs2820037	A	0.300	0.497	16.1		0.422	12.9	0.089	8.9		0.159	6.3
	rs6997709	G	0.320	1.000	3.5		0.077	24.4	0.710	1.6		0.599	2.0
	rs7961152	C	0.060	0.406	-14.1		0.256	13.2	0.325	3.6		0.477	2.3
	rs11110912	C	0.420	0.668	-9.4		0.439	13.6	0.497	-3.8		0.599	-2.5
	rs1937506	G	0.060	0.005	49.9		0.024	27.7	0.325	4.0		0.164	4.9

MERLIN 1.1-alpha (association option) was used for this analysis
SBP: systolic blood pressure
DBP: diastolic blood pressure
asb regressed: regressed on age, sex, and bmi
AA: African Americans; EA: European Americans; HA: Hispanic Americans
phenotype r^2: correlation between the phenotype to the left and right (asb regressed)
P-values were not adjusted for multiple testing.
An additive model was assumed for all analyses.
P-value <0.05
P-value <0.01

nonsignificant for HA. There is no significant result for this SNP in AA.

We also investigated association with hypertension status using the Lamp software for discrete traits (see 'Methods'). No significant result was obtained in any population studied (Table 2), although several variants show a trend.

Genomic position of rs1937506 and binding motif analysis
The SNP rs1937506 is located in a 500 kb gene desert on chromosome 13q21. The adjacent genes are a predicted olfactory receptor and the protocadherin 9 isoform 1 precursor genes. Neither gene has been reported to be involved in EH. Sequence analysis reveals an Oct-1 and an ICSBP binding site in close proximity to rs1937506.

Discussion
New genomic tools now enable geneticists to assess genetic variability as the basis of common disease in an unbiased fashion. The results of the WTCCC study on EH using 2000 cases and 3000 controls is the first such example in the field of hypertension. None of the 469 557 SNPs the WTCCC tested (post-genotyping quality control) were significantly associated with hypertension status after correction for multiple testing. Six SNPs with moderate evidence for association were identified ($5 \times 10^{-7} < P < 1 \times 10^{-5}$) based on a Bayesian's interpretation of detection thresholds. We attempted to replicate these six SNPs in the participants of the FBPP in this study.

The design of the WTCCC genome-wide association study on hypertension has been criticized by the WTCCC investigators and others.[6] Although cases were recruited at the extremes of the BP distribution, were diagnosed before 60 years of age, and were mostly non-obese (BMI < 30 kg/m^2) – all of which are believed to increase power by focusing on more genetic forms – analysis was based on hypertension status as opposed to BP levels. A quantitative phenotype is generally thought to significantly improve power, but only as long as medication does not confound BP or can be adequately controlled for. The absence of genome-wide significant findings in the WTCCC study on HTN might also be due to the following reasons: (1) variants associated with EH were assayed but the results are not significant because of small effect sizes, the experi-
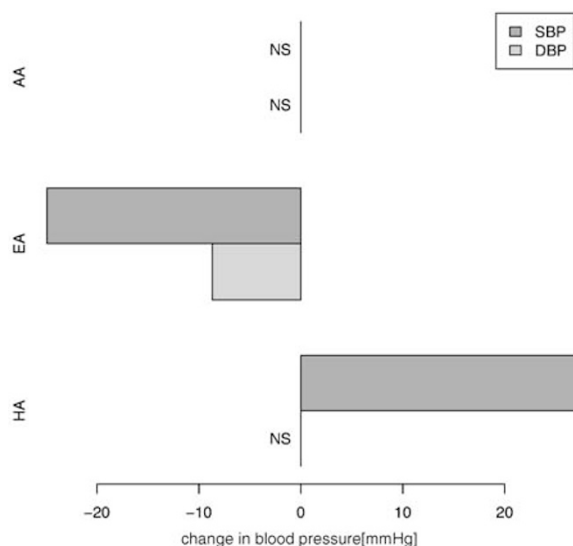
Figure 1 Effect sizes of each additional copy of the *G* allele in rs1937506 on age, sex, and BMI-regressed systolic blood pressure (SBP) in African-American (AA), European-American (EA), and Hispanic-American (HA) samples. Only the significant effect sizes are depicted. NS, nonsignificant.

not in HA and this might be related to the lower power in HA given the smaller sample size. At this point it is unclear if rs1937506 is the causal variant for the association observed and it is also unclear if ethnicity-specific, opposite-directed effects exist. Molecular biology experiments will confirm or refute this hypothesis. These studies will also need to take into account the genomic position of this SNP in a gene desert. Other variants associated with disease have also been identified in gene deserts by genome-wide association studies.[14]

Conflict of interest/Disclosures
None.

mental design (eg controls are from the general population and will likely contain a significant proportion of untreated hypertensives); (2) the variants associated with EH were not assayed because of insufficient density of markers or markers that are of intermediate or rare frequency. Quantile–quantile plots of each SNP in the WTCCC study show no data point being significantly different from the expected χ^2-distribution so that even the most significant variants could simply be chance findings. In addition, the true genome-wide significance level in the WTCCC study might be significantly lower than 5×10^{-7} which is derived from a Bayesian argument rather than a frequentist argument. If this is the case, the probability of the findings being false positives would be even greater.

In this replication study we identify one of the six SNPs as potentially associated with SBP and DBP in 4979 EA and 2646 HA from the FBPP (rs1937506). The variant is not significantly associated with hypertension status. This might appear surprising because the WTCCC study used hypertension status as primary phenotype, but several important differences might explain the result: (1) the WTCCC study focused on non-obese individuals and the FBPP has many obese participants (Table 1); this is in line with our observation that regression of SBP or DBP on BMI leads to lower *P*-values in EA for rs1937506. (2) Using a quantitative phenotype can increase power.

Considering SBP and DBP as phenotypes, the effect sizes are large but they are in opposite directions in EA and HA. Some less significant results are only observed in EA and

References
1 Oldham PD, Pickering G, Roberts JA, Sowry GS: The nature of essential hypertension. *Lancet* 1960; **1**: 1085–1093.
2 Rosamond W, Flegal K, Friday G *et al.* Heart Disease and Stroke Statistics — 2007 Update. *Circulation* 2007; **115**: e69–e171.
3 Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S, Murray CJ: Selected major risk factors and global and regional burden of disease. *Lancet* 2002; **360**: 1347–1360.
4 Lifton RP, Gharavi AG, Geller DS: Molecular mechanisms of human hypertension. *Cell* 2001; **104**: 545–556.
5 FBPP: Multi-center genetic study of hypertension: the Family Blood Pressure Program (FBPP). *Hypertension* 2002; **39**: 3–9.
6 WTCCC: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
7 Tobin MD, Sheehan NA, Scurrah KJ, Burton PR: Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005; **24**: 2911–2935.
8 Purcell S, Cherny SS, Sham PC: Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; **19**: 149–150.
9 Wigginton JE, Abecasis GR: PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005; **21**: 3445–3447.
10 Chen WM, Abecasis GR: Family-based association tests for genomewide association scans. *Am J Hum Genet* 2007; **81**: 913–926.
11 Li M, Boehnke M, Abecasis GR: Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. *Am J Hum Genet* 2005; **76**: 934–949.
12 Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM: PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; **74**: 367–369.
13 Abecasis GR, Cardon LR, Cookson WO: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
14 Gudmundsson J, Sulem P, Manolescu A *et al*: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007; **39**: 631–637.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)