

# Whole-genome association study identifies *STK39* as a hypertension susceptibility gene

Ying Wang<sup>a</sup>, Jeffrey R. O'Connell<sup>a</sup>, Patrick F. McArdle<sup>a</sup>, James B. Wade<sup>b</sup>, Sarah E. Dorff<sup>a</sup>, Sanjiv J. Shah<sup>c</sup>, Xiaolian Shi<sup>a</sup>, Lin Pan<sup>d</sup>, Evadnie Rampersaud<sup>a</sup>, Haiqing Shen<sup>a</sup>, James D. Kim<sup>e</sup>, Arohan R. Subramanya<sup>b</sup>, Nanette I. Steinle<sup>a</sup>, Afshin Parsa<sup>f</sup>, Carole C. Ober<sup>d</sup>, Paul A. Welling<sup>b</sup>, Aravinda Chakravarti<sup>g</sup>, Alan B. Weder<sup>h</sup>, Richard S. Cooper<sup>i</sup>, Braxton D. Mitchell<sup>a</sup>, Alan R. Shuldiner<sup>a</sup>, and Yen-Pei C. Chang<sup>a,1</sup>

<sup>a</sup>Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201; <sup>b</sup>Department of Physiology, University of Maryland School of Medicine, Baltimore, MD 21201; <sup>c</sup>Section of Cardiology, Department of Medicine, University of Chicago, Chicago, IL 60611; <sup>d</sup>Department of Human Genetics, University of Chicago, Chicago, IL 60637; <sup>e</sup>Georgetown University School of Medicine, Washington, DC 20057; <sup>f</sup>Division of Nephrology, University of Maryland School of Medicine, Baltimore, MD 21201; <sup>g</sup>Mckusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205; <sup>h</sup>Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, MI 48106; and <sup>i</sup>Department of Preventive Medicine and Epidemiology, Loyola Stritch School of Medicine, Maywood, IL 60153

Edited by Gregg L. Semenza, Johns Hopkins University School of Medicine, Baltimore, MD, and approved November 11, 2008 (received for review August 27, 2008)

Hypertension places a major burden on individual and public health, but the genetic basis of this complex disorder is poorly understood. We conducted a genome-wide association study of systolic and diastolic blood pressure (SBP and DBP) in Amish subjects and found strong association signals with common variants in a serine/threonine kinase gene, *STK39*. We confirmed this association in an independent Amish and 4 non-Amish Caucasian samples including the Diabetes Genetics Initiative, Framingham Heart Study, GenNet, and Hutterites (meta-analysis combining all studies:  $n = 7,125$ ,  $P < 10^{-6}$ ). The higher BP-associated alleles have frequencies  $> 0.09$  and were associated with increases of 3.3/1.3 mm Hg in SBP/DBP, respectively, in the Amish subjects and with smaller but consistent effects across the non-Amish studies. Cell-based functional studies showed that *STK39* interacts with WNK kinases and cation-chloride cotransporters, mutations in which cause monogenic forms of BP dysregulation. We demonstrate that in vivo, *STK39* is expressed in the distal nephron, where it may interact with these proteins. Although none of the associated SNPs alter protein structure, we identified and experimentally confirmed a highly conserved intronic element with allele-specific in vitro transcription activity as a functional candidate for this association. Thus, variants in *STK39* may influence BP by increasing *STK39* expression and consequently altering renal  $\text{Na}^+$  excretion, thus unifying rare and common BP-regulating alleles in the same physiological pathway.

blood pressure | essential hypertension | genome-wide association study | SPAK | *STK39*

Hypertension is a leading cause of morbidity and mortality worldwide, contributing to cardiovascular disease, stroke, and end-stage kidney disease (1). The most common form, essential hypertension (EH), is widely believed to involve multiple genes with variant alleles. Like other complex disorders, manifestation of EH in any single individual is likely dependent on a variety of genetic and environment factors, making identification of EH susceptibility genes in the general population a major challenge. Studies on the genetic basis of rare monogenetic BP disorders, in contrast, have identified mutations in genes affecting a single physiological pathway, renal salt transport (2, 3). Although such studies highlight the importance of this pathway to BP regulation, the underlying genetic basis for EH remains poorly understood.

To identify common EH susceptibility genes, we conducted a genome-wide association (GWA) analysis in the Old Order Amish, a closed founder population who immigrated to the Lancaster, Pennsylvania area from Switzerland in the early 1700s. The members of this community are descendants from a small number of common founders and share a relatively homogeneous lifestyle, making this population ideal for identifying genes that underlie

complex diseases. We analyzed genotype data (Affymetrix GeneChip Human Mapping 100K platform) for systolic blood pressure (SBP) and diastolic blood pressure (DBP) measurements obtained in 542 subjects of the Amish Family Diabetes Study (AFDS) (Table 1) (4). This approach identified a novel hypertension susceptibility gene, *STK39*, in which common variants are associated with BP levels.

## Results

**GWAS in 542 AFDS Subjects.** In our initial GWA scan, we analyzed 79,447 autosomal SNPs that passed our genotype quality control procedure (median and mean inter-SNP distances, 11.7 kb and 30 kb, respectively). All association results are available upon request and those with  $P$  values  $< .0005$  are provided in [supporting information \(SI\) Table S1](#). Our top signal, rs4977950, is significant at the genome-wide level but is located in a gene desert 900 kb away from known genes (9p21.3;  $P = 9.1 \times 10^{-8}$  with SBP). But many of the strongest signals from the GWA of SBP were with a cluster of SNPs located on chromosome 2q24.3 (Fig. 1;  $P = 8.9 \times 10^{-6}$  to  $9.1 \times 10^{-5}$ ) within the *STK39* (serine threonine kinase 39) gene, which encodes the SPAK (Ste20-related proline-alanine-rich kinase) protein. These SNPs also are associated with DBP (Table S1B), albeit at slightly lower levels of statistical significance.

