ORIGINAL ARTICLE

Evolutionary history of disease-susceptibility loci identified in longitudinal exome-wide association studies

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

[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

[9]Department of Cardiovascular Medicine, Gifu Prefectural Tajimi Hospital, Tajimi, Japan

Correspondence
Yoshiki 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.
Email: hyasukou@proof.ocn.ne.jp

Funding information
Hitachi Global Foundation, Grant/Award Number: 1323; Okasan Kato Culture Promotion Foundation, Grant/Award Number: OKF-17-1-1; Core Research for Evolutional Science and Technology, Grant/Award Number: JPMRJCR1302; Japan Society for the Promotion of Science, Grant/Award Number: JP15H04772 and JP17H00758

Abstract

Background: Our longitudinal exome-wide association studies previously detected various genetic determinants of complex disorders using ~26,000 single-nucleotide polymorphisms (SNPs) that passed quality control and longitudinal medical examination data (mean follow-up period, 5 years) in 4884–6022 Japanese subjects. We found that allele frequencies of several identified SNPs were remarkably different among four ethnic groups. Elucidating the evolutionary history of disease-susceptibility loci may help us uncover the pathogenesis of the related complex disorders.

Methods: In the present study, we conducted evolutionary analyses such as extended haplotype homozygosity, focusing on genomic regions containing disease-susceptibility loci and based on genotyping data of our previous studies and datasets from the 1000 Genomes Project.

Results: Our evolutionary analyses suggest that derived alleles of rs78338345 of *GGA3*, rs7656604 at 4q13.3, rs34902660 of *SLC17A3*, and six SNPs closely located at 12q24.1 associated with type 2 diabetes mellitus, obesity, dyslipidemia, and three complex disorders (hypertension, hyperuricemia, and dyslipidemia), respectively, rapidly expanded after the human dispersion from Africa (Out-of-Africa). Allele frequencies of *GGA3* and six SNPs at 12q24.1 appeared to have remarkably changed in East Asians, whereas the derived alleles of rs34902660 of *SLC17A3* and rs7656604 at 4q13.3 might have spread across Japanese and non-Africans, respectively, although

we cannot completely exclude the possibility that allele frequencies of disease-associated loci may be affected by demographic events.

Conclusion: Our findings indicate that derived allele frequencies of nine disease-associated SNPs (rs78338345 of *GGA3*, rs7656604 at 4q13.3, rs34902660 of *SLC17A3*, and six SNPs at 12q24.1) identified in the longitudinal exome-wide association studies largely increased in non-Africans after Out-of-Africa.

1 | INTRODUCTION

Recent genome-wide association studies (GWASs) have identified various genetic variants that confer susceptibility to hypertension (International Consortium for Blood Pressure Genome-Wide Association Studies et al., 2011; Kato et al., 2011; Levy et al., 2009; Liu et al., 2016; Newton-Cheh et al., 2009; Simino et al., 2014), chronic kidney disease ([CKD]; Böger et al., 2011; Chambers et al., 2010; Okada, Sim, et al., 2012; Pattaro et al., 2016), hyperuricemia (Chittoor et al., 2016; Köttgen et al., 2013; Matsuo et al., 2008; Nakayama et al., 2017; Phipps-Green et al., 2016), type 2 diabetes mellitus ([T2DM]; Cho et al., 2012; Hara et al., 2014; Imamura & Maeda, 2011; Imamura et al., 2016; Mahajan et al., 2014; Matsuba et al., 2016; Morris et al., 2012), obesity (Dorajoo et al., 2017; Fall & Ingelsson, 2014; Locke et al., 2015; Okada, Kubo, et al., 2012; Scott et al., 2016), metabolic syndrome ([MetS]; Fall & Ingelsson, 2014; Kraja et al., 2011; Povel, Boer, Reiling, & Feskens, 2011; Zabaneh & Balding, 2010), dyslipidemia [hypertriglyceridemia, hyper-low density lipoprotein (LDL)-cholesterolemia, and hypo-high density lipoprotein (HDL)-cholesterolemia] (Aguilar-Salinas, Tusie-Luna, & Pajukanta, 2014; Asselbergs et al., 2012; Barber et al., 2010; Jia et al., 2016; Kathiresan et al., 2009; Kooner et al., 2008; Kurano et al., 2016; Lange et al., 2014; Teslovich et al., 2010), coronary artery disease ([CAD]; Golbus et al., 2016; Lu et al., 2012; Nikpay et al., 2015; Schunkert et al., 2011; Wakil et al., 2016), and cerebral infarction (CI; Akinyemi et al., 2015; Kubo et al., 2007; Meschia et al., 2011; Yamada et al., 2009) in diverse ethnic groups. In addition, genetic variants associated with hematological traits have been identified by previous GWASs (Kamatani et al., 2010; Li et al., 2013; Lo et al., 2011; Mousas et al., 2017; Okada & Kamatani, 2012; van Rooij et al., 2017). Information on the association of numerous genetic variants with various disease-related traits is publicly available in several databases such as Genome-Wide Repository of Associations Between SNPs and Phenotypes (GRASP) (https://grasp.nhlbi.nih.gov/Overview.aspx; Leslie, O'Donnell, & Johnson, 2014), GWAS

Catalogue (https://www.ebi.ac.uk/; MacArthur et al., 2017), and DisGeNET (http://www.disgenet.org/web/DisGeNET/; Piñero et al., 2015). Most conventional GWASs, however, have been conducted in a cross-sectional manner that measured traits at a single point in time.

In our previous studies, therefore, longitudinal exome-wide association studies were performed to explore novel genetic determinants of complex disorders (hypertension, T2DM, dyslipidemia, CAD, CI, CKD, hyperuricemia, obesity, and MetS) as well as 13 hematological traits in 4884–6022 Japanese subjects who had undergone annual health checkups for several years (Yasukochi et al., 2017, 2018a, 2018b, 2018c, 2018d, 2018e, 2019). In these studies, we identified 272 single-nucleotide polymorphisms (SNPs) associated with complex disorders or hematological markers, using ~26,000 SNPs that passed quality control and longitudinal medical examination data of the Japanese subjects. Of these SNPs, 28 (one SNP for hypertension, two for T2DM, two for hyper-LDL-cholesterolemia, four for hypo-HDL-cholesterolemia, three for CAD, three for CI, one for hyperuricemia, three for obesity, and nine for hematological traits) were identified as novel genetic determinants (Table 1). A whole genome sequence analysis facilitates to account for the missing heritability of complex traits and diseases (Wainschtein et al., 2019) because it covers coding and noncoding variants, including rare variants in regions of low linkage disequilibrium (LD). The generalized estimating equation method implemented in our previous studies can increase the chance of type I error due to a small effective sample size (Sitlani et al., 2015). After quality controls for the genotyping data in the longitudinal exome-wide association studies, SNPs with the small effective sample size (minor allele frequency [MAF] of <5%) were filtered out. Nevertheless, our longitudinal exome-wide association studies identified 28 SNPs that were not found in previous GWASs (Yasukochi et al., 2017, 2018a, 2018b, 2018c, 2018d, 2018e, 2019). This suggests that the longitudinal exome-wide association study is useful for finding the missing heritability of complex traits and diseases even though rare variants are removed.

TABLE 1 Novel genetic determinants of complex disorders or hematological markers identified by longitudinal exome-wide association studies in the Inabe cohort

Disease or Trait	RefSNP ID	Allele[a]	Position[b]	Position[c]	Gene or locus	Amino acid change[d]	Reference
Hypertension	rs11917356	G → A	3:130110550	3:130391707	COL6A5	G982D	Yasukochi et al., 2017
Hyperuricemia	rs55975541	G → A	11:64597201	11:6482929	CDC42BPG	R1237W	Yasukochi et al., 2018a
Hematological traits	rs3917688	G → A	1:169591080	1:169621842	SELP		Yasukochi et al., 2018b
	rs7584099	G → A	2:148478336	2:147720767	2q22.3		
	rs4686683	C → A	3:185307363	3:185589575	SENP2		
	rs13121954	G → A	4:148023829	4:147102677	4q31.2		
	rs395967	A → G	5:38842959	5:38842857	OSMR-AS1		
	rs1579219	G → A	6:30224305	6:30256528	HCG17		
	rs12338	G → C	8:11710888	8:11853379	CTSB	L26V	
	rs3133745	C → T	8:96534806	8:95522578	C8orf37-AS1		
	rs10757049	A → G	9:19281501	9:19281503	DENND4C		
Dyslipidemia	rs34902660	C → A	6:25851102	6:25850874	SLC17A3	G239V	Yasukochi et al., 2018c
	rs1042127	T → G	6:31084170	6:31116393	CDSN	S408A	
	rs74416240	A → G	12:110342598	12:109904793	GIT2	N387S	
	rs925368	G → A	12:110390979	12:109953174	TCHP	E152E	
	rs7969300	T → C	12:111993712	12:111555908	ATXN2	N248S	
	rs12231744	C → T	12:112477055	12:112039251	NAA25	R876K	
Obesity	rs9491140	C → T	6:124691237	6:124370091	NKAIN2		Yasukochi et al., 2018d
	rs145848316	C → A	7:151882672	7:152185587	KMT2C	A1685S	
	rs7863248	T → C	9:88308127	9:85693212	AGTPBP1		
CAD/CI	rs4606855	G → C	19:14769339	19:14658527	ADGRE3	E75Q	Yasukochi et al., 2018e
	rs7132908	G → A	12:50263148	12:49869365	FAIM2		
	rs6580741	G → C	12:50727706	12:50333923	FAM186A	H2228Q	
	rs1324015	G → A	13:43727849	13:43153713	LINC00400		
	rs884205	G → T	18:60054857	18:62387624	TNFRSF11A		
	rs3746414	G → A	20:50769379	20:52152840	ZFP64	S451N	
T2DM	rs6414624	C → T	4:5743512	4:5741785	EVC	H258Y	Yasukochi et al., 2019
	rs78338345	C → G	17:73238509	17:75242428	GGA3	E147Q	

