ORIGINAL ARTICLE

Longitudinal exome-wide association study to identify genetic susceptibility loci for hypertension in a Japanese population

Yoshiki Yasukochi[1,2], Jun Sakuma[2,3,4], Ichiro Takeuchi[2,4,5], Kimihiko Kato[1,6], Mitsutoshi Oguri[1,7], Tetsuo Fujimaki[8], Hideki Horibe[9] and Yoshiji Yamada[1,2]

Genome-wide association studies have identified various genetic variants associated with complex disorders. However, these studies have commonly been conducted in a cross-sectional manner. Therefore, we performed a longitudinal exome-wide association study (EWAS) in a Japanese cohort. We aimed to identify genetic variants that confer susceptibility to hypertension using ~ 244 000 single-nucleotide variants (SNVs) and physiological data from 6026 Japanese individuals who underwent annual health check-ups for several years. After quality control, the association of hypertension with SNVs was tested using a generalized estimating equation model. Finally, our longitudinal EWAS detected seven hypertension-related SNVs that passed strict criteria. Among these variants, six SNVs were densely located at 12q24.1, and an East Asian-specific motif (haplotype) 'CAAAA' comprising five derived alleles was identified. Statistical analyses showed that the prevalence of hypertension in individuals with the East Asian-specific haplotype was significantly lower than that in individuals with the common haplotype 'TGGGT'. Furthermore, individuals with the East Asian haplotype may be less susceptible to the adverse effects of smoking on hypertension. The longitudinal EWAS for the recessive model showed that a novel SNV, rs11917356 of *COL6A5*, was significantly associated with systolic blood pressure, and the derived allele at the SNV may have spread throughout East Asia in recent evolutionary time.

INTRODUCTION

Hypertension is a major complex disorder affected by both genetic background and environmental factors, including lifestyle. In 2014, the Ministry of Health, Labour and Welfare reported that the estimated number of Japanese patients receiving medical treatment for hypertension had increased by ~ 10% since 2011 and was more than 10 million[1] (http://www.mhlw.go.jp/english/database/db-hss/). Therefore, the exploration of genetic variants that confer susceptibility to hypertension is important for the personalized prevention of hypertension.

In recent decades, genome-wide association studies (GWASs) in various ethnic populations have identified genes, or loci, associated with hypertension.[2–7] Two GWASs for systolic (SBP) and diastolic blood pressures (DBP) in

~ 30 000 individuals of European ancestry identified 13 hypertension-associated loci.[2,3] Subsequently, a large-scale GWAS using a multistage design to examine DBP and SBP in 200 000 individuals of European ancestry detected an additional 16 novel single-nucleotide variants (SNVs).[4] However, the latter GWAS detected fewer than 10 SNVs associated with DBP or SBP in non-European populations. This discrepancy could reflect the small sample size or differences in genetic backgrounds among populations, or this finding could be a true and valid result.

For the most part, conventional GWASs have been conducted in a cross-sectional manner that commonly measures traits at a single point in time. SBP, DBP and the prevalence of hypertension are significantly correlated with age. Therefore, control individuals who have normal blood pressure at a

[1]Department of Human Functional Genomics, Advanced Science Research Promotion Center, Mie University, Tsu, Japan; [2]CREST, Japan Science and Technology Agency, Kawaguchi, Japan; [3]Computer Science Department, College of Information Science, University of Tsukuba, Tsukuba, Japan; [4]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan; [5]Department of Computer Science, Nagoya Institute of Technology, Nagoya, Japan; [6]Department of Internal Medicine, Meitoh Hospital, Nagoya, Japan; [7]Department of Cardiology, Kasugai Municipal Hospital, Kasugai, Japan; [8]Department of Cardiovascular Medicine, Inabe General Hospital, Inabe, Japan and [9]Department of Cardiovascular Medicine, Gifu Prefectural Tajimi Hospital, Tajimi, Japan
Correspondence: Dr Y Yasukochi, Department of Human Functional Genomics, Advanced Science Research Promotion Center, Organization for the Promotion of Regional Innovation, Mie University, 1577 Kurima-machiya, Tsu, Mie 514-8507, Japan.
E-mail: hyasukou@proof.ocn.ne.jp
Received 21 April 2017; revised 1 June 2017; accepted 12 June 2017

certain point will potentially display hypertension after a few years. Conventional cross-sectional GWASs do not consider this possibility. Given that longitudinal GWASs evaluate temporal changes in blood pressure and the prevalence of hypertension related to age, this analysis increases the statistical power to detect these associations. Therefore, we traced disease progression and physiological changes in 6026 Japanese individuals who underwent annual health check-ups for several years. We performed a longitudinal exome-wide association study (EWAS), which is suited to studying the effects of low frequency or rare variants within a gene of interest, to explore novel genetic variants that confer susceptibility to hypertension, focusing on rare coding variants, such as East Asian-specific alleles. The results of this longitudinal EWAS enhance the current understanding of the relationship between ethnic-specific genetic factors and diseases.

MATERIALS AND METHODS

Ethics statement
The study protocol complied with the Declaration of Helsinki and was approved through the Committees on the Ethics of Human Research of Mie University Graduate School of Medicine and Inabe General Hospital. Written informed consent was obtained from all subjects before enrolment in the present study.

Study population
The 6026 community-dwelling individuals in Inabe City, Mie Prefecture, Japan, were recruited from individuals who visited the health-care center of Inabe General Hospital for an annual health check-up and were followed-up yearly (Inabe cohort). We refer to this cohort as the 'Inabe cohort'. Individuals in the Inabe cohort were registered between March 2010 and September 2012, and genomic DNA was extracted from venous blood cells using a DNA extraction kit (SMITEST EX-R&D; Medical and Biological Laboratories, Nagoya, Japan) and stored in the genomic DNA bank of the Research Center for Genomic Medicine at Mie University. For all participants, medical examination data obtained from April 2003 to March 2014 (11 years) were deposited into a database. Each subject had one set of health data for each year of attendance at the clinic. Therefore, all participants had undergone 1–11 medical examinations (a total of 28 529 examinations), and the average follow-up period was 5 years. The detailed characteristics of the study subjects in a 5-year follow-up are described elsewhere.[8]

