Observational Study

Medicine®

OPEN

Contributions of rare coding variants in hypotension syndrome genes to population blood pressure variation

Priyanka Nandakumar, PhD[a,b], Alanna C. Morrison, PhD[c], Megan L. Grove, MS[c], Eric Boerwinkle, PhD[c,d], Aravinda Chakravarti, PhD[a,e,*]

Abstract

Rare variants, in particular renal salt handling genes, contribute to monogenic forms of hypertension and hypotension syndromes with electrolyte abnormalities. A study by Ji et al (2008) demonstrated this effect for rare loss-of-function coding variants in *SLC12A3* (NCCT), *SLC12A1* (NKCC2), and *KCNJ1* (ROMK) that led to reduction of ~6 mm Hg for SBP and ~3 mm Hg for DBP among carriers in 2492 European ancestry Framingham Heart Study (FHS) subjects. These findings support a potentially large role for these variants in interindividual variation in systolic and diastolic blood pressure (SBP, DBP) in the population. The present study focuses on replicating the analyses completed by Ji et al to identify effects of rare variants in the population-based Atherosclerosis Risk in Communities (ARIC) study.

We attempted to replicate the findings by Ji et al by applying their criteria to identify putative loss-of-function variants with allele frequency <0.001 and complete conservation across a set of orthologs, to exome sequencing data from 7444 European ancestry participants of the ARIC study.

Although we failed to replicate the previous findings when applying their methods to the ARIC study data, we observed a similar effect when we restricted analyses to the subset of variants they observed.

These results simultaneously support the utility of exome sequencing data for studying extremely rare coding variants in hypertension and underscore the need for improved filtering methods for identifying functional variants in human sequences.

Abbreviations: AA = African-American, AAF = alternate allele frequency, ANNOVAR = annotate variation, ARIC = Atherosclerosis Risk in Communities, BCM-HGSC = Baylor College of Medicine Human Genome Sequencing Center, BMI = body mass index, BP = blood pressure, CHARGE = Cohorts for Heart and Aging Research in Genomic Epidemiology, DBP = diastolic blood pressure, EA = European-American, ExAC = Exome Aggregation Consortium, FHS = Framingham Heart Study, GWAS = genome-wide association study, HTN = hypertension, HWE = Hardy-Weinberg Equilibrium, KING = Kinship-based INference, LOF = loss-of-function, NFE = non-Finnish European, PANTHER = Protein ANalysis THrough Evolutionary Relationships, PolyPhen-2 = Polymorphism Phenotyping v2, SBP = systolic blood pressure, SD = Standard deviation, SIFT = sorts intolerant from tolerant, SKAT = sequence-kernel association test, SNV = single nucleotide variant, WES = whole exome sequencing.

Keywords: blood pressure, epidemiologic studies, exome, genomics, hypotension

Editor: Ovidiu Constantin Baltatu.

Funding/Support: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). DNA sequencing for this study was carried out at the Baylor College of Medicine Human Genome Sequencing Center and supported by the National Human Genome Research Institute grants U54 HG003273 and UM1 HG008898. This study was funded by NIH grant HL086694 to A.C.

Sources were received from National Institutes of Health (NIH). The funders had no role in study design, collection of data, execution or analysis, or manuscript preparation or review.

The authors have no conflicts of interest.

Supplemental Digital Content is available for this article.

[a] Center for Complex Disease Genomics, [b] Predoctoral Training Program in Human Genetics and Molecular Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, [c] Human Genome Sequencing Center, Baylor College of Medicine, [d] Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, [e] Center for Human Genetics and Genomics, NYU School of Medicine, New York, NY.

[*] Correspondence: Aravinda Chakravarti, Center for Human Genetics and Genomics, NYU School of Medicine Science Building, Room 802/803, 435 E 30th Street New York, NY 10016 (e-mail: aravinda.chakravarti@nyumc.org).

Medicine (2018) 97:33(e11865)

Received: 2 March 2018 / Accepted: 25 July 2018

http://dx.doi.org/10.1097/MD.0000000000011865

1. Introduction

Hypertension (HTN) is a major risk factor for cardiovascular disease and affects ~30% of adults worldwide.[1,2] Recently, genome-wide association studies (GWAS) have identified common variants at ~166 loci associated with systolic blood pressure (SBP), diastolic blood pressure (DBP), and essential HTN, but these variants explain <5% of the phenotypic variance.[3–27] In contrast, rare variants in ~20 genes involved in renal salt handling and water balance have been implicated in monogenic forms of either HTN or hypotension with electrolyte abnormalities.[28,29] However, the contribution of these genes to population level interindividual variation in SBP and DBP is generally unknown. As the existing GWAS loci do not include any of the 20 known HTN or hypotension syndromic genes, it is reasonable to infer that common variation in these genes do not contribute greatly to interindividual BP variation. The question is, why?

In 2008, Ji et al[29] examined the effects of rare variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* on BP in the European ancestry Framingham Heart Study (FHS) subjects. The authors chose these ion channel genes because homozygotes for loss-of-function (LOF) variants in these diuretics targets lead to recessive renal salt-wasting hypotension syndromes [Bartter (prevalence ~1/1,000,000[29,30]) and Gitelman (prevalence ~1/40,000[29,31])], so that, consequently, LOF heterozygotes in these genes could reduce BP substantially and be protective against HTN. Ji et al[29] identified 18 not previously validated and potentially LOF missense variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* in the FHS offspring cohort and demonstrated a significant protective effect (−6.3 mm Hg for SBP and −3.4 mm Hg for DBP) in their carriers in all ages between 25 and 60 years. The expected rate of carriers for these genes is ~1% of the population; the study by Ji et al[29] identified such carries in ~1.5% of their sample.