Abbreviations: CAD, coronary artery disease. CI, cerebral infarction; T2DM, type 2 diabetes mellitus.

[a]Major allele → minor allele.

[b]Position in NCBI build GRCh37.p13.

[c]Position in NCBI build GRCh38.p10.

[d]Splice variants of amino acid substitution are not shown.

Of the 28 SNPs identified in our longitudinal exome-wide association studies, allele frequencies of some disease-susceptibility loci are clearly different between Africans (AFRs) and non-AFRs, according to allele frequency data of four ethnic populations from the 1000 Genomes Project (1KGP; The 1000 Genomes Project Consortium, 2010). This suggests that derived or ancestral allele frequencies of disease-susceptibility loci may have increased after the divergence of non-AFR and AFR ancestral populations. Approximately 135,000–40,000 years ago, anatomically modern humans dispersed throughout the world from Africa (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011; Groucutt et al., 2015; Li & Durbin, 2011; López, van Dorp, & Hellenthal, 2015; Olivieri et al., 2006; Reyes-Centeno, Hubbe, Hanihara, Stringer, & Harvati, 2015; Soares et al., 2009; Wall, 2017), although the dating remains controversial because of a recent finding of older anatomically modern human fossils (194,000–177,000 years ago) outside Africa (Hershkovitz et al., 2018). These migrations are called "Out-of-Africa". After the exodus from Africa, positive selection has acted on several genomic regions in non-AFRs to adapt to new environments (Bigham et al., 2010; Crawford et al., 2017;

Pickrell et al., 2009; Sabeti et al., 2007; Simonson, Huff, Witherspoon, Prchal, & Jorde, 2015; Voight, Kudaravalli, Wen, & Pritchard, 2006). However, it is possible that positively selected variants might increase susceptibility to complex disorders in return for environmental adaptability (Okumiya et al., 2010). In the present study, we have estimated the evolutionary history of genomic regions around disease-associated loci identified in our previous studies. These evolutionary inferences may be helpful in elucidating how and why allelic variants conferring susceptibility to complex disorders have been disseminated in modern humans.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

The study protocol complies 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 prior to enrollment in the present study.

2.2 | Datasets

In previous longitudinal exome-wide association studies (Yasukochi et al., 2017, 2018a, 2018b, 2018c, 2018d, 2018e, 2019), 6026 community-dwelling individuals were recruited from population-based cohort studies (Inabe Health and Longevity Study) in Inabe City, Mie, Japan (Oguri et al., 2017; Yamada, Matsui, Takeuchi, & Fujimaki, 2015a, 2015b; Yamada, Matsui, Takeuchi, Oguri, & Fujimaki, 2015a, 2015b). These individuals visited Inabe General Hospital for an annual health checkup, with a mean follow-up period of 5 ± 3 years (covering April 2003–March 2014). We refer to this cohort as "Inabe cohort", and used SNP data genotyped by Infinium HumanExome-12 ver. 1.2 BeadChip and Infinium Exome-24 ver. 1.0 BeadChip (Illumina) (Grove et al., 2013) in our previous studies to determine their allele frequencies. The frequencies in the Inabe cohort are listed in Table 2 and Table S1.

We conducted evolutionary analyses using datasets retrieved from public databases. Information regarding allele frequencies of the target SNPs within four ethnic groups [East Asian (EAS), South Asian (SAS), European (EUR), and AFR] was obtained from the 1000 Genomes Project (http://www.internationalgenome.org/; The 1000 Genomes Project Consortium, 2010) using the Ensembl genome browser (http://www.ensembl.org; Zerbino et al., 2018). The allele frequencies in the four ethnic groups are listed in Table 2 and Table S1. The categories of the ethnic populations are listed at the following URL: http://www.internationalgenome.org/data-portal/population. We also obtained allele frequency data

of target SNPs using The Exome Aggregation Consortium (ExAC) Browser (http://exac.broadinstitute.org/). Information on the allele frequencies in great apes (chimpanzee, gorilla, and orangutan) was obtained from the Great Ape Genome Project (GAGP) database (http://biologiaevolutiva.org/greatape/; Prado-Martinez et al., 2013). Alleles with high frequency in AFRs of present-day humans, archaic humans, and chimpanzees were defined as ancestral alleles.

A nucleotide at the homologous target SNPs within vertebrate reference genomes was investigated in the Multiz Alignments of 100 Vertebrates (Blanchette et al., 2004) and PhyloP score (Pollard, Hubisz, Rosenbloom, & Siepel, 2010) in the University of California, Santa Cruz (UCSC) Genome Browser database (http://genome.ucsc.edu; Kent et al., 2002). An effect of amino acid substitution on protein function was predicted by SIFT (Ng & Henikoff, 2003; Vaser, Adusumalli, Leng, Sikic, & Ng, 2016), PolyPhen-2 (Adzhubei et al., 2010), and Combined Annotation Dependent Depletion ([CADD]; Rentzsch, Witten, Cooper, Shendure, & Kircher, 2019) scores.

2.3 | Estimates of LD

Genotype data used in the present study were formatted for the programs described below using R software version 3.5.1 (R Foundation for Statistical Computing; R Core Team, 2018) through RStudio 1.1.456 (RStudio, Boston, MA; http://www.rstudio.com/) (RStudio Team, 2016) and Perl script (version 5.26.2; https://www.perl.org/get.html).

We conducted the LD analysis for a candidate SNP whose allele frequency may have largely changed after Out-of-Africa. The LD among SNPs was estimated using JMP Genomics version 9.0 software (SAS Institute). In addition, we surveyed the LD between a target SNP and adjacent SNPs in JPT (Japanese in Tokyo, Japan) from the 1KGP, employing LDlink web-based tools (https://analysistools.nci.nih.gov/LDlink/; Machiela & Chanock, 2015).

2.4 | Inference of evolutionary events

A sliding window analysis (window size = 100 SNPs and step size = 20 SNPs) was implemented to examine the nucleotide diversity (π) among phased haplotypes from the 1KGP (note that the sequence length of haplotypes is not considered). Tajima's D statistics (Tajima, 1989) was used to test the significance of departure from neutral expectations after removing possible recombination sites estimated. These analyses were performed using DnaSP 5.10.01 software (Librado & Rozas, 2009). The extended haplotype homozygosity (EHH; Sabeti et al., 2002) and integrated haplotype score (iHS; Voight et al., 2006) were used to infer recent positive selection using the R package "rehh 2.0" (Gautier, Klassmann, & Vitalis, 2017; Gautier & Vitalis, 2012). A signal of positive selection in ancestral modern human

TABLE 2 Allele frequencies of disease-associated genes or loci in modern human populations