A growing body of evidence indicates that SPAK interacts with cation-chloride cotransporters as a part of an evolutionarily conserved signaling pathway important in controlling salt transport in osmotic cell volume regulation (5). In addition to the known effects of SPAK on the  $\text{Na}^+-\text{K}^+-2\text{Cl}^-$  cotransporter (NKCC1), SPAK also phosphorylates the thiazide-sensitive  $\text{Na}^+-\text{Cl}^-$  cotransporter (NCC) and the loop diuretic-sensitive  $\text{Na}^+-\text{K}^+-2\text{Cl}^-$  cotransporter (NKCC2) (6). Both transporters play major roles in renal salt excretion, as underscored by the direct involvement of NCC and NKCC2 in rare autosomal recessive conditions of hypotension, hypokalemic metabolic alkalosis, and salt wasting, also known as Gitelman and Bartter syndrome (7, 8). SPAK is phosphorylated by WNK1 and WNK4 (lysine-deficient protein kinase 1 and 4) (6, 9), and rare mutations in these 2 genes lead to autosomal dominant pseudohypoal-

Author contributions: Y.W., A.C., A.B.W., B.D.M., A.R. Shuldiner, and Y.-P.C.C. designed research; Y.W., J.B.W., S.E.D., S.J.S., N.I.S., C.C.O., and Y.-P.C.C. performed research; J.R.O., C.C.O., A.B.W., R.S.C., and A.R. Shuldiner contributed new reagents/analytic tools; Y.W., J.R.O., P.F.M., S.J.S., X.S., L.P., E.R., H.S., J.D.K., A.R. Subramanya, A.P., C.C.O., P.A.W., A.C., B.D.M., and Y.-P.C.C. analyzed data; and Y.W., J.B.W., A.R. Shuldiner, and Y.-P.C.C. wrote the paper.

The authors declare no conflicts of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: cchang@medicine.umaryland.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0808358106/DCSupplemental](http://www.pnas.org/cgi/content/full/0808358106/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA

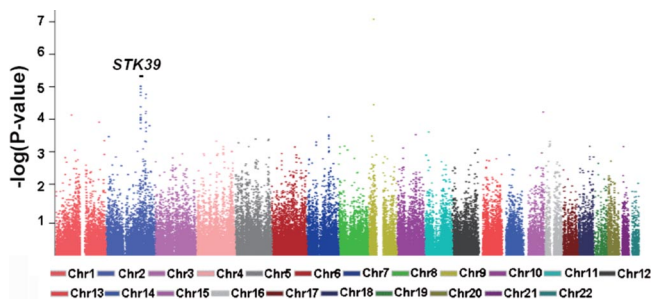
**Table 1. Characteristics of the Amish study subjects**

	Initial screening sample, AFDS	Replication sample 1, AFDS	Replication sample 2, HAPI
Number	542	557	790
Age, years	54.8 ± 13.1	45.8 ± 19.4	42.8 ± 13.9
Sex, % male	44%	45%	54%
Body mass index, kg/m <sup>2</sup>	28.1 ± 5.2	26.4 ± 5.1	26.5 ± 4.5
SBP, mm Hg	128.1 ± 20.9	123.8 ± 17.2	120.7 ± 14.5
DBP, mm Hg	81.0 ± 10.5	76.5 ± 9.8	76.2 ± 8.6
% T2DM	22.2%	9.7%	0.6%
% on antihypertensive medication	11.0%	9.2%	0%

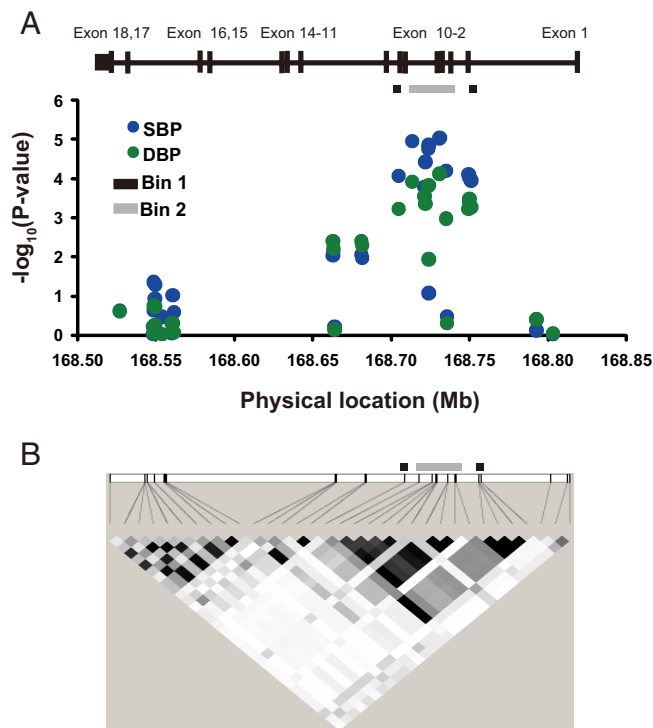
dosteronism type II, characterized by hyperkalemia, hypertension, and sensitivity to thiazide diuretics (Gordon syndrome) (3). Our GWA finding, combined with the emerging role of SPAK in renal electrolyte transport, make *STK39* a logical EH candidate gene.

*STK39* contains 18 exons that span ≈300 kb. There are 39 SNPs on the Affymetrix 100K platform located within 5 kb of *STK39*. The associations of these SNPs with BP and their pairwise linkage disequilibrium (LD) relationships are shown in Fig. 2. Two groups of correlated SNPs ( $r^2 > 0.8$  within groups and  $r^2 > 0.4$  between groups) demonstrated strong evidence of association with SBP ( $P < .0001$ ). The average minor allele frequencies (MAFs) of the 2 groups were 0.2 and 0.1. For simplicity, SNPs that are highly correlated with either rs6749447 or rs3754777 ( $r^2 > 0.8$ ) are referred to as belonging to LD bin 1 and bin 2, respectively (Fig. 2). The association was highly significant when tested using an additive or dominant model (Fig. 3A). The association remained significant when the analysis was restricted to subjects not taking antihypertensive medication (data not shown).