We attempted to investigate *SLC12A3*, *SLC12A1*, and *KCNJ1* using identical methods but using whole exome sequence (WES) data on 7810 European ancestry subjects from the population-based Atherosclerosis Risk in Communities (ARIC) cohort. Like the FHS, ARIC is a longitudinal study with BP measurements over time and, therefore, allows for identical analyses as in the study by Ji et al.[29] Success in such studies could motivate genome-wide screens for genes with analogous effects.

2. Methods

2.1. Study participants

The ARIC (Atherosclerosis Risk in Communities) study is a population-based, prospective study on 15,792 individuals, including 11,478 European and 4266 African ancestry US subjects from Forsyth County, NC, Jackson, MI, Minneapolis, MN, and Washington County, MD.[32,33] There have been 5 visits, with individuals in the first visit (1987–1989) aged between 45 and 64 years. Subsequent visits occurred in 1990 to 1992, 1993 to 1995, 1996 to 1998, and 2011 to 2013. The data from the first 4 visits were used in this study. SBP and DBP were measured thrice at each of the first 3 visits, and twice at the fourth visit, using a random zero sphygmomanometer; the average of the (final) 2 measurements at each visit were used for analysis. In this study, exome sequence data from 7444 European–Americans (EAs) were analyzed (see below). All participants provided written consent, and approval was obtained from the appropriate institutional review boards.

2.2. Whole exome sequencing, variant calling, and quality control

ARIC samples were sequenced at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC), as part of a larger set of CHARGE[34] samples. The exome sequencing protocol, variant calling, and quality control procedures are described elsewhere (Yu et al).[35] Among 14,443 CHARGE samples in the final cleaned set, which had mean depth of 78x coverage, there were 7810 EA and 3180 African–American (AA) ARIC subjects. Only the 7810 EA samples were utilized in our analysis for replication.

We then applied additional stringent filters to the data from these 7810 individuals, and the following observations were culled: individuals with <90% call rate; variants with >10% missing genotype calls among samples; and variants failing Hardy–Weinberg Equilibrium ($P < 1 \times 10^{-6}$). We used genome-wide genotype data to estimate the genetic relationships among the remaining 7767 individuals using KING,[36] retaining only those individuals with third-degree or more remote relationships. This provided a final dataset comprising 7444 individuals.

2.3. Annotation to identify variants of interest

Annotation of variants in *SLC12A3*, *SLC12A1*, and *KCNJ1*, with respect to their predicted deleterious effects, was carried out using ANNOVAR[37] with refGene annotations, to identify putative LOF missense variants ("nonsynonymous") in the selected transcripts used in the downstream analyses.

2.4. Phenotypes

Longitudinal SBP and DBP phenotypes were calculated on the basis of previously described methods,[29,38] using the first 4 visits from the ARIC study. Briefly, SBP and DBP were adjusted in a cubic regression on age within each age group (<35, 35–44, 45–54, 55–64, 65–74, and 75+ years) for the 7444 individuals, and then used to adjust their measurements for those visits that were taken while on antihypertensive medications. Although all SBP measurements were used, DBP measurements were restricted to those taken at age 55 years and below, as DBP is known to decline with increasing age beyond this point.[39] Subsequently, these residuals were adjusted for mean age in a linear regression, separately by sex; in our study, this was done with all individuals with at least 1 visit, regardless of the time span between the first and last visits. This differs from previously described methods in the study by Levy et al[38] where at least 4 examinations were required with a minimum time span of 10 years between the first and final examinations in the FHS and at least 3 examinations in the FHS Offspring Cohort studies. The resulting standardized residuals were used as phenotypes, and the mean of the nonstandardized residuals among carriers of analyzed variants is presented as the effect size in Figs. 1 and 2.

2.5. Transcripts, orthologs, and paralogs selection and alignment

Methods for ortholog and paralog selection were as previously described[29]; they are briefly restated here emphasizing pertinent differences, summarized in Table S1, Supplemental Digital Content, http://links.lww.com/MD/C403. In the original study, the representative transcripts NM_000339 (*SLC12A3*), NM_000338 (*SLC12A1*), and NM_000220/NM_153764-7