Among the 6026 subjects in the Inabe cohort, 2249 patients were affected by essential hypertension, and the remaining 3777 patients were treated as controls. Hypertension was defined as either DBP of ⩾ 90 mm Hg or SBP of ⩾ 140 mm Hg (or both), by virtue of the subject having taken antihypertensive medication. Blood pressure was measured at least twice with the subjects having rested in the sitting position for > 5 min. A skilled physician or nurse obtained the measurements according to the guidelines of the American Heart Association.[9] The mean DBP of all subjects was 74.7 mm Hg, whereas the mean SBP was 120.7 mm Hg. The mean age of subjects was 52.5 years of age (18–91 years of age) and the mean body mass index was 23.0 kg m^{-2}. The percentages of missing data were 2.2% for DBP, SBP and body mass index. There were no missing data for age, gender or smoking status.
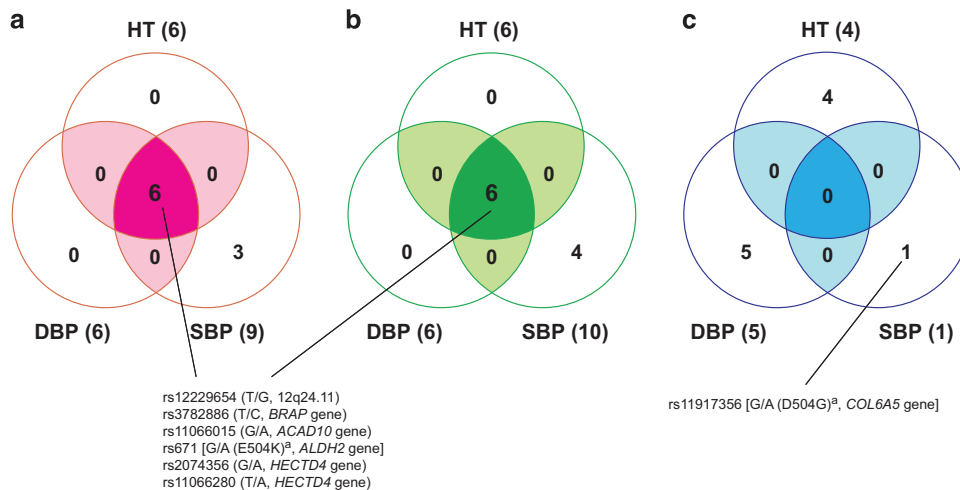
Longitudinal EWAS
A longitudinal EWAS for hypertension was performed on the Inabe cohort, and Infinium HumanExome-12 ver. 1.2 BeadChip or Infinium Exome-24 ver. 1.0 BeadChip (Illumina, San Diego, CA, USA) were used to genotype ~ 244 000 SNVs. The exome array includes putative functional exonic variants selected from > 12 000 ethnically diverse individuals.[10] Previously available marker sets of GWASs were designed to identify common alleles (minor allele frequency (MAF) of ⩾ 5%) and were not well suited to studying the effects of low frequency or rare variants within a gene of interest. The exome arrays used in the present study contained ~ 244 000 SNVs, including common, low-frequency (0.5% ⩽ MAF < 5%) and rare (MAF < 0.5%) variants, although EWAS is a focused genotyping method that differs from GWAS, which includes up to 4.5 million markers for SNVs and copy number variations. Using JMP Genomics version 6.0 (SAS Institute, Cary, NC, USA), the genotyping data of 6026 individuals were converted into binary data with the dominant and recessive models, after removing monomorphic sites among subjects, resulting in 58 563 SNVs for the dominant model and 38 725 SNVs for the recessive model. In a similar manner, the genotyping data were also converted into numerical data with the additive model, generating 58 563 SNVs. The dominant and recessive models were defined as 'AA (0) vs AB+BB (1)' and 'AA+AB (0) vs BB (1)' (A, major allele; B, minor allele), respectively, whereas the additive model was defined as 'AA (0) < AB (1) < BB (2)'. SNVs on the mitochondrial genome and sex chromosomes were removed from these analyses; SNVs in which the genotype distribution significantly deviated from Hardy–Weinberg equilibrium ($P < 0.001$) in controls were also discarded.

Sitlani *et al.*[11] reported that a small effective sample size can increase the chances of generating false positives (type I errors). In the preliminary analysis, a MAF of < 0.01 appeared to increase type I errors in the recessive model because extraordinarily large numbers of SNVs were identified as hypertension-associated SNVs. Therefore, we discarded SNVs with a MAF of < 0.05. Quantile–quantile plots for *P*-values of allele frequencies in the EWASs for DBP, SBP, or the prevalence of hypertension in all genetic models are shown in Supplementary Figure S1. The genomic inflation factor (λ) of *P*-values was 1.10 for the prevalence of hypertension and 1.09 for DBP and SBP in the dominant model (Supplementary Figure S1a). In the additive model, the λ was 1.12 for the prevalence of hypertension and 1.09 for DBP and SBP (Supplementary Figure S1b). In the recessive model, the λ was 1.07 for the prevalence of hypertension, 1.09 for DBP and 1.14 for SBP (Supplementary Figure S1c).

The genotype data of participants in the EWAS were examined for population stratification using principal component analysis according to the EIGENSTRAT method[12] with the JMP Genomics program (Supplementary Figure S2), and we excluded four outliers identified in the analysis. The rearrangement of Inabe longitudinal data was conducted using R software version 3.32[13] via RStudio version 1.0.136[14] and Perl script.

Statistical analyses
The association of longitudinal changes in the prevalence of hypertension, DBP and SBP with genetic variants was examined using the generalized estimating equation (GEE) model[15,16] with adjustments for age, gender, body mass index and smoking status, using the R package 'geepack'.[17] The waves argument was used to specify the ordering of repeated measurements within individuals. The statistical significance of the association was $P < 8.54 \times 10^{-7}$ (0.05/58 563 SNVs) for dominant and additive models and $P < 1.29 \times 10^{-6}$ (0.05/38 725

a HT (6) DBP (6) SBP (9)
0 / 0 / 6 / 0 / 0 / 0 / 3

b HT (6) DBP (6) SBP (10)
0 / 0 / 6 / 0 / 0 / 0 / 4

c HT (4) DBP (5) SBP (1)
4 / 0 / 0 / 0 / 5 / 0 / 1

rs12229654 (T/G, 12q24.11)
rs3782886 (T/C, BRAP gene)
rs11066015 (G/A, ACAD10 gene)
rs671 [G/A (E504K)[a], ALDH2 gene]
rs2074356 (G/A, HECTD4 gene)
rs11066280 (T/A, HECTD4 gene)

rs11917356 [G/A (D504G)[a], COL6A5 gene]

Figure 1 Single-nucleotide variants (SNVs) showing significant associations with hypertension. The associations were examined using the generalized estimating equation (GEE) model in the dominant (**a**), additive (**b**) and recessive (**c**) genetic models. The seven candidates of hypertension-associated SNVs shown in the figure were determined when the association was supported by the GEE model with normal and t reference distributions, and the approxdf value of these SNVs was higher than 30 (see text). The three GEE analyses independently tested the association between SNVs and three clinical parameters: prevalence of hypertension (HT); systolic blood pressure (SBP); and diastolic blood pressure (DBP). On the basis of Bonferroni's correction, P-values of $< 8.54 \times 10^{-7}$ for the dominant and additive models and $P < 1.29 \times 10^{-6}$ for the recessive model were considered statistically significant. [a]Amino-acid change reflecting a point mutation at each SNV (nonsynonymous substitution).

SNVs) for the recessive model after applying Bonferroni's correction to compensate for the multiple comparison of genotypes with the clinical parameters. Sitlani and colleagues proposed that compared with the use of normal distribution, the use of a t reference distribution with estimated degrees of freedom for the distributional assumption of the ratio of coefficient estimate to its standard error can improve the accuracy of the significance detected with the GEE. In addition, these authors recommended the use of approxdf, a scale of small effective sample size:[11] approxdf $= 2 \times MAF \times Nindep$, where Nindep is the sum of the estimated number of independent observations per person. We also computed the P-value via GEE using a t reference distribution with Satterthwaite estimates of degrees of freedom and approxdf using the R package 'bosswithdf',[11,18] and discarded SNVs with approxdf < 10, as approxdf $\geqslant 10$ decreases the chances of generating type I errors.[11] The GEE with t reference distribution was implemented using the clinical parameters of subjects in the last 5 years for which data were available.

The significance of difference of DBP or SBP between groups of subjects with different genotypes was assessed using Welch's t-test (two-sided P-value < 0.05). Haplotype phase and linkage disequilibrium (LD) among SNVs were estimated using IMPUTE2 version 2.3.2[19] and Haploview version 4.2[20] programs, respectively. Perl and R scripts were written to convert the SNV data used in the present study into suitable formats for each program.