**Replication in Remaining AFDS Subjects and an Independently Collected Amish Cohort.** We genotyped 1 tag SNP from each LD bin, rs6749447 and rs3754777, in the remaining 557 AFDS subjects; the associations also were significant ( $P = .003$  for rs6749447 [Fig. 3B] and  $.02$  for rs3754777). In addition, we selected a second Amish replication set from the Heredity and Phenotype Intervention (HAPI) Heart Study (10), an independently collected Amish sample for which subjects were ascertained without considering diabetes status (Table 1). We genotyped rs6749447 and rs3754777 in this cohort and found strong evidence of association between rs6749447 and SBP ( $n = 790$ ;  $P = .0001$ ; Fig. 3C). In this younger and less hypertensive cohort, the association was in the same direction as that observed in the AFDS, but was more significant when tested under the recessive model. After combining all 3 sets of Amish subjects to determine the effect size of this association, we estimated that having 1 copy of the minor allele was associated with a 3-mmHg higher SBP and a 1-mmHg higher DBP (Table 2).

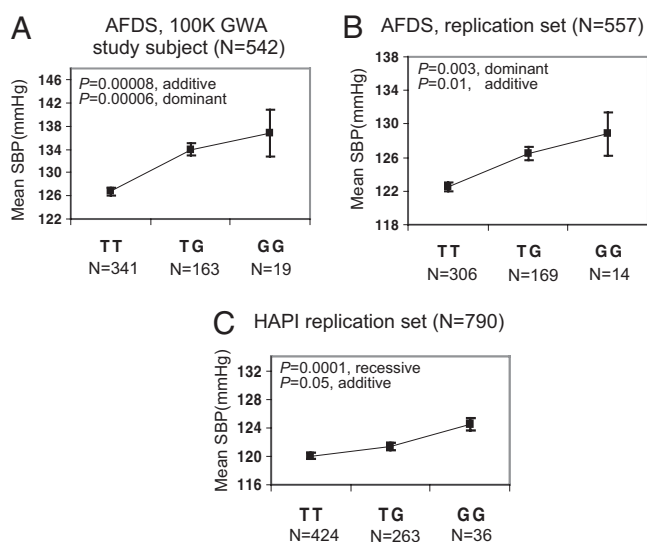


**Fig. 1.** Whole genome association scan results for SBP in 542 AFDS subjects. SNPs from each chromosome are represented by a different color and ordered by physical location.



**Fig. 2.** (A) Exonic structure, association between *STK39* SNPs and SBP/DBP. (B) LD relationship among SNPs. The locations of associated SNPs in LD bins 1 and 2 are denoted by black and gray bars, respectively. Pairwise LD relationships ( $r^2$ ) among SNPs are indicated by color (black = 1; white = 0; shades of gray =  $0 < r^2 < 1$ ).

**Replication in Non-Amish Studies.** Seeking additional replication and extension of our findings in non-Amish populations, we accessed publicly available GWA data from the Framingham Heart Study (FHS) (11) and the Diabetes Genetics Initiative (DGI) (12), with 1,345 and 3,082 subjects, respectively. Based on the criteria that significance levels ( $P$  values) are  $< .10$  and the associations are in



**Fig. 3.** rs6749447 genotype-specific mean SBP among AFDS subjects in the initial GWA scan (A), replication AFDS set (B), and HAPI subjects (C). The number of subjects, mean, and SEM are shown for each genotype group. Genotype-specific means were calculated using SBP adjusted for age, sex, and diabetes status. Significance levels of association to SBP are shown along with genetic models.

**Table 2. Effect sizes of LD bin 1 SNP rs6749447 and LD bin 2 SNP rs3754777 with BP in Amish and in non-Amish studies**

	Study	MAF	Sample size	Effect size (mm Hg)*			
				SBP	P†	DBP	P†
rs6749447	Amish	0.21	1,745	3.0 [1.7–4.4]	.00001	1.3 [0.4–2.1]	.003
	DGI‡	0.28	2,842	1.5 [0.4–2.7]	.02	0.1 [–0.8–1.0]	.81¶
	FHS§	0.29	1,232	1.7 [–0.4–3.8]	.08	1.1 [0.1–2.1]	.03
	GenNet	0.27	751	1.2 [–0.3–3.7]	.12	0.5 [–0.6–1.5]	.35
	Hutterites‡	0.23	555	1.9 [–0.4–4.0]	.20	1.0 [–0.7–2.7]	.69
	Combined		7,125	1.9 [1.2–2.6]	1.6e-7	0.8 [0.3–1.2]	.003
rs3754777	Amish	0.11	1,850	3.3 [1.7–5.0]	.0001	1.3 [0.2–2.3]	.02
	DGI‡	0.14	2,895	1.7 [0.3–3.1]	.03	0.7 [–0.5–1.8]	.24¶
	FHS§	0.18	1,244	1.9 [–0.2–4.0]	.12	0.9 [–0.1–1.9]	.09
	GenNet	0.17	748	1.6 [–0.2–3.4]	.09	0.6 [–0.6–1.8]	.31
	Hutterites‡	0.10	534	2.0 [–1.4–5.4]	.45	0.7 [–1.9–3.1]	.26
	Combined		7,271	2.1 [1.2–3.0]	2.3e-6	0.9 [0.4–1.4]	.001

AFDS and HAPI subjects were combined to estimate effect size.

\*Effect sizes of SBP and DBP are the mm Hg difference attributable to 1 copy of the minor allele followed by 95% confidence intervals.

†All *P* values are based on the additive model.

‡Because rs6749447 and rs3754777 were not genotyped in DGI and the Hutterites, data from rs6714609 and rs1517329, which are in near-complete or complete LD with rs6749447 and rs3754777, respectively, are presented. In both the HapMap CEU and Amish subjects,  $r^2 = 0.92$  for rs6749447 and rs6714609,  $r^2 = 1$  for rs3754777 and rs1517329.