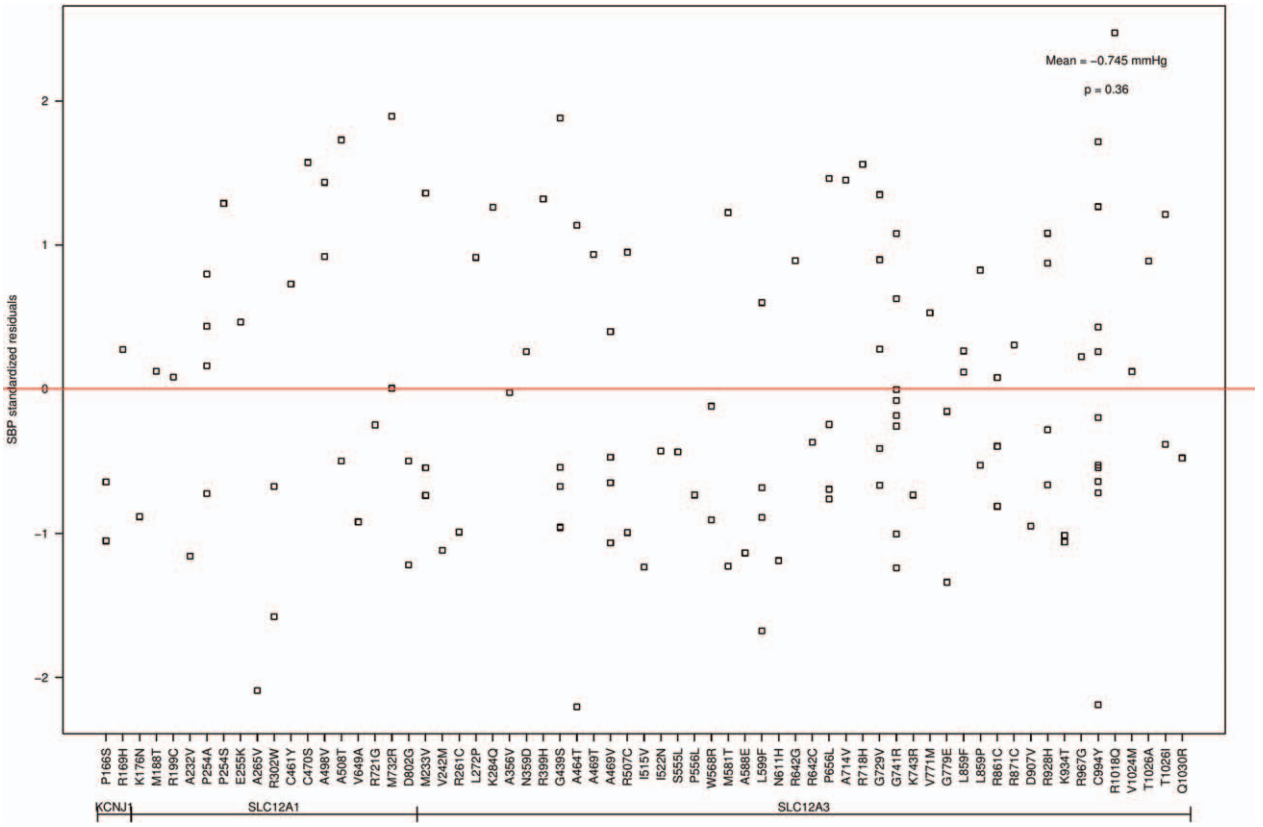

Figure 1. Age- and sex-adjusted SBP residuals for 121 carriers of 65 single nucleotide coding variants depicting the mean effect of carriers.

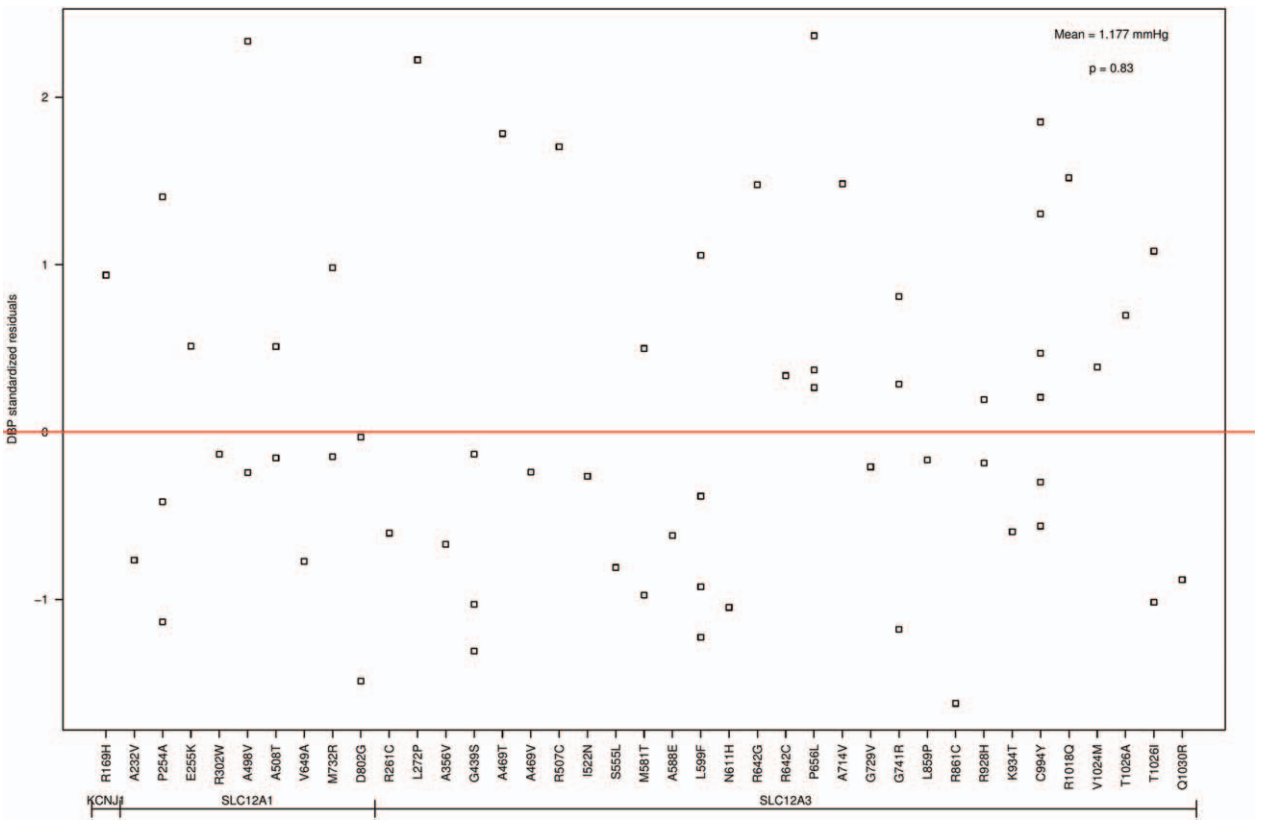

Figure 2. Age- and sex-adjusted DBP residuals for 62 carriers of 39 single nucleotide coding variants depicting the mean effect of carriers.