Disease	RefSNP ID	Position[a]	EAS All	JP-Inabe	JPT[b]	CDX[b]	CHB[b]	CHS[b]	KHV[b]	SAS[b]	EUR[b]	AFR[b]
Hypertension-Hyperuricemia-Dyslipidemia	rs12229654	12:11414461	T: 0.777 (13,319); G: 0.223 (3,829)	T: 0.747 (9,001); G: 0.253 (3,043)	T: 0.837 (174); G: 0.163 (34)	T: 0.957 (178); G: 0.043 (8)	T: 0.835 (172); G: 0.165 (34)	T: 0.724 (152); G: 0.276 (58)	T: 0.869 (172); G: 0.131 (26)	T: 1.000 (978)	T: 1.000 (1,006)	T: 1.000 (1,322)
	rs3782886 *BRAP*	12:12110489	A: 0.719 (12,323); G: 0.281 (4,825)	A: 0.682 (8,220); G: 0.318 (3,824)	A: 0.750 (156); G: 0.250 (52)	A: 0.957 (178); G: 0.043 (8)	A: 0.845 (174); G: 0.155 (32)	A: 0.729 (153); G: 0.271 (57)	A: 0.864 (171); G: 0.136 (27)	A: 1.000 (978)	A: 1.000 (1,006)	A: 1.000 (1,322)
	rs11066015 *ACAD10*	12:12218009	G: 0.735 (12,582); A: 0.265 (4,548)	G: 0.700 (8,425); A: 0.300 (3,619)	G: 0.864 (171); A: 0.136 (27)	G: 0.957 (178); A: 0.043 (8)	G: 0.840 (173); A: 0.160 (33)	G: 0.719 (151); A: 0.281 (59)	G: 0.760 (158); A: 0.240 (50)	G: 1.000 (978)	G: 1.000 (1,006)	G: 1.000 (1,322)
	rs671 *ALDH2*	12:12241766	G: 0.714 (12,353); A: 0.286 (4,960)	G: 0.699 (8,422); A: 0.301 (3,622)	G: 0.760 (158); A: 0.240 (50)	G: 0.957 (178); A: 0.043 (8)	G: 0.840 (173); A: 0.160 (33)	G: 0.729 (153); A: 0.271 (57)	G: 0.864 (171); A: 0.136 (27)	G: 1.000 (978)	G: 1.000 (1,006)	G: 0.998 (1,320); A: 0.002 (2)
	rs2074356 *HECTD4*	12:12645401	C: 0.757 (12,991); T: 0.243 (4,164)	C: 0.722 (8,696); T: 0.278 (3,348)	C: 0.788 (164); T: 0.212 (44)	C: 0.968 (180); T: 0.032 (6)	C: 0.898 (185); T: 0.102 (21)	C: 0.819 (172); T: 0.181 (38)	C: 0.894 (177); T: 0.106 (21)	C: 1.000 (978)	C: 1.000 (1,006)	C: 1.000 (1,322)
	rs11066280 *HECTD4*	12:12817783	T: 0.719 (12,326); A: 0.281 (4,824)	T: 0.687 (8,272); A: 0.313 (3,772)	T: 0.764 (159); A: 0.236 (49)	T: 0.866 (161); A: 0.134 (25)	T: 0.811 (167); A: 0.189 (39)	T: 0.705 (148); A: 0.295 (62)	T: 0.783 (155); A: 0.217 (43)	T: 0.996 (974); A: 0.004 (4)	T: 1.000 (1,006)	T: 0.999 (1,321); A: 0.001 (1)
T2DM	rs78338345 *GGA3*	17:73238.09	C: 0.888 (11,586); G: 0.112 (1,464)	C: 0.886 (10,670); G: 0.114 (1,372)	C: 0.861 (179); G: 0.139 (29)	C: 0.973 (181); G: 0.027 (5)	C: 0.869 (179); G: 0.131 (27)	C: 0.914 (192); G: 0.086 (18)	C: 0.934 (185); G: 0.066 (13)	C: 0.998 (976); G: 0.002 (2)	C: 1.000 (1,006)	C: 1.000 (1,322)
Obesity	rs7656604	4:72547436	G: 0.921 (11,997); A: 0.079 (1,029)	G: 0.916 (11,031); A: 0.084 (1,013)	G: 0.923 (192); A: 0.077 (16)	G: 0.989 (184); A: 0.011 (2)	G: 0.971 (200); A: 0.029 (6)	G: 0.957 (201); A: 0.043 (9)	G: 0.955 (189); A: 0.045 (9)	G: 0.947 (926); A: 0.053 (52)	G: 0.994 (1,000); A: 0.006 (6)	G: 0.200 (264); A: 0.800 (1,058)
Dyslipidemia	rs34902660 *SLC17A3*	6:25851102	C: 0.907 (11,831); A: 0.093 (1,219)	C: 0.900 (10,836); A: 0.100 (1,206)	C: 0.938 (195); A: 0.062 (13)	C: 1.000 (186)	C: 1.000 (206)	C: 1.000 (210)	C: 1.000 (198)	C: 1.000 (978)	C: 1.000 (1,006)	C: 0.984 (1,301); A: 0.016 (21)

Note: Values indicate the allele frequency, with the observed numbers in parentheses. The upper and lower alleles are the major and minor alleles in the Inabe cohort, respectively. JP-Inabe is Japanese in the Inabe cohort; JPT is Japanese in Tokyo, Japan; CDX is Chinese Dai in Xishuangbanna, China; CHB is Han Chinese in Beijing, China; CHS is Southern Han Chinese; KHV is Kinh in Ho Chi Minh City, Vietnam. T2DM, type 2 diabetes mellitus. AFR, Africans; EAS, East Asians; SAS, South Asians; EUR, Europeans.

[a]Position in NCBI build GRCh37.p13.

[b]Allele frequency obtained from the 1000 Genomes Project through the Ensembl genome browser.

populations was surveyed by the Selective Sweep Scan on human versus Neanderthal polymorphisms (Green et al., 2010) in the UCSC Genome Browser.

3 | RESULTS

To explore SNPs whose allele frequencies have remarkably changed after Out-of-Africa, based on information available on the 1KGP database and our genotype data for the Inabe cohort, we firstly examined allele frequencies of SNPs closely located in genomic regions using 272 disease-associated SNPs identified in our previous studies, because it is possible that SNPs densely located are in strong LD. As a result, 52 SNPs were clustered in six genomic regions: more than five SNPs were located within 100 kb genomic intervals (Table S1). In addition, we also surveyed allele frequencies of 28 SNPs that have been newly identified in our longitudinal exome-wide association studies (Table 1). Of the 28 SNPs, 17 were removed from further analyses because their allele frequencies were not largely different among the four ethnic groups, and three were included in the SNP clusters mentioned above. A total of 60 SNPs in 10 genomic regions were used for further analyses. Genome scans to detect positive selection were comprehensively conducted throughout the 10 genomic regions around disease-susceptibility loci, using genotype data from the 1KGP. After candidate SNPs were filtered out by the EHH analyses (Figures S1–S5), potential signatures of positive selection were detected in four genomic regions around nine genetic determinants of T2DM [rs6414624 (C → T, p.H258Y) of EvC ciliary complex subunit 1 (*EVC*) and rs78338345 (C → G, p.E147Q) of golgi associated, gamma adaptin ear containing, ARF binding protein 3 (*GGA3*) (Yasukochi et al., 2019)], obesity [rs7656604 (G → A) at 4q13.3 (Yasukochi et al., 2018d)], dyslipidemia [rs34902660 of solute carrier family 17 member 3 (*SLC17A3*) and rs1042127 of corneodesmosin (*CDSN*) (Yasukochi et al., 2018c)], or hypertension-dyslipidemia-hyperuricemia [rs12229654 (T → G) at 12q24.11, rs3782886 (T → C, p.R241R) of BRCA1 associated protein (*BRAP*), rs11066015 (G → A) of acyl-CoA dehydrogenase family member 10 (*ACAD10*), rs671 (G → A, p.E504K) of aldehyde dehydrogenase 2 family member (*ALDH2*), and rs2074356 (G → A) and rs11066280 (T → A) of HECT domain E3 ubiquitin protein ligase 4 (*HECTD4*) (Yasukochi et al., 2017, 2018a, 2018c)].

3.1 | Susceptibility loci for T2DM

According to information on the 1KGP database (Table 2), the frequency of the minor allele "G" of rs78338345 in *GGA3* in the Inabe cohort was remarkably low outside of East Asia (EAS = 0.079, SAS = 0.002, and EUR and AFR = 0.000). Based on information available on the ExAC Browser, the

MAF was also extremely low in the three non-EAS populations (0.0002–0.0012) while it was relatively high in EAS (0.1016). Analyses implemented in our previous study suggested that the MAF of this SNP has increased across East Asians in recent evolutionary time, although a strong signal of recent positive selection was not detected (Yasukochi et al., 2019).

The LDproxy search in the LDlink web-based application indicated that rs78338345 of *GGA3* and rs28372681 near nucleoporin 85 (*NUP85*) were in significant LD ($r^2 = 0.882$) in JPT (Table S2). However, rs28372681 has not been reported to be associated with T2DM. The LDproxy also showed no SNP in LD with rs6414624 of *EVC* ($r^2 < 0.2$) in JPT (Table S3). These results suggest that the two SNPs identified in our longitudinal exome-wide association studies affect the prevalence of T2DM independently.