§Effect sizes and significance levels are from examination 6.

¶Only 1,247 DGI subjects were analyzed for DBP.

the same direction as in our Amish studies, the *STK39* association was replicated with either SBP or DBP in both groups (Table 2 for LD bin 1 and 2 SNPs). In addition, 2 studies with smaller sample sizes provided modest evidence of an association between *STK39* and BP (Table 2). We observed the same trend and effect size in a sample from another founder population, the Hutterites ( $n = 575$ ) (13), but the difference did not reach significance level (LD bin 1 and 2 SNPs;  $P > .10$ ). We found the same result when we genotyped and analyzed rs6749447 and rs3754777 in the GenNet European-American subjects ( $n = 802$ ) (14).

#### Meta-Analysis Combining Association Evidence from All 6 Studies.

Among the 4 studies for which primary data were available (AFDS, HAPI, GenNet, and Hutterites), *STK39* SNPs were more significantly associated with SBP under the recessive model than under the additive model in both HAPI (Fig. 3C) and the Hutterites (LD bin 1 SNPs;  $P = .11$  and  $.20$  for recessive and additive models, respectively). For the sake of consistency, we performed meta-analysis using the additive model only, because test results based on the dominant and recessive models were unavailable for DGI and FHS. Nevertheless, combining the associations from all 6 studies ( $n > 7,000$ ) provided compelling evidence of association ( $P = 1.6 \times 10^{-7}$  for rs6749447 and  $2.3 \times 10^{-6}$  for rs3754777), with an estimated allelic effect size of 2 mmHg SBP and 1 mmHg DBP (Table 2).

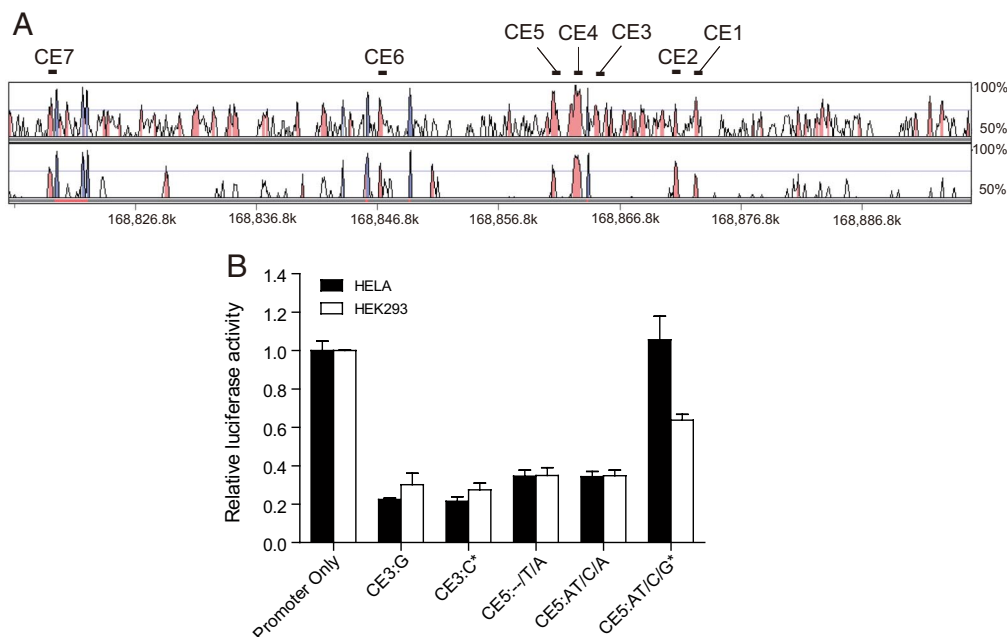
#### Identification of a Functional Candidate for the GWAS Signal.

The associated SNPs in LD bins 1 and 2 are located in a 79.5-kb region (chr2:168,699,002–168,778,544) spanning introns 1–8 and containing multiple intronic conserved elements (CEs). We sequenced all 7 exons (including exon–intron junctions) and 7 CEs, which, based on multispecies sequence alignment, are equally or more conserved evolutionarily compared with the coding exons (Fig. 4A and Table S2). Although we noted no coding or splicing variants, we identified 5 relatively common intronic polymorphisms (MAF > 0.05) within 3 CEs: rs12692877 in CE3, rs10930309 in CE4, and a dinucleotide AT deletion, rs11892008, rs35929607, in CE5. Among these, rs12692877 and rs35929607 were in complete LD with rs6749447 and rs3754777, respectively, and were strongly associated with SBP in the Amish ( $P < .00001$  with AFDS subjects). CE4 was not included in further analyses because it contained no common variant associated with BP.

Because neither of the CEs harboring BP-associated SNPs demonstrated evidence of non-protein-coding RNAs, we hypothesized that these CEs might have a regulatory function and that the sequence variants in CE3 or CE5 may influence *STK39* expression. We tested 5 luciferase reporter gene constructs, each with minimal promoter and a different allele at the polymorphic sites of CE3 and CE5, to determine whether these BP-associated variants altered luciferase expression. The constructs tested included CE3:G and CE3:C for alleles of rs12692877 and CE5:del/T/A, CE5:AT/T/A, and CE5:AT/C/G for the 3 common haplotypes of dinucleotide AT deletion, rs11892008, and rs35929607 in CE5. The major allele-bearing constructs of both CE3 and CE5 demonstrated decreased luciferase activity compared with the control vector (Fig. 4B). The presence of the G allele at rs35929607 in CE5 restored luciferase activity. In contrast, minor alleles of the other 3 sites in either CE3 or CE5 had no impact on luciferase activity. To summarize, rs35929607 is a conserved nucleotide within a highly conserved sequence element (Fig. S1). The less common G allele, associated with higher BP, enhances transcription in vitro, which is predicted to up-regulate expression of *STK39*.