(*KCNJ1*) were analyzed; in our study, we analyzed the same transcripts for *SLC12A3* and *SLC12A1*, and analyzed the transcript NM_153766 for *KCNJ1*. This *KCNJ1* transcript encodes an identical protein sequence to those encoded by other transcript variants NM_153764 and NM_153767, and presented with maximum positional similarity through multiple alignment to variants listed as conserved in Ji et al,[29] as compared with the canonical transcript, NM_000220.

The orthologs analyzed in the study by Ji et al[29] for *SLC12A3* were from human, mouse, rat, rabbit, dog, cow, chicken, zebrafish, and winter flounder. In our study, the obsolete record XP_871112 from cow was replaced with NP_001193107. The *SLC12A1* orthologs analyzed in their study were from human, mouse, rat, rabbit, dog, chicken, zebrafish, and *Tetraodon*. The record NP_062007 from rat was removed from Refseq records, as it is a nonsense-mediated mRNA decay (NMD) candidate, and the record XP_850426 from dog was removed from Refseq records as part of standard genome annotation processing (SGAP). These orthologs were dropped in our study as well. Finally, the orthologs studied for *KCNJ1* in the study by Ji et al[29] were from human, mouse, rat, dog, chicken, cow, frog, zebrafish, fugu, *Caenorhabditis elegans*, sea urchin, and *Drosophila melanogaster*. In our study, XP_546403 from dog and XP_425795 from chicken were dropped from Refseq records for SGAP, XP_684541 from zebrafish was replaced with the updated record NP_001092204, XP_585917 from cow was replaced with the updated record NP_0011179136, and the human protein NP_722450 was used, corresponding to the selected transcript for *KCNJ1*.

The paralogs analyzed for *SLC12A3* and *SLC12A1* (*SLC12A* family) in both studies were from *SLC12A1*, *SLC12A3*, and additionally, *SLC12A2*, *SLC12A4*, *SLC12A5*, *SLC12A6*, and *SLC12A7*. The paralogs analyzed for KCNJ1 (*KCNJ* family) in their study include *KCNJ2*, *KCNJ3*, *KCNJ4*, *KCNJ5*, *KCNJ6*, *KCNJ8*, *KCNJ9*, *KCNJ10*, *KCNJ11*, *KCNJ12*, *KCNJ13*, *KCNJ14*, *KCNJ15*, and *KCNJ16*. In our study, as for the orthologs, we used the human protein NP_722450 in accordance with the *KCNJ1* transcript analyzed as stated above, and additionally, NP_733838 from *KCNJ14* was removed due to inadequate support for its transcript, NM_170720.1 (http://www.ncbi.nlm.nih.gov/nuccore/).

Multiple alignments were performed with ClustalW v2.1[40,41] with default parameters for all orthologs per gene, and all paralogs per gene family (1 for the *KCNJ* family and 1 for the *SLC12A* family), with 5 alignments in total. This differs slightly from the original alignment procedure described in Ji et al,[29] in which it was stated that pairwise alignments were used in addition to multiple alignments to determine conserved positions. While the many pairwise alignments may have some differences to a multiple sequence alignment, this difference is not expected to lead to any significant change in our downstream assessment of overall conservation of residues.

2.6. Selection criteria and "validation" annotation:

We studied missense variants only because no nonsense variants ultimately met the filtering criteria of frequency or conservation. The criteria used by Ji et al[29] to identify putative LOF missense variants included allele frequency <0.001 and complete conservation across the selected orthologs. The exceptions to this were variants that were conserved in orthologs, but with the mutant residue present in paralogs, as these were thought to be sustainable within the species. Also, as in their study, we further annotated the analyzed variants from ARIC and the SNVs from FHS using PANTHER[42] (v9.0), SIFT[43] (with GRCh37/Ensembl 63), and PolyPhen-2[44] to determine predicted deleterious effects of variants.

2.7. Statistical analyses

All individuals in this study were unrelated or related more remotely than third-degree relatives. Standard 1-tailed *t* tests, under the assumption that the alternate alleles are blood pressure (BP)-lowering, were carried out on standardized residuals as final phenotype to compare carriers with noncarriers, assuming equal variance for both groups. Noncarriers were defined as those who have nonmissing genotypes for all 65 variants; those with missing genotypes for any of the 65 were excluded from all calculations.

2.8. Protein plots

Protein plots in Figures S1–S3, Supplemental Digital Content, http://links.lww.com/MD/C403, were created with the Protter software.[45]

3. Results

We examined rare missense variants in 7444 EA ARIC subjects with WES data. The phenotypic (BP) and risk factor (age, sex, BMI) characteristics for these individuals are summarized in Table 1. We identified a total of 216 missense variants (rate of 0.029 per individual) in the cleaned exome sequence data for *SLC12A3*, *SLC12A1*, and *KCNJ1*. To assess their properties, we aligned these genes to their orthologs and paralogs as described in the Methods section with ClustalW2, and discovered overall conservation (protein sequence identity) of 41.5%, 40.6%, and 18.3% across orthologs of *SLC12A3*, *SLC12A1*, and *KCNJ1*, respectively, across all species considered in each of the multiple alignments. Of these, 65 variants (46, 17, and 2 for *SLC12A3*, *SLC12A1*, and *KCNJ1*, respectively) had alternate allele frequencies (AAFs) <0.001 in the 7444 ARIC EA samples. These occurred at highly sequenced bases (median depth 75X; range: 27–196X in 7810 ARIC EA individuals) demonstrating high data quality.

We annotated these 65 variants with SIFT, PANTHER, and Poly-Phen2, which confirmed that their vast majority (57, or

Table 1

Clinical characteristics of 7444 ARIC subjects analyzed in this study.