Survey of allele frequencies in human populations and variants in nonhuman primates
Information on allele frequencies of target SNVs within human populations was obtained from the 1000 Genomes Project[21] (http://www.internationalgenome.org/) and the Integrative Japanese Genome Variation Database (iJGVD) from the Tohoku Medical Megabank Organization of Tohoku University[22] (https://ijgvd.megabank.tohoku.ac.jp/author/tommo/). In addition, a genetic variant at the homologous SNVs within vertebrates was investigated in the UCSC Genome

Browser (for archaic humans) or Multiz Alignments of 100 Vertebrates in the UCSC database (http://genome.ucsc.edu), and information on allele frequencies in great apes was obtained from the Great Ape Genome Project database[23] (http://biologiaevolutiva.org/greatape/).

RESULTS
Association of SNVs with hypertension in the Inabe cohort
In the present study, the GEE model with adjustments for age, gender, body mass index and smoking status was used to examine the association between 58 563 SNVs for the dominant and additive models, and 38 725 SNVs for the recessive model and clinical parameters related to hypertension in the Inabe cohort (after removing four outliers identified using the principal component analysis; Supplementary Figure S2). In the dominant and additive models, six SNVs located at 12q24.11–12q24.13 were significantly ($P < 8.54 \times 10^{-7}$) associated with all three clinical parameters: the prevalence of hypertension; SBP; and DBP (Figure 1; Table 1). In addition, three SNVs located on chromosome 10 were significantly associated with SBP in the dominant and additive models, and one additional SNV located on the same chromosomal region showed a significant association with SBP in the additive model (Figure 1; Table 1). In the recessive model, four SNVs were significantly ($P < 1.29 \times 10^{-6}$) associated with the prevalence of hypertension, and one and five SNVs were significantly associated with SBP and DBP, respectively (Figure 1; Table 1).

To confirm the associations between candidate SNVs detected in the longitudinal EWAS and clinical parameters related to hypertension, we applied the GEE model with t reference distribution and threshold of approxdf ($\geqslant 10$). Additional analysis showed that all six SNVs at 12q24.11–12q24.13

Table 1 Candidate SNVs showing significant associations with clinical parameters associated with hypertension using the generalized estimating equation model with adjustments for age, gender, BMI and smoking status

Model[a]	Trait/disease	Ref SNP ID	Allele	Position[b]	Gene	Genotype[a]	Estimate[c]	P-value (n ref. dist)[d]	MAF	Approxdf[e]	P-value (t ref. dist)[f]
Dominant	HT	rs12229654	T→G	12: 111 414 461		TT vs TG+GG	−0.35	**2.4×10^{-8}**	0.253	**256.4**	**1.3×10^{-5}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT vs TC+CC	−0.39	**9.8×10^{-11}**	0.318	**338.6**	**6.6×10^{-9}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG vs GA+AA	−0.41	**1.9×10^{-11}**	0.300	**336.5**	**2.7×10^{-9}**
		rs671	G→A	12: 112 241 766	ALDH2	GG vs GA+AA	−0.41	**3.0×10^{-11}**	0.301	**337.4**	**3.9×10^{-9}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG vs GA+AA	−0.37	**1.9×10^{-9}**	0.278	**304.3**	**1.7×10^{-7}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT vs TA+AA	−0.37	**1.7×10^{-9}**	0.313	**338.1**	**4.1×10^{-8}**
	SBP	rs12413409	G→A	10: 104 719 096	CNNM2	GG vs GA+AA	−1.62	**8.1×10^{-7}**	0.273	**253.9**	0.014
		rs11191548	T→C	10: 104 846 178		TT vs TC+CC	−1.62	**7.7×10^{-7}**	0.273	**252.4**	0.014
		rs11191580	T→C	10: 104 906 211	NT5C2	TT vs TC+CC	−1.63	**7.4×10^{-7}**	0.273	**250.7**	0.016
		rs12229654	T→G	12: 111 414 461		TT vs TG+GG	−1.89	**1.1×10^{-8}**	0.253	**224.6**	**3.8×10^{-5}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT vs TC+CC	−1.84	**2.1×10^{-8}**	0.318	**297.3**	**3.3×10^{-6}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG vs GA+AA	−2.03	**5.6×10^{-10}**	0.300	**295.4**	**1.2×10^{-7}**
		rs671	G→A	12: 112 241 766	ALDH2	GG vs GA+AA	−1.99	**1.1×10^{-9}**	0.301	**296.3**	**2.3×10^{-7}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG vs GA+AA	−1.91	**5.5×10^{-9}**	0.278	**266.7**	**2.2×10^{-6}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT vs TA+AA	−1.73	**1.3×10^{-7}**	0.313	**296.8**	**1.5×10^{-5}**
	DBP	rs12229654	T→G	12: 111 414 461		TT vs TG+GG	−1.60	**2.8×10^{-10}**	0.253	**256.4**	**1.4×10^{-6}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT vs TC+CC	−1.60	**3.6×10^{-10}**	0.318	**338.6**	**1.9×10^{-7}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG vs GA+AA	−1.73	**8.4×10^{-12}**	0.300	**336.5**	**9.8×10^{-9}**
		rs671	G→A	12: 112 241 766	ALDH2	GG vs GA+AA	−1.72	**1.3×10^{-11}**	0.301	**337.4**	**1.5×10^{-8}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG vs GA+AA	−1.59	**3.5×10^{-10}**	0.278	**304.3**	**5.8×10^{-7}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT vs TA+AA	−1.54	**1.3×10^{-9}**	0.313	**338.1**	**5.0×10^{-7}**
Additive	HT	rs12229654	T→G	12: 111 414 461		TT<TG<GG	−0.30	**7.7×10^{-9}**	0.270	**35.3**	**2.4×10^{-5}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT<TC<CC	−0.31	**6.6×10^{-11}**	0.337	**73.3**	**3.6×10^{-9}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG<GA<AA	−0.33	**2.4×10^{-11}**	0.320	**61.1**	**2.1×10^{-9}**
		rs671	G→A	12: 112 241 766	ALDH2	GG<GA<AA	−0.32	**4.0×10^{-11}**	0.320	**61.2**	**2.6×10^{-9}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG<GA<AA	−0.31	**1.5×10^{-11}**	0.295	**50.7**	**3.4×10^{-8}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT<TA<AA	−0.30	**4.7×10^{-10}**	0.332	**71.8**	**1.2×10^{-8}**
	SBP	rs11191454	A→G	10: 104 660 004	AS3MT	AA<AG<GG	−1.29	**7.2×10^{-7}**	0.290	**45.1**	0.029
		rs12413409	G→A	10: 104 719 096	CNNM2	GG<GA<AA	−1.35	**2.0×10^{-7}**	0.285	**43.7**	0.017
		rs11191548	T→C	10: 104 846 178		TT<TC<CC	−1.34	**2.3×10^{-7}**	0.285	**43.7**	0.019
		rs11191580	T→C	10: 104 906 211	NT5C2	TT<TC<CC	−1.34	**2.4×10^{-7}**	0.284	**44.2**	0.020
		rs12229654	T→G	12: 111 414 461		TT<TG<GG	−1.62	**8.4×10^{-10}**	0.270	**30.3**	**2.0×10^{-5}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT<TC<CC	−1.55	**5.5×10^{-10}**	0.337	**64.0**	**2.2×10^{-8}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG<GA<AA	−1.69	**3.5×10^{-11}**	0.320	**53.2**	**1.8×10^{-9}**
		rs671	G→A	12: 112 241 766	ALDH2	GG<GA<AA	−1.67	**4.9×10^{-11}**	0.320	**53.2**	**2.5×10^{-9}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG<GA<AA	−1.68	**1.3×10^{-10}**	0.295	**44.3**	**2.0×10^{-8}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT<TA<AA	−1.50	**2.1×10^{-9}**	0.332	**62.8**	**4.9×10^{-8}**
	DBP	rs12229654	T→G	12: 111 414 461		TT<TG<GG	−1.33	**3.9×10^{-11}**	0.270	**30.3**	**4.2×10^{-6}**
		rs3782886	T→C	12: 112 110 489	BRAP	TT<TC<CC	−1.36	**1.6×10^{-12}**	0.337	**64.0**	**1.8×10^{-9}**
		rs11066015	G→A	12: 112 168 009	ACAD10	GG<GA<AA	−1.45	**1.4×10^{-13}**	0.320	**53.2**	**4.3×10^{-10}**
		rs671	G→A	12: 112 241 766	ALDH2	GG<GA<AA	−1.45	**1.2×10^{-13}**	0.320	**53.3**	**3.6×10^{-10}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG<GA<AA	−1.39	**2.8×10^{-12}**	0.295	**44.3**	**9.6×10^{-9}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT<TA<AA	−1.32	**6.9×10^{-12}**	0.332	**62.9**	**3.1×10^{-9}**
Recessive	HT	rs3766871	G→A	1: 237 778 084	RYR2	GG+GA vs AA	−37.54	**$<2.0 \times 10^{-16}$**	0.051	0.3	**4.4×10^{-4}**
		rs11258194	T→A	10: 13 152 400	OPTN	TT+TA vs AA	−39.49	**$<2.0 \times 10^{-16}$**	0.075	0.2	**2.7×10^{-4}**
		rs10519996	G→C	15: 35 665 765	DPH6	GG+GC vs CC	−38.52	**$<2.0 \times 10^{-16}$**	0.065	0.1	**7.4×10^{-4}**
		rs5763919	G→A	22: 30 662 763	OSM	GG+GA vs AA	−38.14	**$<2.0 \times 10^{-16}$**	0.055	0.2	**4.0×10^{-7}**
	SBP	rs11917356	G→A	3: 130 110 550	COL6A5	GG+GA vs AA	−2.93	**2.3×10^{-8}**	0.320	**60.3**	**6.4×10^{-5}**
	DBP	rs79450009	C→T	4: 25 673 284	SLC34A2	CC+CT vs TT	−12.79	**2.1×10^{-9}**	0.050	0.4	NA
		rs3782886	T→C	12: 112 110 489	BRAP	TT+TC vs CC	−2.11	**3.7×10^{-7}**	0.318	**64.0**	**5.6×10^{-6}**
		rs671	G→A	12: 112 241 766	ALDH2	GG+GA vs AA	−2.16	**6.7×10^{-7}**	0.301	**53.3**	**1.4×10^{-5}**
		rs2074356	G→A	12: 112 645 401	HECTD4	GG+GA vs AA	−2.26	**1.1×10^{-6}**	0.278	**44.3**	**1.9×10^{-5}**
		rs11066280	T→A	12: 112 817 783	HECTD4	TT+TA vs AA	−2.07	**6.9×10^{-7}**	0.313	**62.9**	**3.7×10^{-6}**