The SNPs in LD bins 1 and 2 were correlated (rs6749447 and rs3754777;  $r^2 = 0.45$  in the Amish and 0.41 in the HapMap CEU subjects), and their association with BP vanished when the genotypes of another SNP from either bin were included as covariates in the analysis. Our data suggest that these SNPs are most likely surrogates for the functional candidate, rs35929607. Some SNPs in LD bin 2 (e.g., rs3754777, rs35929607) were in complete LD in the Amish ( $r^2 = 1$  in the 1,889 Amish subjects genotyped); thus, the association between SBP and rs35929607 was essentially the same as that between SBP and rs3754777, with minor differences due to genotyping completeness. In contrast, rs3754777 and rs35929607 were partially correlated in the non-Amish subjects ( $r^2 = 0.87$  in the 802 GenNet subjects genotyped). The functional SNP rs35929607 provided a more significant association than the SNPs genotyped in our GWA effort (between rs35929607 and rs6749447;  $r^2 = 0.57$ ;  $P = .04$  for rs35929607,  $.09$  for rs3754777, and  $.12$  for rs6749447), consistent with this SNP being either a better surrogate for the functional variant or the functional variant itself.

**Segment-Specific Renal Expression of *STK39*.** The identification of *STK39* as a susceptibility gene for EH raises the intriguing possi-



**Fig. 4.** (A) Pairwise alignment of human-mouse (*Top*) and human-opossum (*Bottom*) genomic sequence of the BP-associated region in *STK39*. Percent sequence identity is shown on the right. Exons are in blue; intronic CEs are in red. The locations of 7 CEs are indicated by bars above the graph. (B) Luciferase activity of CE3 and CE5 bearing different alleles. The 2 alleles associated with higher BP are indicated by an asterisk (\*).

bility that SPAK controls BP by regulating renal ion transport. Consistent with this idea, recent *in vitro* studies have demonstrated that SPAK binds to and phosphorylates NKCC2 and NCC, 2 Na<sup>+</sup>-dependent cation chloride cotransporters that mediate renal NaCl reabsorption (6). To explore whether SPAK has the capacity to regulate these transporters *in vivo*, we determined the renal expression pattern of SPAK through immunofluorescence microscopy of rat kidney sections. We observed SPAK immunostaining at the site of NKCC2 expression, the thick ascending limb (TAL) of the loop of Henle (Fig. 5A), as well as at the predominant site of NCC expression, the distal convoluted tubules (DCTs) (Fig. 5C). In addition, we observed robust SPAK expression in the cortical collecting ducts (CCDs) (Fig. 5A). The immunolocalization of SPAK to these nephron segments provides supporting evidence that higher transcriptional activity SPAK due to the presence of the G allele of rs35929607 may increase the activity of downstream NKCC2 and NCC, which would promote Na<sup>+</sup> reabsorption, thereby increasing intravascular volume and BP. More studies are needed to define the role of SPAK in BP regulation, to establish whether rs35929607 alone explains the association signal detected in the 79.5-kb region, and to determine whether the *in vitro* allele-specific transcriptional activity of CE5 is relevant to *STK39* activity and renal Na<sup>+</sup> excretion *in vivo*.

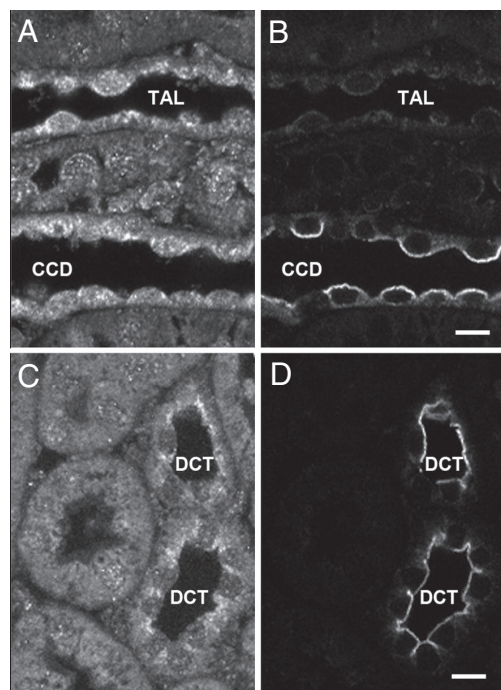
## Discussion

Based on the GWA scanning set, the initial association in *STK39* was no longer significant at the genome-wide level after Bonferroni correction was applied, adjusting for the number of SNPs analyzed. But this approach is overly conservative, because 40,000 SNPs on the 100K arrays are highly correlated with 1 or more SNPs (15). False-positive findings due to multiple testing is unlikely here, because the association between *STK39* and BP is consistent across 2 independent Amish studies and non-Amish populations, and the effect size and statistical evidence for association in the combined sample is compelling. The associations in the Amish and non-Amish studies are all in the same direction. The studies differ in the genetic models that provided the most significant association (additive vs. recessive), possibly due to increased arterial stiffness in the studies

with older subjects. Thus, the relationship between *STK39* genotypes and BP is recessive in the younger HAPI subjects but additive or dominant in the older AFDS subjects. Although the GWA approach has generated several successes for such traits as diabetes, coronary artery disease, and obesity (12, 16–19), no convincing susceptibility locus for EH has been identified to date. The initial strong GWA signal found in the Amish may have been facilitated by the relative genetic and, more importantly, lifestyle homogeneity and somewhat greater LD in this founder population compared with the general population. In fact, the association between *STK39* and BP was stronger when sodium intake was standardized (HAPI study, data not shown).