Visit/Phase	Age (SD)	BMI (SD)	SBP unadjusted (SD)	SBP adjusted (SD)	DBP unadjusted (SD)	DBP adjusted (SD)	% medicated
1	54.288 (5.665)	26.967 (4.815)	118.204 (16.487)	121.848 (19.649)	71.765 (9.678)	73.055 (10.416)	18.8
2	57.223 (5.64)	27.285 (4.887)	119.754 (17.299)	124.107 (21.135)	71.356 (9.698)	73.096 (10.379)	20.6
3	60.289 (5.609)	27.91 (5.141)	122.597 (17.696)	128.491 (22.436)	70.779 (9.932)	74.032 (10.908)	25.6
4	63.095 (5.592)	28.25 (5.209)	125.916 (18.187)	133.533 (23.517)	70.048 (9.823)	73.092 (10.587)	30.9

Of 7444 individuals (53.76% female), SBP residuals were calculated for 7443 and DBP residuals were calculated for 3835 subjects.
BMI = body mass index, DBP = diastolic blood pressure, SBP = systolic blood pressure.

Table 2

Comparison of analyzed SNVs in extracellular, intracellular, and transmembrane domains across the *SLC12A3*, *SLC12A1*, and *KCNJ1* proteins for ARIC and FHS.

Number of variants	ARIC	FHS	Total
No. of var extracellular	16	5	21
No. of var intracellular	31	15	46
No. of var Transmembrane	18	8	26
Total	65	28	93

ARIC = Atherosclerosis Risk in Communities, FHS = Framingham Heart Study .

86.2%) were predicted to be pathogenic by all three prediction programs (Table S2, Supplemental Digital Content, http://links.lww.com/MD/C403). We also used these versions of the software programs to obtain updated annotations for the 28 FHS SNVs, enabling comparisons between the 2 sets of variants (Table S3, Supplemental Digital Content, http://links.lww.com/MD/C403). All but 3 variants (89.3%) were predicted to be pathogenic by all three programs. Further, allele frequencies of the 65 ARIC variants in non-Finnish Europeans from the Exome Aggregation Consortium (ExAC) are shown in Table S4, Supplemental Digital Content, http://links.lww.com/MD/C403, and those for the 28 FHS SNVs are presented in Table S5, Supplemental Digital Content, http://links.lww.com/MD/C403. There are 40 of 65 ARIC variants (~61.5%) and 17 of 28 FHS SNVs (60.7%) with nonzero non-Finnish ExAC allele frequencies, providing similar evidence for the existence of these genotype calls across the 2 studies [Fisher exact test in R reported as $P=1$ (due to rounding)]. Further, examining the variant distributions across the 3 genes (Figures S1–S3, Supplemental Digital Content, http://links.lww.com/MD/C403) shows them to be evenly distributed across the domains in both studies (Fisher exact test $P=.80$, Table 2).

Longitudinal SBP and DBP values were then analyzed by standard 1-tailed *t* tests to test for differences between carriers and noncarriers of the variants. The standardized residual phenotype values of carriers, modeled after Fig. 2 in Ji et al,[29] are shown in Figs. 1 and 2 for SBP and DBP, respectively. The mean effect was -0.745 mm Hg for SBP (*t* test $P=.36$) considering all 65 variants in 121 carriers and 6750 noncarriers, and 1.177 mm Hg for DBP (*t* test $P=.83$) considering 39 variants in 62 carriers and 3478 noncarriers. As only DBP measurements under the age of 55 years were analyzed in this study (see Methods), there were fewer variants in fewer individuals analyzed. Thus, these results neither show a prominent direction of effect, nor are they statistically significant.

We also examined a set of 10 variants from the 28 FHS SNVs that were also present in the ARIC exome sequencing dataset. Of these, 9 met the conservation and AAF criteria within ARIC (except the proline at position 348 in *SLC12A1* has an alanine substitution in *Danio rerio*). Standardized residuals for carriers of all 10 variants are shown for SBP, and for carriers of the subset of 8 variants with DBP measurements available under age 55, in Figs. 3 and 4, respectively. These figures indicate that the primary positive effect sizes among this small subset of variants are from the P254A and G741R variants. Overall, these results clearly replicate those for the 30 variants in the study by Ji et al[29] with similar mean effect sizes among carriers (SBP: −6.888 mm Hg, $P=.02$; DBP: −3.120 mm Hg, $P=.11$).