Abbreviations: BMI, body mass index; DBP, diastolic blood pressure; HT, prevalence of hypertension; MAF, minor allele frequency; NA, not available; ref. dist, reference distribution; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism; SNV, single-nucleotide variant.
[a]Genetic model, dominant model: AA vs AB+BB (A, major allele; B, minor allele); additive model (AA<AB<BB); recessive model (AA+AB vs BB).
[b]Position in NCBI build GRCh37.
[c]Estimate of coefficient.
[d]P-values via generalized estimating equation (GEE) model using a normal reference distribution (a statistically significant P-value corrected using Bonferroni's method is shown in bold).
[e]A scale of small effective sample size: approxdf $= 2 \times$ MAF $\times$ Nindep, where Nindep is the sum of the estimated number of independent observations per person (approxdf of $\geqslant 10$ is shown in bold).
[f]P-values via the GEE model using a t reference distribution with Satterthwaite estimates of degrees of freedom (a statistically significant P-value corrected using Bonferroni's method is shown in bold).

Table 2 Seven candidate SNVs showing significant associations with the clinical parameters associated with hypertension

Trait/disease	Ref SNP ID	Position[a]	Gene or locus	Mutation	Genotype	All subjects[b]	Hypertension[b]	Controls[b]
HT–SBP–DBP	rs12229654	Chr12: 111 414 461	12q24.11	Silent	TT	15 908 (55.8%)	5829 (60.5%)	10 079 (53.3%)
	(T → G)				TG	10 796 (37.8%)	3323 (34.5%)	7473 (39.6%)
					GG	1823 (6.4%)	482 (5.0%)	1341 (7.1%)
HT–SBP–DBP	rs3782886	Chr12: 112 110 489	BRAP	Synonymous (Arg241Arg)	TT	13 297 (46.6%)	5035 (52.3%)	8262 (43.7%)
	(T → C)				TC	12 380 (43.4%)	3826 (39.7%)	8554 (45.3%)
					CC	2852 (10.0%)	773 (8.0%)	2079 (11.0%)
HT–SBP–DBP	rs11066015	Chr12: 112 168 009	ACAD10	Silent in intron	GG	13 950 (46.6%)	5258 (54.6%)	8692 (46.0%)
	(G → A)				GA	12 026 (43.4%)	3691 (38.3%)	8335 (44.1%)
					AA	2553 (10.0%)	685 (7.1%)	1868 (9.9%)
HT–SBP–DBP	rs671	Chr12: 112 241 766	ALDH2	Missense (Glu303Lys)	GG	13 939 (48.9%)	5250 (54.5%)	8689 (46.0%)
	(G → A)				GA	12 026 (42.2%)	3697 (38.4%)	8329 (44.1%)
					AA	2564 (8.9%)	687 (7.1%)	1877 (9.9%)
HT–SBP–DBP	rs2074356	Chr12: 112 645 401	HECTD4	Silent in intron	GG	14 876 (48.9%)	5498 (57.1%)	9378 (49.6%)
	(G → A)				GA	11 388 (42.2%)	3546 (36.8%)	7842 (41.5%)
					AA	2265 (9.0%)	590 (6.1%)	1675 (8.9%)
HT–SBP–DBP	rs11066280	Chr12: 112 817 783	HECTD4	Silent in intron	TT	13 409 (52.1%)	5066 (52.6%)	8343 (44.2%)
	(T → A)				TA	12 338 (39.9%)	3816 (39.6%)	8522 (45.1%)
					AA	2782 (7.9%)	752 (7.8%)	2030 (10.7%)
SBP	rs11917356	Chr3: 130 110 550	COL6A5	Missense (Asp982Gly)	GG	2917 (48.4%)	1787 (47.4%)	1130 (50.2%)
	(G → A)				GA	2494 (41.4%)	1556 (41.2%)	938 (41.7%)
					AA	611 (10.1%)	430 (11.4%)	181 (8.0%)

Abbreviations: DBP, diastolic blood pressure; HT, prevalence of hypertension; SBP, systolic blood pressure; SNP, single-nucleotide polymorphism; SNV, single-nucleotide variant.
[a]Position in NCBI build GRCh37.
[b]Values indicate the numbers of measurements obtained, and the percentages are indicated in parentheses.