We conducted sliding window analyses for nucleotide diversities in a 24.6 kb genomic region around *GGA3* in the four ethnic groups from the 1KGP (Figure 1). The π values in non-AFRs were entirely lower than those in AFRs. In particular, π values of five windows in a 5.2 kb region containing rs78338345 in non-AFRs ($\pi = 0.002$–0.015) were significantly ($p = 5.4 \times 10^{-7}$, Welch's *t* test) lower than those in AFRs ($\pi = 0.027$–0.034). Tajima's *D* test for the 5.2 kb region indicated significant ($p < .05$) negative *D* values in all the four ethnic groups (AFR = −2.33, EAS = −2.04, EUR = −2.52, SAS = −2.49). A negative *D* value can result from selective sweep (recent positive selection), purifying selection, or recent population growth.

We additionally performed EHH analyses in a 2.5 Mb genomic region around rs78338345, using genotype data from the 1KGP. The analyses indicated that EHH values of the derived allele "G" were considerably higher than those of the ancestral allele "C" throughout the region examined in EAS whereas high EHH values of the derived allele were not observed in SAS (Figure 2), suggesting that *GGA3* can be a candidate target of positive selection in EAS.

Using variant call format datasets of 38 chimpanzees, *Pan troglodytes* and *Pan paniscus*, from the GAGP database, we examined genotypes in the site homologous to human rs78338345. No chimpanzee examined had the minor allele "G". In addition, the major allele "C" might be evolutionarily conserved in vertebrates including archaic humans, according to information on the UCSC Genome Browser database (Table S4). These results suggest that the "G" allele may be unique to modern humans, although further research for genetic polymorphisms in each vertebrate species is required.

According to information on the 1KGP database, the major (ancestral) allele frequency of rs6414624 ("C") in *EVC* in the Inabe cohort was high outside Africa (EAS = 0.918, SAS = 0.831, EUR = 0.793, and AFR = 0.502), indicating that the ancestral allele frequencies may have increased after Out-of-Africa. We mainly focused on SNPs whose frequencies of derived allele rapidly increased because it is difficult to detect signals of recent positive selection
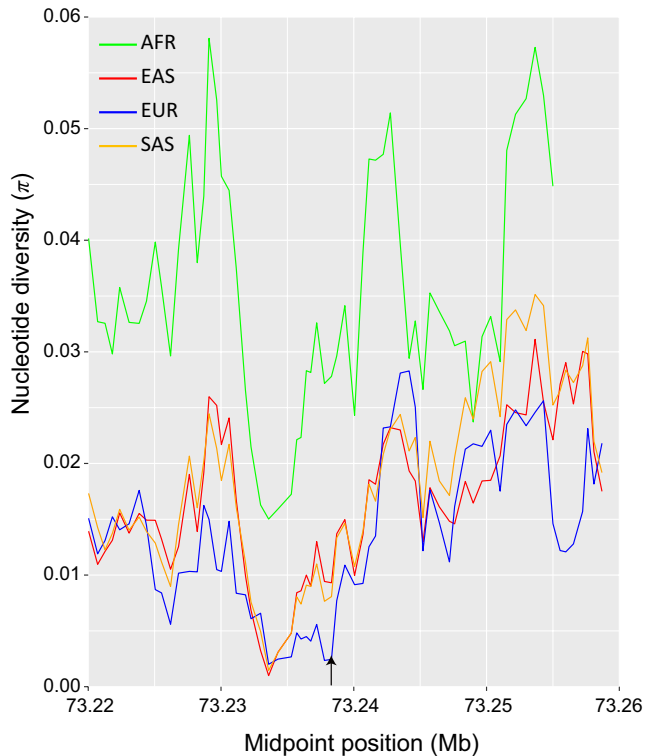
FIGURE 1 Nucleotide diversity (π) in a 40-kb genomic region around *GGA3* within four ethnic groups based on phased haplotype data from the 1000 Genomes Project. Values of π are shown by a sliding window with a window size of 100 SNPs and a step size of 20 SNPs. Green, red, yellow, and blue lines represent π values in AFR, EAS, SAS, and EUR, respectively. The position of SNP associated with type 2 diabetes mellitus (rs78338345) is indicated by an arrow. The horizontal axis represents the chromosomal position (NCBI build GRCh37). AFR, Africans; EAS, East Asians; SAS, South Asians; EUR, Europeans

FIGURE 2 Extended haplotype homozygosity (EHH) plots in a 2.5-Mb genomic region around *GGA3* within two ethnic groups. A blue line represents the EHH value of the ancestral allele "C". A red line represents the EHH value of the derived allele "G". A vertical dashed line indicates the position of the core SNP (rs78338345). The horizontal axis represents the chromosomal position (NCBI build GRCh37). The vertical axis represents the EHH value. EAS, East Asians; SAS, South Asians

targeting the ancestral allele, compared with the derived allele. However, we preliminary conducted the iHS analysis, which examines the ratio of EHH decay between the ancestral and derived alleles, for rs6414624 in *EVC*. The iHS value around rs6414624 was negative (iHS = −4.22, $p = 2.4 \times 10^{-5}$), suggesting a long-range haplotype carrying the derived allele. In addition, EHH values for ancestral and derived core alleles were not largely different within the four ethnic groups (Figures S6). At least, therefore, recent positive selection is unlikely to target the ancestral allele in *EVC* although further analyses such as genome scans for detecting signals of positive selection on standing genetic variation are required.

3.2 | Susceptibility loci for obesity

The frequency of major allele "G" at rs7656604 located between solute carrier family 4 member 4 (*SLC4A4*) and vitamin D binding protein (*GC*) genes in the Inabe cohort (0.916) was remarkably low in AFRs (0.200; Table 2), suggesting
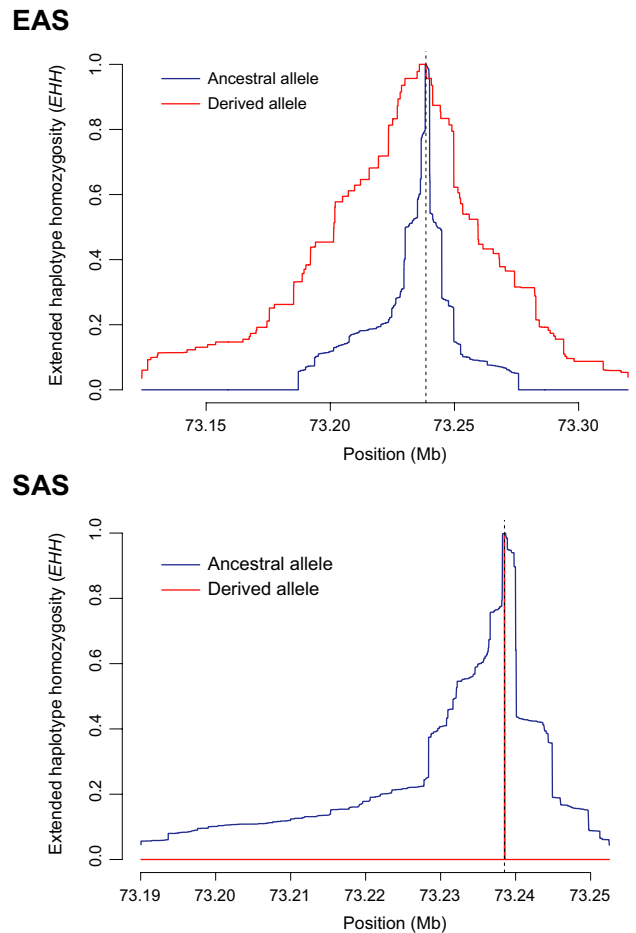
recent positive selection at this locus after Out-of-Africa. To estimate positive selection, we conducted sliding window and LD analyses using SNP data in a ~100 kb genomic region at 4q13.3 from the 1KGP. The sliding window analysis indicated that π values among phased haplotypes in a 2 kb region (four windows) around rs7656604 were significantly ($p = .0005$, Welch's *t* test) low in non-AFRs ($\pi = 0.001$–0.012) compared with AFR ($\pi = 0.033$–0.046; Figure 3). This suggests that the observed nucleotide diversities were greatly reduced by positive selection (selective sweep) in non-AFRs. However, LDs among SNPs in the genomic region appeared to be disrupted in EAS (Figure S7). Additionally, the EHH test did not show a wide range of high EHH values (EHH >0.9 across ~7.5 kb in EAS at maximum) for the major (derived) allele "G" (Figure S8). We further conducted Tajima's *D* tests for the 2 kb region nearby rs7656604 to examine the significance of deviation
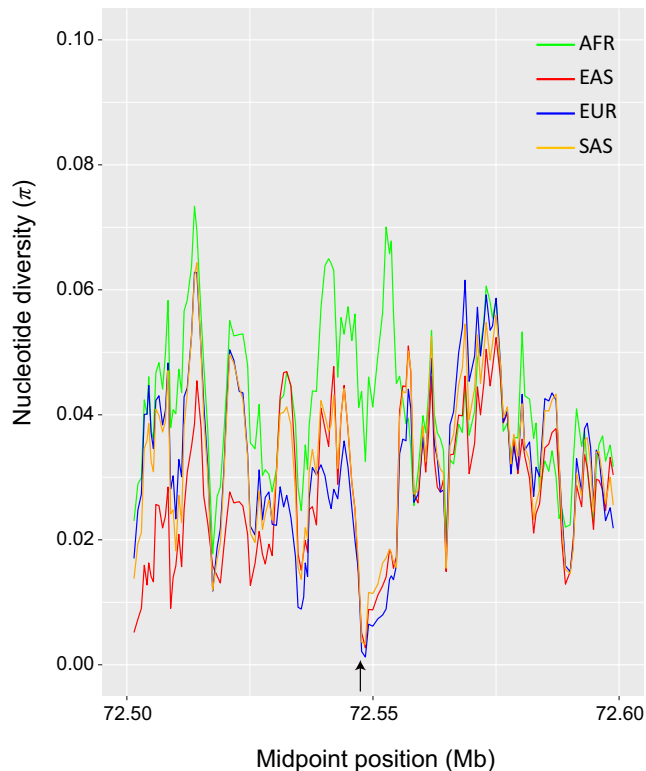
FIGURE 3 Nucleotide diversity (π) in a 100 kb genomic region around *SLC4A4* and *GC* at 4q13.3 within four ethnic groups based on phased haplotype data from the 1000 Genomes Project. Values of π are shown by a sliding window with a window size of 100 SNPs and a step size of 20 SNPs. Green, red, yellow, and blue lines represent π values in AFR, EAS, SAS, and EUR, respectively. The position of the SNP associated with body mass index in Japanese men (rs7656604) is indicated by an arrow. The horizontal axis represents the chromosomal position (NCBI build GRCh37). AFR, Africans; EAS, East Asians; SAS, South Asians; EUR, Europeans