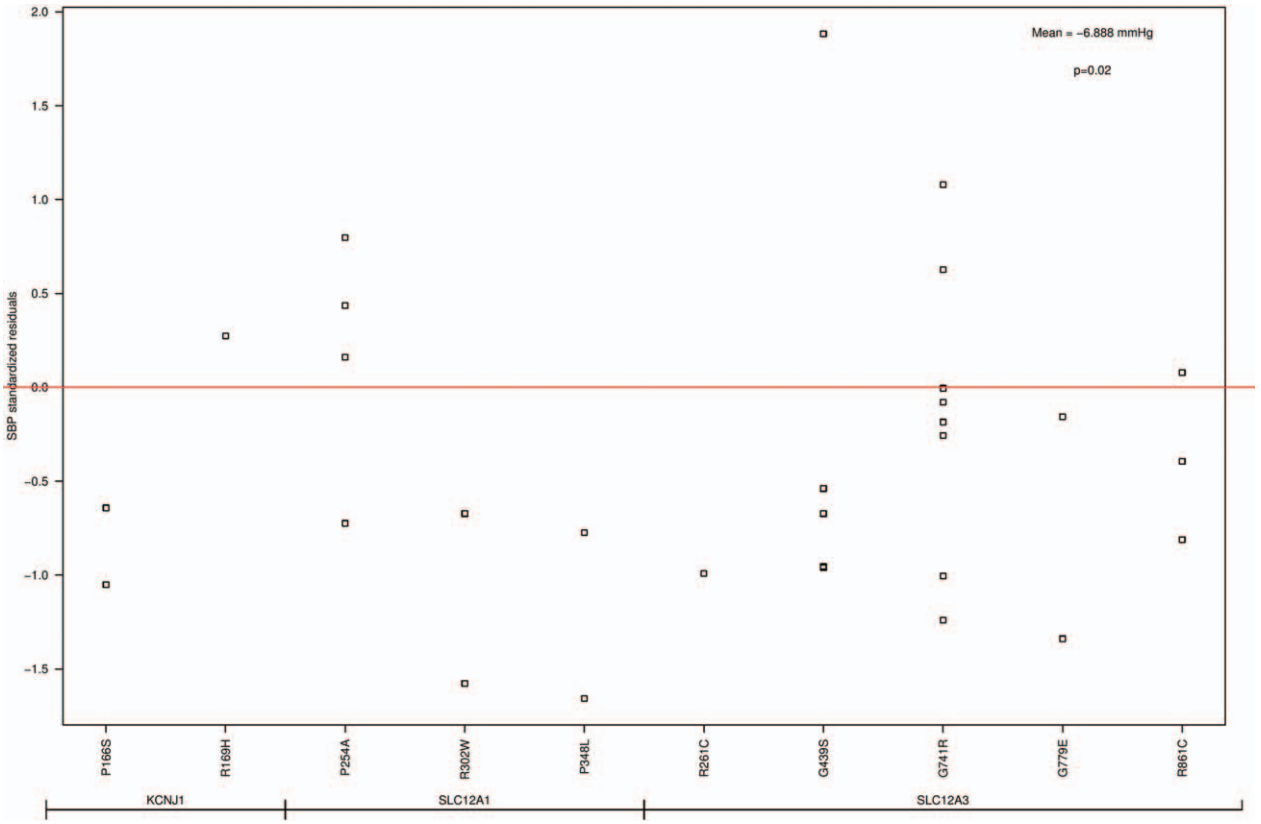

Figure 3. Age- and sex-adjusted SBP residuals for 10 FHS single nucleotide coding variants in ARIC depicting the mean effect of carriers.

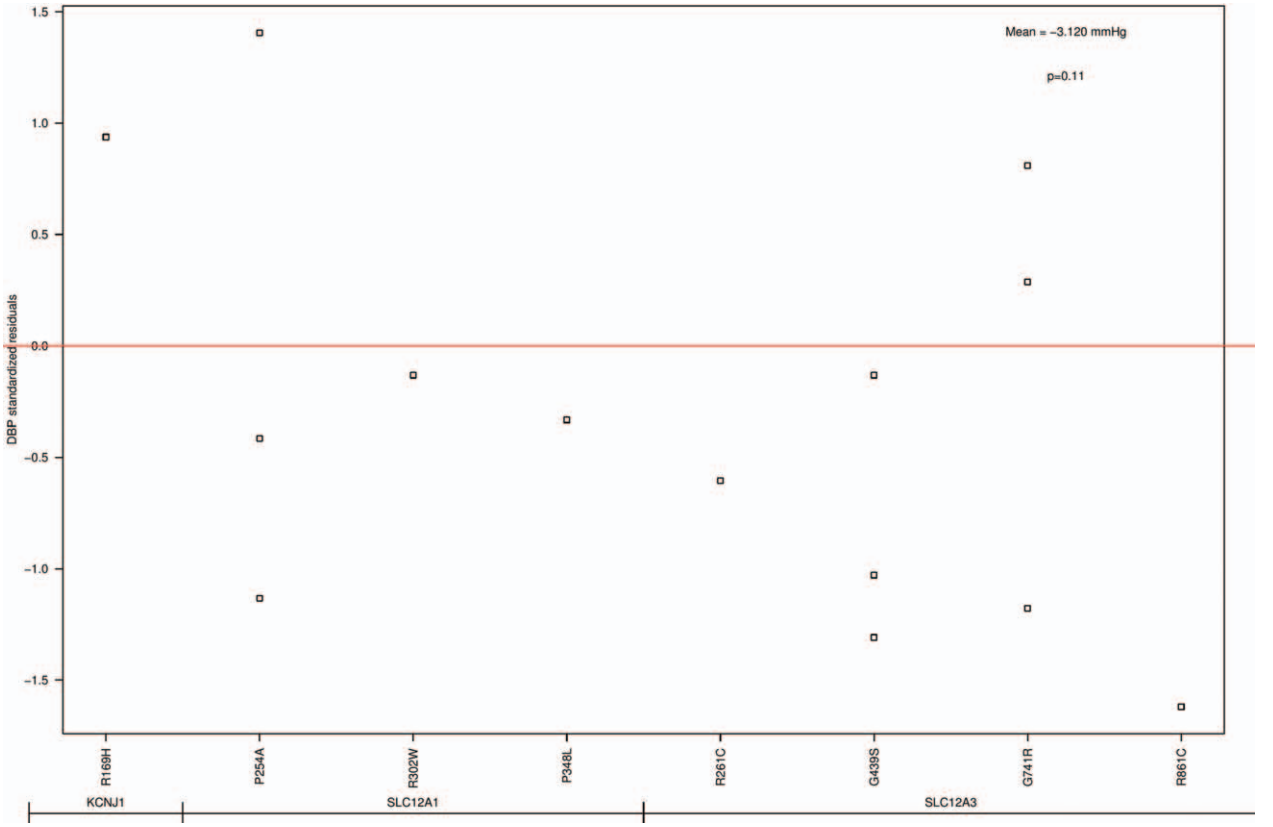

Figure 4. Age- and sex-adjusted DBP residuals for 8 FHS single nucleotide coding variants in ARIC depicting the mean effect of carriers.