(rs12229654 at 12q24.11, rs3782886 of BRAP, rs11066015 of ACAD10, rs671 of ALDH2, and rs2074356 and rs11066280 of HECTD4) detected in the dominant and additive models and one SNV (rs11917356 of COL6A5) in the recessive model showed significant associations ($P < 0.05$/the number of candidate SNVs in each genetic model) and values of approxdf were > 30 (Table 1). Consequently, we identified the seven SNVs as genetic variants that confer susceptibility to hypertension (Table 2).

Identification of East Asian-specific haplotype on chromosome 12
Our longitudinal EWAS for hypertension in the dominant and additive models revealed six significant SNVs. On the basis of the information obtained from the 1000 Genomes Project[21] and iJGVD databases, we examined allele frequencies among human populations (Table 3). Interestingly, minor (derived) alleles of all six SNVs were specifically detected in East Asia. This result suggests that the derived alleles were recently expanded throughout East Asia. On the basis of clinical parameters of subjects in the latest year for which data were available, the prevalence of hypertension in subjects with the East Asian-specific allele was lower than that of those with

common (ancestral) alleles at each SNV site (mean odds ratio = 0.78, $P < 1.0 \times 10^{-8}$ using Fisher's exact test). Among local Japanese populations, the frequencies of the East Asian alleles were highest in the Inabe cohort (Table 3).

The six hypertension-associated SNVs were located at 12q24.1. Therefore, using the SNV data obtained in the present study, we estimated the LD among these SNVs based on 232 biallelic sites spanning an ~ 2.0 Mb genomic region on chromosome 12 (Supplementary Figure S3). The LD among the six SNVs was relatively strong for the physical distance. Particularly, five SNVs at 12q24.12–12q24.13 (excluding rs12229654 at 12q24.11) showed strong LDs ($r^2 > 0.8$). Therefore, we used these five SNVs for a haplotype estimate.

Using the IMPUTE2 program, we generated phased haplotypes comprising 232 biallelic sites, and examined the genotype containing the five hypertension-related SNV motifs (from left, rs3782886 of BRAP, rs11066015 of ACAD10, rs671 of ALDH2, and rs2074356 and rs11066280 of HECTD4) for all subjects. In the Inabe cohort, the proportion of diplotype 'TGGGT/TGGGT' was the highest (2274 subjects, 46.1%), followed by 'TGGGT/CAAAA' (2109 subjects, 35.0%) and 'TGGGT/CAAAA' (460 subjects, 7.6%; Supplementary Table S1). The 'TGGGT' haplotype comprises all ancestral alleles at each

Table 3 Allele frequencies of hypertension-associated SNPs in human populations

Ref SNP ID	East Asian								South Asian[a]	European[a]	African[a]
	All	JP-Inabe	JPT[a]	JP-Tohoku[b]	CDX[a]	CHB[a]	CHS[a]	KHV[a]			
rs12229654	T: 0.777 (13 319)	T: 0.747 (9001)	T: 0.837 (174)	T: 0.847 (3470)	T: 0.957 (178)	T: 0.835 (172)	T: 0.724 (152)	T: 0.869 (172)	T: 1.000 (978)	T: 1.000 (1006)	T: 1.000 (1322)
Near MYL2	G: 0.223 (3829)	G: 0.253 (3043)	G: 0.163 (34)	G: 0.153 (626)	G: 0.043 (8)	G: 0.165 (34)	G: 0.276 (58)	G: 0.131 (26)			
rs3782886	T: 0.719 (12 323)	T: 0.682 (8220)	T: 0.750 (156)	T: 0.799 (3271)	T: 0.957 (178)	T: 0.845 (174)	T: 0.729 (153)	T: 0.864 (171)	T: 1.000 (978)	T: 1.000 (1006)	T: 1.000 (1322)
BRAP	C: 0.281 (4825)	C: 0.318 (3824)	C: 0.250 (52)	C: 0.201 (825)	C: 0.043 (8)	C: 0.155 (32)	C: 0.271 (57)	C: 0.136 (27)			
rs11066015	G: 0.735 (12 582)	G: 0.700 (8425)	G: 0.864 (171)	G: 0.816 (3326)	G: 0.957 (178)	G: 0.840 (173)	G: 0.719 (151)	G: 0.760 (158)	G: 1.000 (978)	G: 1.000 (1006)	G: 1.000 (1322)
ACAD10	A: 0.265 (4548)	A: 0.300 (3619)	A: 0.136 (27)	A: 0.184 (752)	A: 0.043 (8)	A: 0.160 (33)	A: 0.281 (59)	A: 0.240 (50)			
rs671	G: 0.714 (12 353)	G: 0.699 (8422)	G: 0.760 (158)	G: 0.727 (3098)	G: 0.957 (178)	G: 0.840 (173)	G: 0.729 (153)	G: 0.864 (171)	G: 1.000 (978)	G: 1.000 (1006)	G: 0.998 (1320)
ALDH2	A: 0.286 (4960)	A: 0.301 (3622)	A: 0.240 (50)	A: 0.273 (1163)	A: 0.043 (8)	A: 0.160 (33)	A: 0.271 (57)	A: 0.136 (27)			A: 0.002 (2)
rs2074356	G: 0.757 (12 991)	G: 0.722 (8696)	G: 0.788 (164)	G: 0.834 (3417)	G: 0.968 (180)	G: 0.898 (185)	G: 0.819 (172)	G: 0.894 (177)	G: 1.000 (978)	G: 1.000 (1006)	G: 1.000 (1322)
HECTD4	A: 0.243 (4164)	A: 0.278 (3348)	A: 0.212 (44)	A: 0.166 (686)	A: 0.032 (6)	A: 0.102 (21)	A: 0.181 (38)	A: 0.106 (21)			
rs11066280	T: 0.719 (12 326)	T: 0.687 (8272)	T: 0.764 (159)	T: 0.797 (3264)	T: 0.866 (161)	T: 0.811 (167)	T: 0.705 (148)	T: 0.783 (155)	T: 0.996 (974)	T: 1.000 (1006)	T: 0.999 (1321)
HECTD4	A: 0.281 (4824)	A: 0.313 (3772)	A: 0.236 (49)	A: 0.204 (834)	A: 0.134 (25)	A: 0.189 (39)	A: 0.295 (62)	A: 0.217 (43)	A: 0.004 (4)		A: 0.001 (1)
rs11917356	A: 0.308 (5274)	A: 0.309 (3716)	A: 0.284 (59)	A: 0.288 (1180)	A: 0.462 (86)	A: 0.369 (76)	A: 0.362 (76)	A: 0.409 (81)	A: 0.723 (707)	A: 0.847 (852)	A: 0.604 (799)
COL6A5	G: 0.692 (11 872)	G: 0.691 (8328)	G: 0.716 (149)	G: 0.712 (2914)	G: 0.538 (100)	G: 0.631 (130)	G: 0.638 (134)	G: 0.591 (117)	G: 0.277 (271)	G: 0.153 (154)	G: 0.396 (523)

Abbreviations: CDX, Chinese Dai in Xishuangbanna, China; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; JP-Inabe, Japanese in Inabe city, Mie, Japan; JP-Tohoku, Japanese in Tohoku, Japan; JPT, Japanese in Tokyo, Japan; KHV, Kinh in Ho Chi Minh City, Vietnam; SNP, single-nucleotide polymorphism.
Values indicate the allele percentages, and the observed numbers are indicated in parentheses.
[a]Allele frequency obtained from 1000 Genomes Project database.
[b]Allele frequency obtained from Integrative Japanese Genome Variation Database (2KJPN) from the Tohoku Medical Megabank Organization of Tohoku University (ToMMo).