from neutral expectations. The results indicate significant ($p < 0.05$) negative D values in all the four ethnic groups (AFR = −1.91, EAS = −2.08, EUR = −1.76, SAS = −2.03).

As mentioned above, rs7656604 is located near *GC*, and, in a previous study (Mozzi et al., 2014), a signal of recent positive selection was detected in this gene within JPT/CHB (Han Chinese in Beijing, China), CEU (Utah Residents with Northern and Western European Ancestry from the CEPH collection), and YRI (Yoruba in Ibadan, Nigeria). According to LDpair, a LDlink web-based application, rs7656604 and five positively selected SNPs in JPT/CHB (rs2298850, rs11723621, rs62302167, rs1155563, and rs61743452) identified by the previous study were not in LD ($r^2 = 0.014$–0.036), suggesting that the allele frequency of rs7656604 was not affected by the selected SNPs.

According to information available on the UCSC Genome Browser (Table S5) and GAGP databases, the minor allele "A" of rs7656604 in non-AFRs appears to be evolutionarily conserved in primates including archaic humans. Thus,

the majority of the "G" allele is likely human-specific. The Selective Sweep Scan (Neanderthal vs. human polymorphisms) in the UCSC Genome Browser implied that rs7656604 has undergone positive selection in ancestral modern human populations since Out-of-Africa.

## 3.3 | Susceptibility loci for dyslipidemia

The LDproxy analyses indicated that rs34902660 of *SLC17A3* and rs1042127 of *CDSN* were in significant LD ($r^2 > 0.8$) with 281 and 87 neighbor SNPs, respectively (Tables S6–S7), although the neighbor SNPs are not associated with dyslipidemia-related phenotypes, according to the GRASP database (Yasukochi et al., 2018c). This suggests that these SNPs can be independently associated with the prevalence of dyslipidemia. Based on allele frequency data from the 1KGP, the minor allele "A" at rs34902660 of *SLC17A3* in the Inabe cohort was observed in JPT and AFR only (Table 2). According to information available on the ExAC Browser, the minor allele was present at an extremely low frequency in SAS and EUR (MAF < 0.0007). We conducted the EHH analysis in a 100-kb genomic region containing *SLC17A3* using genotype data from the 1KGP. The analyses displayed high EHH values (>0.8) of the minor (derived) allele "A" across 90 kb and 61 kb genomic regions in JPT and YRI, respectively (Figure 4). As the derived allele was not observed in East Asia outside Japan, no high EHH values of the derived allele were observed in CHB based on SNP data from the 1KGP. The extent of high EHH values of the major (ancestral) allele "C" (<2 kb) was remarkably small in JPT and YRI compared with the EHH values of the derived allele (Figure 4). However, Tajima's D test for the 20 kb genomic region containing rs34902660 in JPT did not show a significant D value ($D = −1.25$, $p > .10$).

According to information available on the GAGP database, no derived "A" allele was observed at the homologous target SNP in 38 chimpanzees. In addition, Multiz Alignments of 100 Vertebrates in the UCSC database indicate that all primates including archaic humans possess the ancestral allele "C" (Table S8). These results suggest that the "C" allele may have been conserved in primates, although further research for genetic polymorphisms in each primate species is required.

3.4 | | Common susceptibility loci for several complex disorders

Six SNPs (rs12229654 at 12q24.11, rs3782886 of *BRAP*, rs11066015 of *ACAD10*, rs671 of *ALDH2*, and rs2074356 and rs11066280 of *HECTD4*) associated with several complex disorders in the Inabe cohort (Yasukochi et al., 2017, 2018a, 2018c) are closely located in the chromosomal region 12q24.11–q24.13. The LDproxy analyses indicated that rs671 in *ALDH2* and five SNPs, including three disease-associated

FIGURE 4 Extended haplotype homozygosity (EHH) plots in a 100-kb genomic region around *SLC17A3* within four populations. A blue line represents the EHH value of the ancestral allele "C". A red line represents the EHH value of the derived allele "A". A vertical dashed line indicates the position of the core SNP (rs34902660). The horizontal axis represents the chromosomal position (NCBI build GRCh37). The vertical axis represents the EHH value. JPT, Japanese in Tokyo, Japan; CHB, Han Chinese in Beijing, China; CEU, Utah residents with Northern and Western European Ancestry from the CEPH collection; YRI, Yoruba in Ibadan, Nigeria
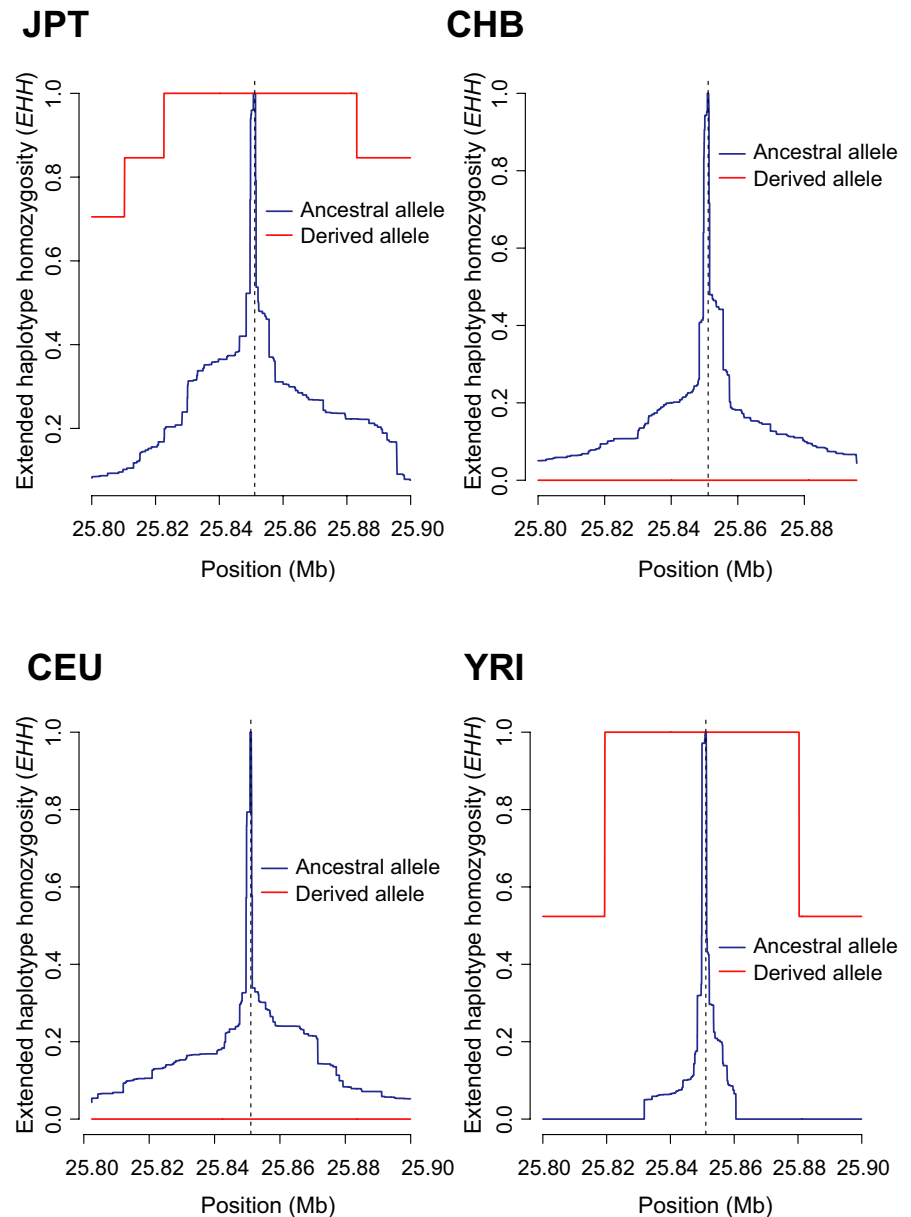
SNPs detected by our longitudinal exome-wide association studies, were in significant LD ($r^2 > 0.8$) in JPT from the 1KGP (Table S9). Recent positive selection is likely to have acted on the genomic region containing rs671, and the minor alleles have expanded throughout East Asia in recent evolutionary time (Koganebuchi et al., 2017; Luo et al., 2009; Oota et al., 2004; Yasukochi et al., 2017).