4. Discussion

We attempted to apply criteria identical to Ji et al[29] to data from ARIC exome sequencing to identify putative LOF variants within the Bartter and Gitelman syndrome *SLC12A3*, *SLC12A1*, and *KCNJ1* genes. Although we failed to demonstrate an overall BP-lowering effect of all rare, coding LOF variants we did replicate the LOF effect in our data of several specific rare variants analyzed in the study by Ji et al.[29] Thus, deleterious variants, or at least some specific alleles, in these genes are protective against essential HTN in the general population. Nevertheless, the lack of overall replication merits further discussion, as it demonstrates that not all alleles annotated as deleterious indeed are so. This is not surprising, but the problem may be more widespread than is acknowledged.

It is usually expected that identical computational genomic analyses should return similar, if not identical, results; however, many inadvertent differences can occur. First, there are differences in the sequence versions of the orthologs and paralogs used in the 2 studies because many records were either updated or dropped from the NCBI Refseq records over the past 9 years. We had 18% to 42% sites conserved across all species in the multiple alignments, while Ji et al[29] state 18% to 24% conservation rate for each of the 3 genes. Sequence and ClustwalW software version differences, as well as differences in calculation methods, are likely to have led to some discrepancies. In our multiple alignments, 21 of 28 FHS SNVs are completely conserved, of which 17 are listed in Table 1 of the study by Ji et al[29] as completely conserved (the remaining 4 are listed as conserved in vertebrates in their study). In addition, dropping sequence records that were obsolete by the time we conducted this study likely contributed to the higher percentages of conserved sites in the species studied, though these positions still represent a conserved set.

The 2 studies also used different variant detection methods: we used exome sequencing, whereas Ji et al[29] used temperature gradient capillary electrophoresis (TGCE) with confirmation by polymerase chain reaction (PCR) amplification and Sanger sequencing from the original DNA sample. The original study stated a "high" sensitivity of detection, having identified known single nucleotide variants (SNVs) in FHS, buttressed by a high concordance with frequencies in previous studies. Follow-up studies state that the vast majority (14 of 18, with conflicting results for an addition 2 variants) of the variants in their study were shown in *Xenopus laevis* oocytes and HEK293 cells to be LOF.[46–48] We use WES data in this study, at high sequencing depths that are in the range of previous estimates deemed sufficient for detection of the vast majority of heterozygous variants.[49–51] However, even greater accuracies may be warranted for genotype-phenotype correlation studies using rare variants. It should also be noted, however, that large-scale sequencing data often contain both false positives and false negatives,[52,53] despite a low error rate, for the rarest of variants. Our annotation of the 65 analyzed ARIC variants and the 28 FHS SNVs analyzed in the study by Ji et al[29] with the software programs SIFT, PANTHER, and PolyPhen support the predicted pathogenicity of both variant sets, and similar fractions of variants in both sets were present in the ExAC non-Finnish European population, demonstrating high concordance with their properties. However, it is not necessarily surprising for sequencing errors to also display this profile of pathogenicity due

to the increased probability that rare variants are sequencing errors and that the analyzed variants were specifically chosen at sites preselected for their conservation among a set of species.

Despite these differences, the carriers of the subset of variants from the study by Ji et al[29] that were also in the ARIC WES data present similar effect sizes for SBP (10 variants) and DBP (8 variants) as in the study by Ji et al,[29] and nearly all passed the same selection criteria in ARIC. As the distributions of variants in both studies seem to show no distinct or study-specific patterns or differences, it remains possible that only certain LOF variants at mechanistically specific locations are essential for BP regulation. This replication is especially notable because the FHS offspring cohort is younger in their visits as compared to ARIC individuals, who were minimum 45 years of age at baseline, and the age difference affects the proportion of individuals whose measurements necessitated adjustment for medication use. Further, the methods detailed in Levy et al[38] indicate that the effects reflect a longer time span of measurements, as they required samples to have at least 3 to 4 visits over at least 10 years, while our samples had a maximum of 4 visits over 12 years. These differences can affect analyses, but in this case the effect is still visible. Additionally, this represents a replication of the effects of rare variants with very low minor allele counts, which demonstrates the utility of exome sequencing data for study of such rare variants, though greater study is required to determine more appropriate filtering methods.

In summary, our attempt to replicate the methods in the study by Ji et al[29] to detect rare and potentially LOF variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* reducing SBP and DBP in variant carriers as compared with noncarriers using WES data provided important lessons. Although the 2 studies are comparable in numerous ways, there are also pertinent differences that can lead to their discrepant outcomes. Regardless of this, we successfully replicated the reduction effects in SBP and DBP with a subset of variants from the study by Ji et al[29] that were present in the ARIC study, which upholds the use of such sequencing data for the study of very rare variants and confirms that *SLC12A3*, *SLC12A1*, and *KCNJ1* are indeed genes protective of HTN in the general population. Though HTN is a common disease, as the particular renal salt wasting syndromes of interest in this study (Bartter and Gitelman) are rare, the method is applicable for identifying rare LOF variants in other rare diseases. Nevertheless, our study highlights the need for improved methods for predicting variants effects and functional tests to identify them equivocally.

Author contributions

Conceptualization: Aravinda Chakravarti.
Data curation: Alanna C. Morrison, Megan L. Grove, Eric Boerwinkle.
Formal analysis: Priyanka Nandakumar, Aravinda Chakravarti.
Funding acquisition: Eric Boerwinkle, Aravinda Chakravarti.
Investigation: Priyanka Nandakumar, Megan L. Grove, Aravinda Chakravarti.
Methodology: Priyanka Nandakumar, Aravinda Chakravarti.
Project administration: Aravinda Chakravarti.
Resources: Alanna C. Morrison, Megan L. Grove, Eric Boerwinkle, Aravinda Chakravarti.
Software: Priyanka Nandakumar.
Supervision: Aravinda Chakravarti.
Visualization: Priyanka Nandakumar.
Writing – original draft: Priyanka Nandakumar.
Writing – review & editing: Priyanka Nandakumar, Alanna C. Morrison, Megan L. Grove, Eric Boerwinkle, Aravinda Chakravarti.