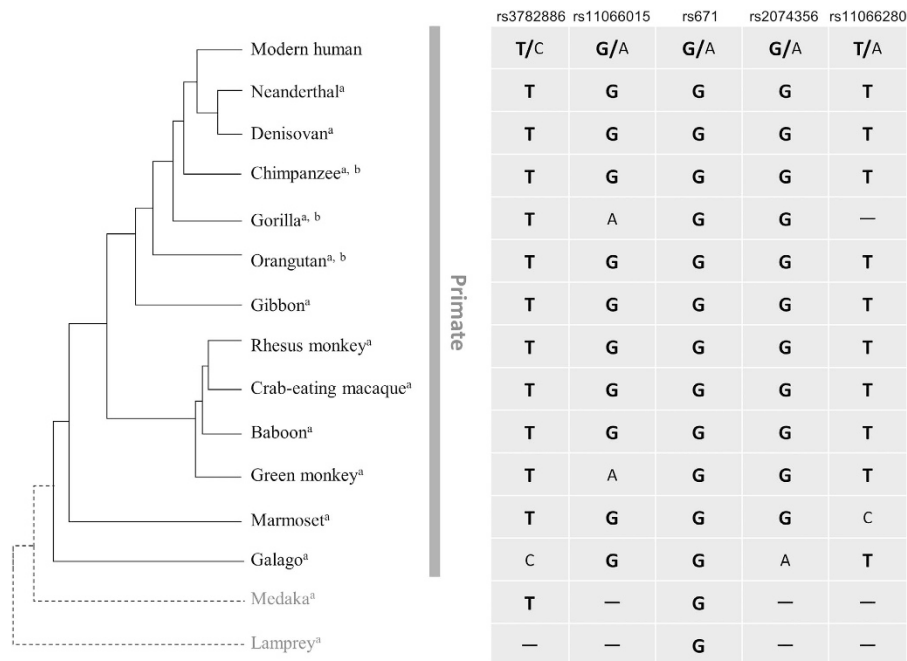SNV, while the 'CAAAA' haplotype comprises all derived alleles.

The haplotype 'CAAAA' is East Asian-specific. However, whether this haplotype is specific to the modern human remains unknown. To address this question, we examined putative haplotypes in other vertebrates based on information in the UCSC Genome Browser or Multiz Alignments of 100 Vertebrates in the UCSC database (Figure 2; Supplementary Table S2). No other species with the East Asian-specific haplotype of modern humans were observed and the common haplotype was predominantly observed in Haplorhini, and even archaic humans, including Neanderthal and Denisovan. Notably, a guanine (ancestral allele) at rs671 in ALDH2 appeared to be consistently conserved in all vertebrates, including a jawless vertebrate, sea lamprey (Petromyzon marinus). Therefore, the nonsynonymous substitution at this site might be disadvantageous for survival in vertebrates.

To examine genetic polymorphisms within nonhuman primates, variant call format data sets of 38 chimpanzees (Pan troglodytes and Pan paniscus), 31 gorillas (Gorilla gorilla and Gorilla beringei) and 10 orangutans (Pongo pygmaeus and Pongo abelii) were retrieved from the Great Ape Genome Project database.[23] We subsequently searched allele frequencies at the counterparts to five human hypertension-related SNVs densely located at 12q24.12–12q24.13. However, the variant call format data sets did not contain any information for the allele frequencies, suggesting that the three primates examined do not have genetic polymorphisms at these sites. Therefore, it is likely that the East Asian-specific haplotype identified in the present study is human-specific.

Comparison of clinical parameters between East Asian-specific and common haplotypes

In the Inabe cohort, the prevalence of hypertension in subjects with the East Asian haplotype (case = 1052 and control = 2111) was significantly lower than in subjects with the common haplotype (case = 4875 and control = 3141; odds ratio = 0.32, $P < 2.2 \times 10^{-16}$ by Fisher's exact test). The odds ratio was lower than that of each SNV (the mean odds ratio was 0.78, see above), suggesting that the effect of haplotype on hypertension is greater than that of individual SNVs. In the combined group of homozygotes with the East Asian haplotype (CAAAA/CAAAA) and heterozygotes with East Asian and common haplotypes (CAAAA/TGGGT), the mean SBP and DBP values were 119.6 ± 0.32 and 73.6 ± 0.24, respectively. These values were significantly lower than the mean SBP (121.5 ± 0.31) and DBP (75.5 ± 0.24) values of homozygotes with the common haplotype ($P < 1.8 \times 10^{-5}$ and $< 1.6 \times 10^{-8}$ for SBP and DBP, respectively, using Welch's t-test), according to cross-sectional analysis using clinical parameters of subjects in the latest year for which data were available. These results suggest that the East Asian haplotype is protective against hypertension.

To survey the effect of different lifestyle variables (drinking, smoking and exercise status) on hypertension among diplotypes, we compared the SBP and DBP values of each diplotype using cross-sectional analysis data in a pairwise fashion

	rs3782886 T/C	rs11066015 G/A	rs671 G/A	rs2074356 G/A	rs11066280 T/A
Modern human	T	G	G	G	T
Neanderthal[a]	T	G	G	G	T
Denisovan[a]	T	G	G	G	T
Chimpanzee[a,b]	T	G	G	G	T
Gorilla[a,b]	T	A	G	G	—
Orangutan[a,b]	T	G	G	G	T
Gibbon[a]	T	G	G	G	T
Rhesus monkey[a]	T	G	G	G	T
Crab-eating macaque[a]	T	G	G	G	T
Baboon[a]	T	G	G	G	T
Green monkey[a]	T	A	G	G	T
Marmoset[a]	T	G	G	G	C
Galago[a]	C	G	G	A	T
Medaka[a]	T	—	G	—	—
Lamprey[a]	—	—	G	—	—

Primate

Figure 2 Comparison of five hypertension-associated single-nucleotide variants (SNVs) among primates. The primate allele identical to the human ancestral allele is shown in bold. [a]The putative allele was estimated based on the UCSC Genome Browser or Multiz Alignments of 100 Vertebrates in the UCSC database. [b]The putative allele was estimated based on allele frequencies deposited in the Great Ape Genome Project database.

(for example, smokers and non-smokers). The mean SBP and DBP values of smokers and drinkers were significantly higher than those of non-smokers and non-drinkers among subjects homozygous for the common diplotypes (Supplementary Figure S4a–d). Moreover, the comparison of DBP values in heterozygotes with the common and East Asian haplotypes showed significant differences between groups based on smoking or drinking status. The differences in the DBP values of smokers and non-smokers with East Asian haplotypes were smaller than those in subjects with a common haplotype. Stratification based on smoking status revealed the degree of difference between the two groups with the 'CAAAA/CAAAA' diplotype was 0.35 mm Hg, and the degree of difference between the two groups with 'TGGGT/CAAAA' and 'TGGGT/TGGGT' diplotypes was 3.01 and 4.33 mm Hg, respectively (Supplementary Figure S4b). This finding suggests that compared with the common haplotype, the East Asian-specific haplotype is less susceptible to the adverse effects of smoking. Using all six detected SNVs in the dominant model, we also estimated phased haplotypes and repeated the stratification analysis. Similar results were observed in analyses based on six SNVs (including rs12229654 at 12q24.11; Supplementary Figure S5a–d).