In the present study, the sliding window analyses using SNP data in a ~2.5 Mb genomic region at 12q24.11–q24.13 from the 1KGP indicated that π values among phased haplotypes in a ~40 kb region (eight windows) around rs671 were not significantly ($p = .299$, Welch's t test) different between EAS (mean $\pi = 0.028 \pm 0.006$) and AFR (mean $\pi = 0.026 \pm 0.005$; Figure S9). However, Tajima's D test for the 40 kb region indicated a significant negative D value in EAS ($D = -2.79$, $p < .001$) after possible recombination sites were removed. We also

conducted the EHH analysis for core SNP rs671 to show a signal of positive selection using genotype data across ~2.5 Mb genomic region at 12q24.11–q24.13 in EAS and SAS from the 1KGP (Figure S10). In EAS, EHH values of the rs671-derived allele "A" were considerably higher than those of the ancestral allele "G" across the genomic region examined. In contrast to EAS, no high EHH value of the derived allele was observed in SAS. We thus confirmed the presence of positive selection operating on the genomic region around *ALDH2* in EAS.

4 | | DISCUSSION

4.1 | Susceptibility loci for T2DM

Individuals with East Asian ancestry are more susceptible to T2DM than those with European ancestry despite a

similar BMI (Ma & Chan, 2013). It has been reported that the prevalence of T2DM in Japanese Americans who have western dietary habits was higher than that in European Americans because of reduced insulin secretory capacity in Japanese or East Asian populations (Yabe, Seino, Fukushima, & Seino, 2015). Differences in genetic background and dietary habits may underlie the different insulin secretory capacity between individuals with East Asian and European ancestries.

The GGA3 is involved in the intracellular trafficking of insulin-like growth factor 2 receptor (IGF2R) protein between the trans-Golgi network and early endosomes (Scott, Fei, Thomas, Medigeshi, & Thomas, 2006). This protein also interacts with ADP ribosylation factor 6 (ARF6; Parachoniak, Luo, Abella, Keen, & Park, 2011) that regulates insulin secretion in the pancreatic β cell (Jayaram, Syed, Kyathanahalli, Rhodes, & Kowluru, 2011; Lawrence & Birnbaum, 2003). Given that GGA3 is related to the insulin secretion pathway with IGF2R or ARF6, the EAS-specific nonsynonymous substitution in rs78338345 might be involved in the insulin secretory capacity. According to the PhyloP score, the ancestral allele "C" at rs78338345 appears to have been evolutionarily conserved in vertebrates (score = 6.08), consistent with the result of Multiz Alignments of 100 Vertebrates. PolyPhen prediction suggested deleterious effects on the protein function (PolyPhen score = 1.00) although two other protein function prediction tools suggested that the amino acid replacement of rs78338345 (p.E147Q) in GGA3 might not disrupt protein function (SIFT score = 0.134 and CADD scaled C-score = 24.8). However, the derived allele "G" was identified as a protective allele against T2DM (Yasukochi et al., 2019). Functional analyses are required to elucidate relationships between the candidate SNP in GGA3 and insulin secretory ability.

The rs78338345 of GGA3 was identified as a novel genetic determinant of T2DM in our previous study (Yasukochi et al., 2019), and the frequency of the derived allele "G" was higher in EAS than in other ethnic groups. In the present study, our estimates suggest that this allele may have expanded throughout East Asia. As the GGA3 protein may be involved in immune responses mediated by target proteins such as clathrin (Benedicto et al., 2015) and ARF6 (Wu & Kuo, 2012), rs78338345 might have spread across EAS to cope with infectious diseases although the biological significance remains unclear.

Because Tajima's D tests suggested that allele frequencies of SNPs around GGA3 are skewed in the populations examined, we cannot exclude the possibility that demographic events affected the allele frequencies within EAS. In addition, it is possible that the haplotype diversity of core allele that emerged by the recent new mutation can be low, resulting in a high EHH without positive selection due to few recombination events. According to ExAC

Browser, the derived allele "G" of rs78338345 is also observed outside of East Asia although the frequencies are very low. One may hypothesize that the derived allele has been maintained at an extremely low frequency during the course of modern human evolution, and the frequency may have largely increased in East Asia in very recent evolutionary time. It will be intriguing to elucidate the evolutionary forces shaping the genetic variation in the genomic region around GGA3. Further analyses based on population genomics are required to verify the evolutionary history of GGA3.

4.2 | Susceptibility loci for obesity

The frequency of the derived allele "G" (major allele in non-AFRs) at rs7656604 near GC appears to have increased considerably outside of Africa. Our previous study indicated that this allele is protective against increased BMI (Yasukochi et al., 2018d). The sliding window analysis indicated a pronounced decrease in nucleotide diversity across a ~2 kb region around rs7656604 in non-AFRs, implying recent positive selection operating on this locus after Out-of-Africa. A signal of recent positive selection on five SNPs in GC of EAS was also detected in a previous study (Mozzi et al., 2014). It is conceivable that positive selection has independently operated on rs7656604 and the positively selected sites in GC because no pairs of the different sets of SNPs were in significant LD ($r^2 < 0.04$). If positive selection has actually operated on rs7656604, one may hypothesize that frequencies of several haplotypes with the derived allele "G" lineage in non-AFRs have immediately increased after Out-of-Africa. This may be caused by selective sweep on standing genetic variation. Further analysis for population genomics is required to accurately elucidate the evolutionary history of the genomic region around rs7656604.

The GC protein is a main carrier for vitamin D and its metabolites. Vitamin D deficiency may be related to the development of adiposity or increased BMI (Snijder et al., 2005). The rs7656604 nearby GC might thus be related to the increased BMI through the effect on vitamin D synthesis. Cutaneous vitamin D production is affected by exposure to ultraviolet (UV) radiation and skin pigments (Jablonski & Chaplin, 2017), and lighter skin color in non-AFRs may have evolved to optimize vitamin D production for shorter exposure to daily sunlight (Wilde et al., 2014). Since frequencies of the derived allele in rs7656604 were similar among non-AFRs, irrespective of geographical distributions based on latitude, the derived allele "G" lineage might have been targeted by natural selection to fine-tune vitamin D homeostasis under UV and food conditions outside Africa. Alternatively, the derived allele might have expanded throughout ancestral modern humans to acquire

resistance to novel pathogens outside Africa, because vitamin D plays an important role in the modulation of the immune system (Calton, Keane, Newsholme, & Soares, 2015; Kamen & Tangpricha, 2010). Further research for the correlation of GC expression profiles with rs7656604 is required to examine whether rs7656604 is involved in the regulation of *GC* expression levels.