References

[1] Centers for Disease Control, Prevention (CDC)Vital signs: prevalence, treatment, and control of hypertension–United States, 1999–2002 and 2005–2008. MMWR Morb Mortal Wkly Rep 2011;60:103–8.
[2] CDC. National Center for Health Statistics – Hypertension. Available at: http://www.cdc.gov/nchs/fastats/hypertension.htm. Updated April 27, 2016. Last accessed August 5, 2016.
[3] Ehret GB, Munroe PB, Rice KM, et al. International Consortium for Blood Pressure Genome-Wide Association StudiesGenetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;478:103–9.
[4] Johnson T, Gaunt TR, Newhouse SJ, et al. Blood pressure loci identified with a gene-centric array. Am J Hum Genet 2011;89:688–700.
[5] 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:e1001177.
[6] 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;43:531–8.
[7] 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;43:1005–11.
[8] 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;93:545–54.
[9] 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;95:49–65.
[10] 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;94:349–60.
[11] 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;95:24–38.
[12] Wang Y, O'Connell JR, McArdle PF, et al. From the cover: whole-genome association study identifies STK39 as a hypertension susceptibility gene. Proc Natl Acad Sci U S A 2009;106:226–31.
[13] Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nat Genet 2016;48:1171–84.
[14] Johnson AD, Newton-Cheh C, Chasman DI, et al. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. Hypertens Dallas Tex 19792011;57:903–10.
[15] Ganesh SK, Tragante V, Guo W, et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum Mol Genet 2013;22:1663–78.
[16] Takeuchi F, Isono M, Katsuya T, et al. Blood pressure and hypertension are associated with 7 loci in the Japanese population. Circulation 2010;121:2302–9.
[17] Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet 2009;41:677–87.
[18] Newton-Cheh C, Larson MG, Vasan RS, et al. Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. Nat Genet 2009;41:348–53.
[19] Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009;41:666–76.
[20] Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet 2015;47:1282–93.
[21] Parmar PG, Taal HR, Timpson NJ, et al. International genome-wide association study consortium identifies novel loci associated with blood pressure in children and adolescents. Circ Cardiovasc Genet 2016;9:266–78.
[22] Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. Nat Genet 2016;48:1162–70.
[23] Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat Genet 2016;48:1151–61.

[24] Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet 2017;49:54–64.
[25] Zhu X, Feng T, Tayo BO, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am J Hum Genet 2015;96:21–36.
[26] 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;20:2273–84.
[27] 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;49:403–15.
[28] Lifton RP. Molecular genetics of human blood pressure variation. Science 1996;272:676–80.
[29] Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 2008;40:592–9.
[30] Rudin A. Bartter's syndrome. A review of 28 patients followed for 10 years. Acta Med Scand 1988;224:165–71.
[31] Simon DB, Nelson-Williams C, Bia MJ, et al. Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter. Nat Genet 1996;12:24–30.
[32] Atherosclerosis Risk in Communities. [Internet]. Available at: https://www2.cscc.unc.edu/aric. Accessed October 23, 2015.
[33] The Atherosclerosis Risk in Communities (ARIC)Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687–702.
[34] Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2009;2:73–80.
[35] Yu B, Pulit SL, Hwang S-J, et al. Rare exome sequence variants in CLCN6 reduce blood pressure levels and hypertension risk. Circ Cardiovasc Genet 2016;9:64–70.
[36] Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867–73.
[37] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
[38] Levy D, DeStefano AL, Larson MG, et al. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study. Hypertension 2000;36:477–83.
[39] Franklin SS, Gustin W, Wong ND, et al. Hemodynamic patterns of age-related changes in blood pressure. The Framingham Heart Study. Circulation 1997;96:308–15.
[40] Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.
[41] Li W, Cowley A, Uludag M, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res 2015;43: W580–4.
[42] Thomas PD, Kejariwal A, Guo N, et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res 2006;34: W645–50.
[43] Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4.
[44] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;Chapter 7:Unit7.20.
[45] Omasits U, Ahrens CH, Müller S, et al. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 2014;30:884–6.
[46] Fang L, Li D, Welling PA. Hypertension resistance polymorphisms in ROMK (Kir1.1) alter channel function by different mechanisms. Am J Physiol Renal Physiol 2010;299:F1359–64.
[47] Monette MY, Rinehart J, Lifton RP, et al. Rare mutations in the human Na-K-Cl cotransporter (NKCC2) associated with lower blood pressure exhibit impaired processing and transport function. Am J Physiol Renal Physiol 2011;300:F840–7.
[48] Acuña R, Martínez-de-la-Maza L, Ponce-Coria J, et al. Rare mutations in SLC12A1 and SLC12A3 protect against hypertension by reducing the activity of renal salt cotransporters. J Hypertens 2011;29:475–83.
[49] Ajay SS, Parker SCJ, Abaan HO, et al. Accurate and comprehensive sequencing of personal genomes. Genome Res 2011;21:1498–505.
[50] Clark MJ, Chen R, Lam HYK, et al. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 2011;29:908–14.
[51] Meynert AM, Ansari M, FitzPatrick DR, et al. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics 2014;15:247.
[52] Fuentes Fajardo KV, Adams D, Mason CE, et al. NISC, Comparative Sequencing ProgramDetecting false-positive signals in exome sequencing. Hum Mutat 2012;33:609–13.
[53] Huang X-F, Wu J, Lv J-N, et al. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a complementary approach to clinical genetic diagnostic testing. Genet Med Off J Am Coll Med Genet 2015;17:307–11.