Identification of a novel genetic variant in the COL6A5 gene
For the recessive model, one SNV (rs11917356) in COL6A5 showed significant association with SBP in the GEE model with normal and t reference distributions, and the value of approxdf was 60.3. We compared the mean SBP and DBP values between genotypes for the candidate SNV. SBP and DBP

values in subjects with the GG or GA genotype (mean SBP = 120.9 ± 0.22 and mean DBP = 74.8 ± 0.17) were significantly higher than those in subjects with the AA genotype (mean SBP = 117.7 ± 0.64 and mean DBP = 72.7 ± 0.48; $P = 2.4 \times 10^{-6}$ and 5.1×10^{-5} for mean SBP and mean DBP, respectively, by Welch's t-test). This finding suggests that the 'A' allele is protective against hypertension.

On the basis of the information obtained from the 1000 Genomes Project[21] and iJGVD[22] databases, we examined allele frequencies of focal SNVs within human populations (Table 3). Remarkably, the frequency of the derived allele 'G' at rs11917356 was highest in Japan (~70%), and the 'G' allele was major in East Asian populations and minor in non-East Asian populations, suggesting that the frequencies of the derived alleles have rapidly increased since the split of East Asian and South Asian lineages.

DISCUSSION
In the present longitudinal EWAS, the GEE model with normal and t reference distributions demonstrated that six SNVs located on chromosome 12 were significantly associated with the clinical parameters of hypertension, while one SNV of COL6A5 on chromosome 3 was significantly related to SBP. The effective sample size based on approxdf of these SNVs was > 30. In a CHAGE multistudy meta-analysis, the GEE model using t reference distribution with approxdf ≥ 10 provided more accurate P-values than standard GEE models.[11] Therefore, the reliability of associations between the clinical parameters of hypertension and the candidate SNVs detected in the present study was high.

On the basis of the data for the six candidate SNVs on chromosome 12, we showed that East Asian-specific alleles (or haplotype) could be protective against hypertension. Among the six SNVs, a nonsynonymous nucleotide substitution at rs671 in *ALDH2* altered an amino-acid residue (Table 2). However, all nucleotide substitutions in the other SNVs were silent. Therefore, the SNV in *ALDH2* may be an important factor in the onset of hypertension. ALDH2 is a mitochondrial acetaldehyde dehydrogenase involved in the major pathway of alcohol metabolism. The rs671 G to A transition in *ALDH2* exon 12 changes the amino-acid at position 504 (E504K), resulting in defective enzyme activity.[24,25] Previous studies reported that the derived *ALDH2* allele (A) was predominantly observed in East Asians, and this allele has potentially expanded in this area through natural selection in the recent past.[26,27] A meta-analysis of GWASs for blood pressure variation in East Asian populations detected significant associations between rs11066280 of *HECTD4* and SBP or DBP values, and an estimated haplotype comprising eight SNVs showed relatively high LD ($r^2 > 0.7$). Furthermore, haplotypes comprising five of the eight SNVs were clearly differentiated into high- and low-blood-pressure groups.[5] In the present longitudinal EWAS, four of the five SNVs detected in the previous study (rs3782886 of *BRAP*, rs671 of *ALDH2*, and rs2074356 and rs11066280 of *HECTD4*) were significantly associated with hypertension. There was high LD among SNVs densely located at 12q24.1, suggesting that positive selection, acting on the site of rs671 in *ALDH2*, alters the allelic frequency of the other five SNVs through genetic hitchhiking. Indeed, a signal indicative of a selective sweep (recent positive selection) was identified at 12q24.13 near the *ALDH2* in two haplotype-based tests.[5] Nevertheless, in the present study, the ancestral alleles at the six detected SNVs appeared to be conserved during the course of primate evolution. In addition, the protective effect of the East Asian haplotype on hypertension (the odds ratio = 0.3) was greater than that of each East Asian allele (the mean odds ratio = 0.8). Therefore, the effect of genetic factors other than rs671 remains controversial. It is intriguing to determine how and why the haplotype generated by unusual alleles at each SNV has spread in East Asians.

As rs671 of *ALDH2* influences alcohol intake or consumption, the drinking status of subjects is potentially associated with hypertension. Cross-sectional analysis using the clinical parameters for subjects in the latest year for which data were available revealed that the prevalence of hypertension in non-drinkers with the East Asian allele at rs671 was significantly lower than that in non-drinkers with the common allele ($P = 0.009$ by Fisher's exact test, odds ratio = 0.87). To further consider the effect of alcohol intake and consumption, we re-analyzed the relationship between longitudinal changes in the prevalence of hypertension, SBP, or DBP values and six detected SNVs in the dominant model using the GEE method after adjustment for alcohol intake or alcohol consumption (also including age, gender, body mass index and smoking status; Supplementary Table S3). The associations of all six SNVs with clinical parameters of hypertension were supported by the results of the GEE test, suggesting that the identified genetic variants at 12q24.1 could influence susceptibility to hypertension.

Given that their genetic variations influence the incidence of hypertension, it is unclear how these variants are involved in the pathogenesis of hypertension. While the immediate cause is unknown, differences in sodium retention might be associated with hypertension. Hence, we preliminarily analyzed the interaction between molecules encoded by genes with the five hypertension-associated SNVs and molecules directly or indirectly related to the regulation of sodium retention (such as RAC1, ADRB2, WNK4, NR3C2, SLC12A3, SCNN1A and SCNN1B). To perform this analysis, we used Cytoscape version 3.4.0[28] software, which automatically integrates human molecular interaction data from different public databases (Supplementary Figure S6). The integrated network did not show any direct interaction among the analyzed molecules. However, the network displayed some potential indirect interactions. For example, HECTD4 was associated with YWHAZ, which is involved in the regulation of signaling pathways. Meek *et al.*[29] previously showed the physical association of these molecules. According to the molecular INTeraction database (http://mint.bio.uniroma2.it), Pozuelo-Rubio *et al.*[30] reported the association of YWHAZ and WNK4, a regulator of renal electrolyte transport, using tandem affinity purification. WNK4 has an important role in blood pressure regulation.[31] In addition, in the generated network, ALDH2 was indirectly associated with the WNK4 via CUL3, a component of ubiquitin E3 ligase, according to the Biological General Repository for Interaction Data sets (https://thebiogrid.org/). These results are not direct evidence that the four hypertension susceptibility genes identified at 12q24.1 in the present study influence the sodium retention ability, and further analyses are required to reveal the pathogenesis of hypertension.

Comparison of the DBP values between smokers and non-smokers in each diplotype of phased haplotypes (East Asian and common haplotypes) showed a clear pattern. The degree of differences of DBP values between the two groups defined according to smoking status decreased with the number of East Asian haplotypes in the diplotype. These data were derived from the cross-sectional analysis using clinical parameters of subjects in the latest year for which data were available. Cigarette smoking potentially increases the risk of cardiovascular diseases.[32,33] Nevertheless, in the present study, subjects with the East Asian haplotype were less susceptible to the adverse effects of smoking on hypertension compared with those with the common haplotype. Thus, the interaction between the genetic factor of common haplotype and the environmental factor of smoking may intensify the adverse effects on hypertension.