4.3 | Susceptibility loci for dyslipidemia

The rs34902660 "A" allele in *SLC17A3* at 6p22.2 was only observed in JPT and AFR from the 1KGP. This suggests that the derived allele "A" may have increased in JPT in very recent evolutionary time. However, the negative Tajima's *D* value in JPT was not statistically significant. It may be difficult to detect the weak signal of recent positive selection by the frequency spectrum neutrality test because the selective sweep could be occurring in JPT in very recent evolutionary time. According to the ExAC Browser database, the derived allele "A" is observed at an extremely low frequency in SAS and EUR populations. This suggests that the derived allele has existed in common ancestral populations of modern humans. One may hypothesize that the derived allele frequency has recently increased in EAS populations after the frequency largely reduced in non-AFR ancestral populations after Out-of-Africa. To validate our results, further scans for the footprint of positive selection are required.

The SLC17A3 protein is involved in efflux transport of intracellular urate and organic anions from the blood into the renal tubule cells. This protein plays a key role in urate secretion in the renal tubular cell (Jutabha et al., 2010). It is possible that an increase in xanthine oxidoreductase activity is related to elevated serum concentration of uric acid, resulting in the development of dyslipidemia (Maiuolo, Oppedisano, Gratteri, Muscoli, & Mollace, 2016). According to the SIFT prediction, rs34902660 of *SLC17A3* might cause a dysfunctional transcript [p.G239V (SIFT score = 0.018)]. Therefore, the association of rs34902660 of *SLC17A3* with dyslipidemia might be attributable to the effect of this gene on abnormal lipid profiles through impaired urate secretion. The serum concentration of uric acid is associated with the production of pro-inflammatory cytokines (Crişan et al., 2016; Spaetgens et al., 2017). Given that the accelerated production of cytokines through the increased serum uric acid level is important for host defense against viral, bacterial, and parasite infection, the derived allele of rs34902660 might confer resistance to endemic or epidemic diseases in Japan. However, some prediction tools did not support the result of SIFT prediction (PolyPhen = 0.218, CADD scaled C-score = 14.8, and PhyloP = 0.06). Further functional analyses are required to clarify the effect of rs34902660 on the protein function.

4.4 | Common susceptibility loci for several complex disorders

Previous studies indicated that derived alleles of several SNPs at 12q24.11–q24.13 have expanded throughout East Asia in recent evolutionary time (Kato et al., 2011; Yasukochi et al., 2017). Our association studies suggested that derived alleles of six EAS-specific SNPs are protective against hypertension (Yasukochi et al., 2017) and hyperuricemia (Yasukochi et al., 2018a) but increase susceptibility to dyslipidemia (Yasukochi et al., 2018c). This discrepancy might be due to differences in the effects of these SNPs on functional pathways involved in disease-related phenotypes.

Of the six SNPs, the major genetic factor for the prevalence of complex disorders may be rs671 of *ALDH2*, because this SNP only results in an amino acid change at position 504. The amino acid change (p.E504K) is well known to cause defective enzyme activity (Crabb, Edenberg, Bosron, & Li, 1989; Hsu, Bendel, & Yoshida, 1987). In fact, the function prediction tools suggest that the missense variant has an effect on the protein function (PolyPhen score = 0.964, SIFT = 0.019, CADD scaled C-score = 34.0, and PhyloP = 9.71). It is possible that the derived allele of rs671 may have arisen ~8,000 years ago when large-scale rice cultivation started in southern China (Koganebuchi et al., 2017). Previous studies suggested that positive selection has acted on *ALDH2* in EAS because of acquired resistance against parasite infection related to large-scale rice cultivation (Koganebuchi et al., 2017; Luo et al., 2009; Oota et al., 2004). ALDH2 may be involved in the metabolism of 4-hydroxy-2-nonenal, a by-product of lipid peroxidation (Breitzig, Bhimineni, Lockey, & Kolliputi, 2016). Therefore, it is possible that the decreased ALDH2 enzyme activity may be associated with the prevalence of complex disorders through accelerated aging phenotypes because of the increased susceptibility to oxidative stress. Although the functional relevance of rs671 to the development of complex disorders remains unclear, the drastic functional change in the ALDH2 enzyme may have additional effects on functional pathways related to several complex disorders.

4.5 | Study limitations

It is possible that population demographic events such as population growth and bottlenecks as well as recombination rate yield genetic patterns that are similar to those generated by natural selection (Bamshad & Wooding, 2003; Nielsen, 2005). A haplotype diversity of core allele arisen from the recent new mutation can also result in high EHH values without positive selection. Therefore, the fluctuation in allele frequencies, LD, or genetic diversities in the genomic regions examined should be carefully interpreted. Further analyses based on population

genomics are required to verify the evolutionary history of these disease-susceptibility loci.

5 | CONCLUSION

Our findings indicate that derived allele frequencies of nine disease-associated SNPs (rs78338345 of *GGA3*, rs7656604 at 4q13.3, rs34902660 of *SLC17A3*, and six SNPs closely located at 12q24.1) in Japanese populations largely increased after Out-of-Africa.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Y. Yasukochi contributed to design the study, analysis and interpretation of the data, and drafting of the manuscript. J. Sakuma and I. Takeuchi contributed to analysis and interpretation of the data as well as revision of the manuscript. K. Kato, M. Oguri, T. Fujimaki, and H. Horibe each contributed to acquisition of the data and revision of the manuscript. Y. Yamada contributed to acquisition, analysis, interpretation of the data, and revision of the manuscript.

ORCID

Yoshiki Yasukochi [iD] https://orcid.org/0000-0001-5997-959X

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Aguilar-Salinas, C. A., Tusie-Luna, T., & Pajukanta, P. (2014). Genetic and environmental determinants of the susceptibility of Amerindian derived populations for having hypertriglyceridemia. *Metabolism*, *63*(7), 887–894. https://doi.org/10.1016/j.metabol.2014.03.012

Akinyemi, R., Ovbiagele, B., Akpalu, A., Jenkins, C., Sagoe, K., Owolabi, L., … Owolabi, M. (2015). Stroke genomics in people of African ancestry: Charting new paths. *Cardiovascular Journal of Africa*, *26*(2), S39–S49. https://doi.org/10.5830/CVJA-2015-039

Asselbergs, F. W., Guo, Y., van Iperen, E. P. A., Sivapalaratnam, S., Tragante, V., Lanktree, M. B., … Drenos, F. (2012). Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *American Journal of Human Genetics*, *91*(5), 823–838. https://doi.org/10.1016/j.ajhg.2012.08.032

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics*, *4*(2), 99–111. https://doi.org/10.1038/nrg999

Barber, M. J., Mangravite, L. M., Hyde, C. L., Chasman, D. I., Smith, J. D., McCarty, C. A., … Krauss, R. M. (2010). Genome-wide association of lipid-lowering response to statins in combined study populations. *PLoS ONE*, *5*(3), e9763. https://doi.org/10.1371/journal.pone.0009763

Benedicto, I., Gondar, V., Molina-Jiménez, F., García-Buey, L., López-Cabrera, M., Gastaminza, P., & Majano, P. L. (2015). Clathrin mediates infectious hepatitis C virus particle egress. *Journal of Virology*, *89*(8), 4180–4190. https://doi.org/10.1128/JVI.03620-14

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., … Shriver, M. D. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genetics*, *6*(9), e1001116. https://doi.org/10.1371/journal.pgen.1001116

Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F. A., Roskin, K. M., … Miller, W. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, *14*(4), 708–715. https://doi.org/10.1101/gr.1933104

Böger, C. A., Gorski, M., Li, M., Hoffmann, M. M., Huang, C., Yang, Q., … Kao, W. H. (2011). Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. *PLoS Genetics*, *7*(9), e1002292. https://doi.org/10.1371/journal.pgen.1002292

Breitzig, M., Bhimineni, C., Lockey, R., & Kolliputi, N. (2016). 4-Hydroxy-2-nonenal: A critical target in oxidative stress? *American Journal of Physiology-Cell Physiology*, *311*(4), C537–C543. https://doi.org/10.1152/ajpcell.00101.2016

Calton, E. K., Keane, K. N., Newsholme, P., & Soares, M. J. (2015). The impact of Vitamin D levels on inflammatory status: A systematic review of immune cell studies. *PLoS ONE*, *10*(11), e0141770. https://doi.org/10.1371/journal.pone.0141770

Chambers, J. C., Zhang, W., Lord, G. M., van der Harst, P., Lawlor, D. A., Sehmi, J. S., … Kooner, J. S. (2010). Genetic loci influencing kidney function and chronic kidney disease. *Nature Genetics*, *42*(5), 373–375. https://doi.org/10.1038/ng.566

Chittoor, G., Kent, J. W., Almeida, M., Puppala, S., Farook, V. S., Cole, S. A., … Voruganti, V. S. (2016). GWAS and transcriptional analysis prioritize *ITPR1* and *CNTN4* for a serum uric acid 3p26 QTL in Mexican Americans. *BMC Genomics*, *17*(1), 276. https://doi.org/10.1186/s12864-016-2594-5