Other than *STK39*, several intriguing association signals were found elsewhere in the genome, mostly in intergenic regions (Table S1). Other than SNPs in *STK39*, only 1 association signal was located near a gene already implicated in BP regulation; rs10503798 is associated with DBP ( $P = .00034$ ) and is 1.8 kb telomeric to *ADRA1A*, the alpha 1A adrenergic receptor (20, 21). We were unable to confirm or refute any association to other known hypertension candidate genes, because most of these genes were poorly tagged by our genotyping array. In fact, only a fraction of the common variants in *STK39* were analyzed by the Affymetrix 100K array. It is very likely that there exist other common *STK39* variants that are modestly associated with BP levels in the general population and rare variants that have a greater effect on BP levels, similar to the mutations in *WNK1*, *WNK4*, *NKCC2*, and *NCC*. For example, a group of SNPs located in intron 10 of *STK39* also have been associated with BP in the AFDS, FHS, and DGI populations ( $P < .05$  in 2 or more studies; Table S3).

Further analyses revealed that several *STK39* SNPs associated with baseline BP also were associated with hypertension status in our Amish subjects, the DGI, and the Wellcome Trust Case Control Consortium (WTCCC) (even with the different criteria used to define hypertension). Among the AFDS subjects, SNPs in LD bins 1 and 2 were associated with hypertension at  $P < .0001$  (after adjusting for age and sex), and multiple SNPs in LD bin 1 also were associated with hypertension in the DGI subjects ( $P < .05$ ; data not shown). Because a GWA of BP was not carried out in the WTCCC



**Fig. 5.** Immunolocalization of SPAK in kidney. (A) SPAK strongly localizes to the TAL of the loop of Henle and the CCDs. (B) Immunolocalization of aquaporin 2 in the same section as in (A) as a segment-specific marker for CCDs. (C) Immunolocalization of SPAK in the DCTs. (D) Immunolocalization of NCC as a segment-specific marker for DCTs. Bar = 4.3  $\mu\text{m}$ .

(19), we did not include the WTCCC in our effort to replicate the Amish association with BP. Nevertheless, several *STK39* SNPs were associated with hypertension in the WTCCC at  $P < .05$ ; for example, rs3754776, located within our associated region and  $< 20$  kb from the putative functional SNP rs35929607, was associated with hypertension status ( $P = .003$  for the additive models and .0002 for the general models) in the WTCCC. This SNP was in partial LD ( $r^2 = 0.37\text{--}0.38$ ) with the bin 1 SNPs but not with the bin 2 SNPs ( $r^2 = 0.01\text{--}0.02$ ) in both the Amish and the HapMap CEU samples. All of the published studies with hypertension GWA results, including ours, recruited based on either diabetes or hypertension; thus, an unbiased estimate of hypertension risk in the general population associated with the *STK39* genotype remains to be determined.

Besides renal electrolyte transport, SPAK has been linked to such functions as cytoskeleton rearrangement, cell differentiation, transformation, and proliferation (5). In fact, in a subset of Amish subjects with extensive metabolic syndrome-related phenotypes, the BP-associated *STK39* SNPs were modestly associated with fasting glucose level, insulin response to glucose, and plasma triglyceride level (data not shown). Furthermore, *STK39* was found to be located in a genomic region in which multiple BP, obesity, and diabetes-related rodent quantitative trait loci have been mapped [information provided by Mouse Genome Informatics (<http://www.informatics.jax.org>) and the Rat Genome Database (<http://rgd.mcw.edu/>)]. Whether or not *STK39* is involved in other metabolic syndrome traits, after adjusting for its effect on BP, awaits further study. In addition, several SNPs within *STK39* have demonstrated an association with autism (22). Although this association has not yet been replicated, it is intriguing, because several known and predicted phosphorylation targets of SPAK (e.g., NKCC1 and KCC2) are expressed in neuronal tissues (23).

A recent study based on the FHS subjects found that those heterozygous for rare variants in *NCCT*, *NKCC2*, and *ROMK* (frequency  $< 0.0001$ ) had a clinically significant reduction in BP and lower hypertension risk, demonstrating that at least some

fraction of BP variation in the general population is due to rare, independent alleles (24). It is noteworthy that our agnostic GWA study identified a gene that unifies in a single pathway the proteins involved in rare, monogenic forms of hypotension/hypertension, such as Bartter's syndrome and Gordon's syndrome, and that of common EH. *STK39* has not been previously implicated as a susceptibility gene for EH, but its involvement in BP regulation is supported by our current understanding of this protein and its interaction partners. In summary, our data establish that SNPs within the *STK39* gene are strongly associated with BP. The real value of this study is not the identification of SNPs associated with BP levels or EH susceptibility in the general populations; rather, our findings highlight the importance of identifying all common and rare variants in this gene that might influence BP regulation, as well as the need for additional biochemical and physiological studies of its role in BP-related pathways. For example, the *STK39* genotype might be a predictor of EH patients who are more likely to respond to salt reduction or a specific type of diuretic as a measure to control BP. SPAK itself might be an excellent target for novel antihypertensive drug therapies.

## Materials and Methods

**Study Subjects and Phenotypes AFDS.** Recruitment for the AFDS was initiated in 1995 with the goal of identifying susceptibility genes for type 2 diabetes (T2DM) and related traits. The study protocol was approved by the University of Maryland School of Medicine's Institutional Review Board, and informed consent was obtained from each study participant. Details of the AFDS design, recruitment, and pedigree structure have been described previously (4). From the entire AFDS, 542 subjects (119 with T2DM, 132 with impaired glucose tolerance, and 291 with normal glucose tolerance) were genotyped using the Affymetrix 100K set (25). The remaining subjects ( $n = 557$ ) was subsequently genotyped to confirm the *STK39* findings (Table 1).

**HAPI Heart Study.** This study was initiated in 2002 to measure the cardiovascular response to 4 short-term interventions affecting cardiovascular health and to identify the genetic and environmental determinants of these responses (10). Before any intervention, baseline BP was obtained manually with the subject in the sitting position after he or she had been sitting quietly for 5 min, and the average of 3 of these measurements was used for analysis. Hypertension medication was discontinued in all HAPI subjects before the start of the study.