The longitudinal EWAS for the recessive model showed one candidate SNV in *COL6A5*. According to information on GWAS central (http://www.gwascentral.org/) and GWAS catalog (https://www.ebi.ac.uk/gwas/) databases, rs11917356 is a novel genetic variant associated with hypertension. In addition,

the frequency of derived allele 'G' was remarkably higher in East Asia (particularly in Japan) than in the other ethnic populations. This finding suggests that the derived allele rapidly spread throughout East Asia. However, the mean SBP and DBP values in subjects with the derived allele were significantly higher than those in subjects with the ancestral allele 'A'. Given that the 'G' allele in Japanese individuals could be susceptible to hypertension, it remains unclear why the frequency of the allele has increased in East Asia. It is possible that the derived allele confers adaptability to environments in East Asia in return for the risk of hypertension incidence.

COL6A5 (collagen type VI alpha 5 chain) is a member of the collagen superfamily. A previous study reported that COL6A5 is a susceptibility gene for the chronic inflammatory skin disorder atopic dermatitis.[34] The rs11917356 represents a G > A transition that alters the amino-acid residue at position 982 (D982G), but this substitution is unlikely to have a strong impact on protein function, according to SIFT (http://sift.jcvi.org/, score = 0.73) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml, score = 0.093). The COL6A5 gene is primarily expressed in the skin but is also expressed in the lungs, small intestine, colon and testes.[34] In addition, COL6A5 protein is strongly expressed around blood vessels at the interface between the papillary and reticular dermis.[35] Thus, the amino-acid change in COL6A5 may be associated with hypertension through changes in blood vessel elasticity (or stiffness). Although the pathogenesis of hypertension resulting from genetic variation at rs11917356 remains unclear, the results of the present study suggest that the genetic variant of COL6A5 confers susceptibility to hypertension.

There were certain limitations in the present study. First, the longitudinal EWAS was conducted only in a local Japanese population. Given that the subjects were community-dwelling individuals who visited the health-care center for an annual health check-up, the selection bias was small in the present study. Although multicenter longitudinal EWASs improved the accuracy and reliability of these results, those data are not currently available. Therefore, the replication of longitudinal EWASs in other Japanese populations or other ethnic groups is required to clarify the associations of the identified SNVs with hypertension. Second, the functional relevance of candidate SNVs identified in the longitudinal EWAS to the pathogenesis of hypertension remains unclear. Third, the follow-up period of annual health check-ups varied from 1 to 11 years among individuals.

In conclusion, the results of the present study suggest that the East Asian-specific haplotype, comprising SNVs genotyped on 12q24.1 and a COL6A5 ancestral allele of which the frequency may have recently decreased in East Asians, may represent protective genes against essential hypertension. In addition, the East Asian-specific haplotype may be less susceptible to the adverse effects of smoking on hypertension. Clinical genotyping for the SNVs detected in this longitudinal EWAS may be useful for precise and/or personalized medicine.

PUBLISHER'S NOTE

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1 Ministry of Health, Labour and Welfare Patient Survey 2016, Available at http://www.mhlw.go.jp/english/database/db-hss/ accessed 4 January 2017.
2 Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L et al. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 2009; 41: 666–676.
3 Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A et al. Genome-wide association study of blood pressure and hypertension. Nat Genet 2009; 41: 677–687.
4 International Consortium for Blood Pressure Genome-Wide Association Studies, Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011; 478: 103–109.
5 Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat Genet 2011; 43: 531–538.
6 Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, Ban H-J et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat Genet 2009; 41: 527–534.
7 Simino J, Shi G, Bis JC, Chasman DI, Ehret GB, Gu X 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.
8 Yamada Y, Matsui K, Takeuchi I, Oguri M, Fujimaki T. Association of genetic variants with hypertension in a longitudinal population-based genetic epidemiological study. Int J Mol Med 2015; 35: 1189–1198.
9 Perloff D, Grim C, Flack J, Frohlich ED, Hill M, McDonald M et al. Human blood pressure determination by sphygmomanometry. Circulation 1993; 88: 2460–2470.
10 Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD et al. Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium. PLoS ONE 2013; 8: e68095.
11 Sitlani CM, Rice KM, Lumley T, McKnight B, Cupples LA, Avery CL et al. Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. Stat Med 2015; 34: 118–130.
12 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38: 904–909.
13 R Core Team R: A language and environment for statistical computing 2016, Available at http://www.r-project.org/ [21 Dec. 2016].
14 RStudio Team RStudio: Integrated Development Environment for R. RStudio, Inc. 2015, Available at http://www.rstudio.com/ [8 Jan. 2017].
15 Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986; 73: 13–22.
16 Hanley JA, Negassa A, Edwardes MD, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. Am J Epidemiol 2003; 157: 364–375.
17 Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw 2006; 15: 1–11.
18 Voorman A, Rice K, Lumley T. Fast computation for genome-wide association studies using boosted one-step statistics. Bioinformatics 2012; 28: 1818–1822.

19 Howie BN, Donnelly P, Marchini J, Rioux J, Xavier R, Taylor K *et al.* A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.

20 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

21 The 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.

22 Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y *et al.* Rare variant discovery by deep whole-genome sequencing of 1070 Japanese individuals. *Nat Commun* 2015; **6**: 8018.

23 Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B *et al.* Great ape genetic diversity and population history. *Nature* 2013; **499**: 471–475.

24 Hsu LC, Bendel RE, Yoshida A. Direct detection of usual and atypical alleles on the human aldehyde dehydrogenase-2 (ALDH$_2$) locus. *Am J Hum Genet* 1987; **41**: 996–1001.

25 Crabb DW, Edenberg HJ, Bosron WF, Li TK. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive *ALDH2^2* allele is dominant. *J Clin Invest* 1989; **83**: 314–316.

26 Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SLB *et al.* The evolution and population genetics of the *ALDH2* locus: random genetic drift, selection, and low levels of recombination. *Ann Hum Genet* 2004; **68**: 93–109.

27 Luo HR, Wu GS, Pakstis AJ, Tong L, Oota H, Kidd KK *et al.* Origin and dispersal of atypical aldehyde dehydrogenase *ALDH2*487Lys. Gene* 2009; **435**: 96–103.

28 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**: 2498–2504.

29 Meek SEM, Lane WS, Piwnica-Worms H. Comprehensive proteomic analysis of interphase and mitotic 14-3-3-binding proteins. *J Biol Chem* 2004; **279**: 32046–32054.

30 Pozuelo-Rubio M. Proteomic and biochemical analysis of 14-3-3-binding proteins during C2-ceramide-induced apoptosis. *FEBS J* 2010; **277**: 3321–3342.

31 Wu G, Peng JB. Disease-causing mutations in KLHL3 impair its effect on WNK4 degradation. *FEBS Lett* 2013; **587**: 1717–1722.

32 Najem B, Houssière A, Pathak A, Janssen C, Lemogoum D, Xhaët O *et al.* Acute cardiovascular and sympathetic effects of nicotine replacement therapy. *Hypertension* 2006; **47**: 1162.

33 Bowman TS, Gaziano JM, Buring JE, Sesso HD. A prospective study of cigarette smoking and risk of incident hypertension in women. *J Am Coll Cardiol* 2007; **50**: 2085–2092.

34 Söderhäll C, Marenholz I, Kerscher T, Rüschendorf F, Esparza-Gordillo J, Worm M *et al.* Variants in a novel epidermal collagen gene (*COL29A1*) are associated with atopic dermatitis. *PLoS Biol* 2007; **5**: 1952–1961.

35 Sabatelli P, Gara SK, Grumati P, Urciuolo A, Gualandi F, Curci R *et al.* Expression of the collagen VI α5 and α6 chains in normal human skin and in skin of patients with collagen VI-related myopathies. *J Invest Dermatol* 2011; **131**: 99–107.

Supplementary Information accompanies the paper on Experimental & Molecular Medicine website (http://www.nature.com/emm)