**FHS.** Recruitment of men and women from the town of Framingham, Massachusetts began in 1948 with the purpose of investigating the causes of cardiovascular disease and related traits. FHS investigators have recently published results of a 100K GWA study based on 1,345 subjects (310 pedigrees), 17 phenotypes, and the Affymetrix GeneChip Mapping 100K Array (11, 26). All data were available through dbGaP. For replication of our *STK39* signals, we used SBP and DBP from examinations 1–7 and the average of SBP/DBP from available examinations 1–7.

**DGI.** The DGI performed a GWA study in 1,464 patients with T2DM and 1,467 controls to search for genetic variants that influence the risk for diabetes and related traits, such as blood glucose and lipid levels, obesity, and BP (12). For the analysis of BP-related traits, cases and controls were combined. The numbers of subjects analyzed for SBP, DBP, and EH status were 2,895, 1,247, and 3,082, respectively. The size of the DBP sample was considerably smaller than the other 2 samples because subjects over age 60 years were excluded from the analysis. For replication,  $P$  values after correction by genomic control were used, and regression coefficients (labeled as BETA) were used to determine the direction of association.

**Hutterites.** The Hutterites are a religious isolate that originated in the Tyrolean Alps in the 1500s. In the 1870s, approximately 900 Hutterites moved to what is now South Dakota, and their descendants now live in  $> 350$  communal farms in the northern United States and Canada. Because of their communal lifestyles, Hutterites share a relatively uniform environment. The 583 subjects (none of whom take antihypertensive medication) in this study are descendants of 62 ancestors and are related to one another in a 13-generation pedigree (13). To test for the effects of each SNP on SBP and DBP, the general 2-allele model test of association was used in the entire pedigree, with all inbreeding loops kept intact, as described previously (27).

**GenNet.** Details of the GenNet subjects have been published previously (14, 28). We analyzed 802 European-American subjects who were not taking antihypertensive medication at the time of the study, collected from Tecumseh, Michigan.

**Genotyping and Genotype Data Quality Control.** For the initial genome-wide association scan conducted in the AFDS samples, genomic DNA from leukocytes was genotyped using the Affymetrix GeneChip Mapping 100K Array set. The genotyping

protocol and quality control procedures used to identify and remove poor-quality 100K data were described previously (25). More details are given in *SI Materials and Table S4*.

**Sequencing.** The 79.5-kb BP-associated region of *STK39* contains 7 exons (exons 2–8) and 7 intronic CEs. We defined noncoding CEs as intronic elements with sequence identity as high as *STK39* coding exons based on a lod score calculated using the PhastCons program (PhastCons Conserved Elements, 17-Way Vertebrate Multiz Alignment, provided by the UCSC Genome Browser) (29). A total of 47 samples were selected based on the rs6749447 genotype and residual BP levels, including 9 homozygotes of the major allele with the lowest SBP residual, 9 heterozygotes with the highest SBP residual, 9 homozygotes of the minor allele with the highest SBP residual, 10 subjects with the highest overall SBP residuals, and 10 subjects with the lowest overall SBP residuals. Pedigree structure was taken into account for sample selection, to minimize relatedness. Exons and CEs were amplified (primer sequences available on request) and sequenced using BigDye version 3.1 and an ABI 3700 DNA sequencer, and then analyzed using Sequence Analysis 3.2 software (all from Applied Biosystems). Sequencher 4.5 (GeneCodes) was used to align individual sequences and identify variants.

**Reporter Gene Assay.** PCR products from 5 individuals with known homozygous genotypes were cloned into the reporter vector (pGL4.23[Luc2/minP], Promega) using an In-Fusion 2.0 Dry-Down PCR cloning kit (BD Biosciences). The following CEs were cloned and tested: CE3:G and CE3:C for CE3 with different alleles of rs12692877 and CE5:delT/A, CE5:AT/T/A, and CE5:AT/C/G for CE5 with 3 haplotypes involving a dinucleotide (AT) deletion, rs11892008 (C/T), and rs35929607 (A/G) that we observed in the Amish. Haplotype delT/G was not found in the Amish population and was not tested. All recombinant plasmids were sequenced to confirm insert sequence and orientation. For each construct, HEK-293 and HeLa cells (obtained from American Type Culture Collection and cultured according to recommended condition) were cotransfected with 2  $\mu$ g of plasmid DNA, 0.1  $\mu$ g of pGL4.73 (Renilla luciferase), and 7  $\mu$ L of FuGENE HD reagent (Roche), to normalize for transfection efficiency. After transfection and a 24-hour incubation, cells were lysed and luciferase signals were assayed using the Dual-Luciferase Reporter Assay System (Promega) and a TD-20/20 Luminometer (Turner Designs). Relative luciferase activity is given in arbitrary units corrected for Renilla luciferase activity and normalized to the control. All luciferase experiments were performed in duplicate at least twice.

**Statistical Analysis.** Association analyses of BP-related traits were performed using the measured genotype approach that models variation in the trait of interest as a function of measured environmental covariates, measured genotype, and a polygenic component to account for phenotypic correlation due to relatedness. A (n-1)-degree-of-freedom *t* test was used to assess the significance of the measured genotype. Age, age<sup>2</sup>, sex, and T2DM status were included as covariates, and the SNP was coded using additive, dominant, and recessive

modes. When analyzing the Amish subjects, the polygenic component was modeled using the relationship matrix derived from the complete 14-generation pedigree structure, to properly control for the relatedness of all subjects in the study (data not shown). Q-Q plots comparing the observed GWA *P* values with those expected under the null are provided in Fig. S2. The genomic inflation factors ( $\lambda$ ) based on *T*-scores were 0.98 for SBP and 1.02 for DBP, indicating that systematic inflation of our association signals due to undetected genotyping error or hidden family relationship was highly unlikely.

Effect sizes in the Amish studies were computed after the AFDS and HAPI subjects were combined. Age, age<sup>2</sup>, sex, T2DM status, and study designation (AFDS vs. HAPI) were included as covariates, and genotype was coded based on an additive genetic model.

Pairwise *r*<sup>2</sup> values between SNPs were calculated using Haploview (30). LD bin assignment was based on HapMap CEU genotypes, which provided an LD structure of the *STK39* region nearly identical to that observed in the Amish.

The BP association results for FHS and DGI were assessed from the websites referenced from their respective published studies (11, 12). The directionality of the association was determined for each study from either the family-based association test (FBAT) statistics or BETA. Study-specific adjustment for hypertension medication was used when evaluating association signals from non-Amish studies. A weighted z-score-based fixed-effects meta-analysis method was used to combine results across all studies, using the METAL program (<http://www.sph.umich.edu/csg/abecasis/Metal/index.html>). In brief, for each SNP, a reference allele was identified, and a z-statistic summarizing the magnitude of the *P* value for association (under the additive model) and direction of effect was generated for each study. An overall z-statistic was computed as a weighted average of the individual statistics, and then a corresponding *P* value for that statistic was computed. The weights were proportional to the square root of the sample size in each study and scaled such that the squared weights summed to 1.

**Immunolocalization of SPAK.** Kidneys from adult Sprague-Dawley rats were fixed by retrograde perfusion and embedded in paraffin. Once sections (3  $\mu$ m) were picked up on cover slips, heat-induced target retrieval using a citrate buffer (pH 8) was used to unmask epitopes. Sections were then washed and incubated overnight with primary rabbit antibodies to SPAK (Abgent, San Diego, CA) at a concentration of 5  $\mu$ g/mL at 4 °C, followed by secondary antibodies as described previously (31).

**ACKNOWLEDGMENTS.** This work was supported by National Institutes of Health (NIH) research grants R01 DK54261–07, R01 HL076768–02, U01 HL72515–01, R01 HL088120, U01 HL084756, and DK32839. We thank John Shelton and Kathy Ryan for their expert genotyping and data support efforts. We also thank Dr. Martin Larson for providing the FHS phenotype data for effect size estimate, Dr. Lei Lu for assisting with genomic inflation, and Dr. Kirby D. Smith for providing valuable discussions. We gratefully acknowledge the Amish liaisons and fieldworkers, as well as the extraordinary cooperation and support of the Amish community, without whom these studies would not be possible.

- Kearney PM, et al. (2005) Global burden of hypertension: Analysis of worldwide data. *Lancet* 365:217–223.
- Lifton RP, Gharavi AG, Geller DS (2001) Molecular mechanisms of human hypertension. *Cell* 104:545–556.
- Wilson FH, et al. (2001) Human hypertension caused by mutations in WNK kinases. *Science* 293:1107–1112.
- Hsueh WC, et al. (2000) Diabetes in the Old Order Amish: Characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care* 23:595–601.
- Delpire E, Gagnon KB (2008) SPAK and OSR1: STE20 kinases involved in the regulation of ion homeostasis and volume control in mammalian cells. *Biochem J* 409:321–331.
- Moriguchi T, et al. (2005) WNK1 regulates phosphorylation of cation-chloride-coupled cotransporters via the STE20-related kinases, SPAK and OSR1. *J Biol Chem* 280:42685–42693.
- Simon DB, et al. (1996) Bartter's syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. *Nat Genet* 13:183–188.
- Simon DB, et al. (1996) Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter. *Nat Genet* 12:24–30.
- Vitari AC, Deak M, Morrice NA, Alessi DR (2005) The WNK1 and WNK4 protein kinases that are mutated in Gordon's hypertension syndrome phosphorylate and activate SPAK and OSR1 protein kinases. *Biochem J* 391:17–24.
- Mitchell BD, et al. (2008) The genetic response to short-term interventions affecting cardiovascular functions: Rationale and design of the HAPI Heart Study. *Am Heart J* 155:823–828.
- Levy D, et al. (2007) Framingham Heart Study 100K Project: Genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet* 8(Suppl 1):S3.
- Saxena R, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–1336.
- Ober C, et al. (1997) HLA and mate choice in humans. *Am J Hum Genet* 61:497–504.
- Chang YP, et al. (2007) Multiple genes for essential hypertension susceptibility on chromosome 1q. *Am J Hum Genet* 80:253–264.
- Nicolae DL, Wen X, Voight BF, Cox NJ (2006) Coverage and characteristics of the Affymetrix GeneChip Human Mapping 100K SNP set. *PLoS Genet* 2:e67.
- Scott LJ, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345.
- McPherson R, et al. (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316:1488–1491.
- Frayling TM, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894.
- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
- Gu D, et al. (2006) Association of alpha1A adrenergic receptor gene variants on chromosome 8p21 with human stage 2 hypertension. *J Hypertens* 24:1049–1056.
- O'Connell TD, et al. (2006) Alpha1-adrenergic receptors prevent a maladaptive cardiac response to pressure overload. *J Clin Invest* 116:1005–1015.
- Ramoz N, Cai G, Reichert JG, Silverman JM, Buxbaum JD (2008) An analysis of candidate autism loci on chromosome 2q24–q33: Evidence for association to the *STK39* gene. *Am J Med Genet B Neuropsychiatr Genet* 147B:1152–1158.
- Richardson C, Alessi DR (2008) The regulation of salt transport and blood pressure by the WNK-SPAK/OSR1 signalling pathway. *J Cell Sci* 121:3293–3304.
- Ji W, et al. (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 40:592–599.
- Rampersaud E, et al. (2007) Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: Evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 56:3053–3062.
- Cupples LA, et al. (2007) The Framingham Heart Study 100K SNP genome-wide association study resource: Overview of 17 phenotype working group reports. *BMC Med Genet* 8 (Suppl 1):S1.
- Abney M, Ober C, McPeck MS (2002) Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: Fasting serum insulin level in the Hutterites. *Am J Hum Genet* 70:920–934.
- FBPP Investigators (2002) Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* 39:3–9.
- Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15:1034–1050.
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
- Coleman RA, Liu J, Wade JB (2006) Use of anti-fluorophore antibody to achieve high-sensitivity immunolocalizations of transporters and ion channels. *J Histochem Cytochem* 54:817–827.