Cho, Y. S., Chen, C.-H., Hu, C., Long, J., Hee Ong, R. T., Sim, X., … Seielstad, M. (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nature Genetics*, *44*(1), 67–72. https://doi.org/10.1038/ng.1019

Crabb, D. W., Edenberg, H. J., Bosron, W. F., & Li, T. K. (1989). Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive *ALDH2*[2] allele is dominant. *The Journal of Clinical Investigation*, *83*(1), 314–316. https://doi.org/10.1172/JCI113875

Crawford, J. E., Amaru, R., Song, J., Julian, C. G., Racimo, F., Cheng, J. Y., … Nielsen, R. (2017). Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *American Journal of Human Genetics*, *101*(5), 752–767. https://doi.org/10.1016/j.ajhg.2017.09.023

Crişan, T. O., Cleophas, M. C. P., Oosting, M., Lemmers, H., Toenhake-Dijkstra, H., Netea, M. G., … Joosten, L. A. B. (2016). Soluble uric acid primes TLR-induced proinflammatory cytokine production by human primary cells via inhibition of IL-1Ra. *Annals of the Rheumatic Diseases*, *75*(4), 755–762. https://doi.org/10.1136/annrheumdis-2014-206564

Dorajoo, R., Ong, R. T. H., Sim, X., Wang, L., Liu, W., Tai, E. S., … Saw, S. M. (2017). The contribution of recently identified adult BMI risk loci to paediatric obesity in a Singaporean Chinese childhood dataset. *Pediatric Obesity*, *12*(6), e46–e50. https://doi.org/10.1111/ijpo.12175

International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret, G. B., Munroe, P. B., Rice, K. M., Bochud, M., Johnson, A. D., Chasman, D. I., … Johnson, T. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, *478*(7367), 103–109. https://doi.org/10.1038/nature10405

Fall, T., & Ingelsson, E. (2014). Genome-wide association studies of obesity and metabolic syndrome. *Molecular and Cellular Endocrinology*, *382*(1), 740–757. https://doi.org/10.1016/j.mce.2012.08.018

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: A reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17*(1), 78–90. https://doi.org/10.1111/1755-0998.12634

Gautier, M., & Vitalis, R. (2012). Rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, *28*(8), 1176–1177. https://doi.org/10.1093/bioinformatics/bts115

Golbus, J. R., Stitziel, N. O., Zhao, W., Xue, C., Farrall, M., Mcpherson, R., … Reilly, M. P. (2016). Common and rare genetic variation in *CCR2*, *CCR5*, or *CX3CR1* and risk of atherosclerotic coronary heart disease and glucometabolic traits. *Circulation: Cardiovascular Genetics*, *9*(3), 250–258. https://doi.org/10.1161/CIRCGENETICS.115.001374

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1034. https://doi.org/10.1038/ng.937

Groucutt, H. S., Petraglia, M. D., Bailey, G., Scerri, E. M. L., Parton, A., Clark-Balzan, L., … Scally, A. (2015). Rethinking the dispersal of *Homo sapiens* out of Africa. *Evolutionary Anthropology*, *24*(4), 149–164. https://doi.org/10.1002/evan.21455

Grove, M. L., Yu, B., Cochran, B. J., Haritunians, T., Bis, J. C., Taylor, K. D., … Boerwinkle, E. (2013). Best practices and joint calling of the HumanExome BeadChip: The CHARGE Consortium. *PLoS ONE*, *8*(7), e68095. https://doi.org/10.1371/journal.pone.0068095

Hara, K., Fujita, H., Johnson, T. A., Yamauchi, T., Yasuda, K., Horikoshi, M., … Kadowaki, T. (2014). Genome-wide association study identifies three novel loci for type 2 diabetes. *Human Molecular Genetics*, *23*(1), 239–246. https://doi.org/10.1093/hmg/ddt399

Hershkovitz, I., Weber, G. W., Quam, R., Duval, M., Grün, R., Kinsley, L., … Weinstein-Evron, M. (2018). The earliest modern humans outside Africa. *Science*, *359*(6374), 456–459. https://doi.org/10.1126/science.aap8369

Hsu, L. C., Bendel, R. E., & Yoshida, A. (1987). Direct detection of usual and atypical alleles on the human aldehyde dehydrogenase-2 (ALDH$_2$) locus. *American Journal of Human Genetics*, *41*(6), 996–1001.

Imamura, M., & Maeda, S. (2011). Genetics of type 2 diabetes: The GWAS era and future perspectives. *Endocrine Journal*, *58*(9), 723–739. https://doi.org/10.1507/endocrj.EJ11-0113

Imamura, M., Takahashi, A., Yamauchi, T., Hara, K., Yasuda, K., Grarup, N., … Kadowaki, T. (2016). Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nature Communications*, *7*, 10531. https://doi.org/10.1038/ncomms10531

Jablonski, N. G., & Chaplin, G. (2017). The colours of humanity: The evolution of pigmentation in the human lineage. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1724), 20160349. https://doi.org/10.1098/rstb.2016.0349

Jayaram, B., Syed, I., Kyathanahalli, C. N., Rhodes, C. J., & Kowluru, A. (2011). Arf nucleotide binding site opener [ARNO] promotes sequential activation of Arf6, Cdc42 and Rac1 and insulin secretion in INS 832/13 β-cells and rat islets. *Biochemical Pharmacology*, *81*(8), 1016–1027. https://doi.org/10.1016/j.bcp.2011.01.006

Jia, C., Kong, X., Koltes, J. E., Gou, X., Yang, S., Yan, D., … Wei, Z. (2016). Gene co-expression network analysis unraveling transcriptional regulation of high-altitude adaptation of Tibetan pig. *PLoS ONE*, *11*(12), e0168161. https://doi.org/10.1371/journal.pone.0168161

Jutabha, P., Anzai, N., Kitamura, K., Taniguchi, A., Kaneko, S., Yan, K., … Sakuraia, H. (2010). Human sodium phosphate transporter 4 (hNPT4/SLC17A3) as a common renal secretory pathway for drugs and urate. *Journal of Biological Chemistry*, *285*(45), 35123–35132. https://doi.org/10.1074/jbc.M110.121301

Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., … Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genetics*, *42*(3), 210–215. https://doi.org/10.1038/ng.531

Kamen, D. L., & Tangpricha, V. (2010). Vitamin D and molecular actions on the immune system: Modulation of innate and autoimmunity. *Journal of Molecular Medicine*, *88*(5), 441–450. https://doi.org/10.1007/s00109-010-0590-9

Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., … Cupples, L. A. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*, *41*(1), 56–65. https://doi.org/10.1038/ng.291

Kato, N., Takeuchi, F., Tabara, Y., Kelly, T. N., Go, M. J., Sim, X., … He, J. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature Genetics*, *43*(6), 531–538. https://doi.org/10.1038/ng.834

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, A. D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Koganebuchi, K., Haneji, K., Toma, T., Joh, K., Soejima, H., Fujimoto, K., … Oota, H. (2017). The allele frequency of ALDH2*Glu504Lys and ADH1B*Arg47His for the Ryukyu islanders and their history

of expansion among East Asians. *American Journal of Human Biology*, 29(2), 1–14. https://doi.org/10.1002/ajhb.22933

Kooner, J. S., Chambers, J. C., Aguilar-Salinas, C. A., Hinds, D. A., Hyde, C. L., Warnes, G. R., … Thompson, J. F. (2008). Genome-wide scan identifies variation in *MLXIPL* associated with plasma triglycerides. *Nature Genetics*, 40(2), 149–151. https://doi.org/10.1038/ng.2007.61

Köttgen, A., Albrecht, E., Teumer, A., Vitart, V., Krumsiek, J., Hundertmark, C., … Gieger, C. (2013). Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nature Genetics*, 45(2), 145–154. https://doi.org/10.1038/ng.2500

Kraja, A. T., Vaidya, D., Pankow, J. S., Goodarzi, M. O., Assimes, T. L., Kullo, I. J., … Borecki, I. B. (2011). A bivariate genome-wide approach to metabolic syndrome: STAMPEED Consortium. *Diabetes*, 60(4), 1329–1339. https://doi.org/10.2337/db10-1011

Kubo, M., Hata, J., Ninomiya, T., Matsuda, K., Yonemoto, K., Nakano, T., … Kiyohara, Y. (2007). A nonsynonymous SNP in *PRKCH* (protein kinase C η) increases the risk of cerebral infarction. *Nature Genetics*, 39(2), 212–217. https://doi.org/10.1038/ng1945

Kurano, M., Tsukamoto, K., Kamitsuji, S., Kamatani, N., Hara, M., Ishikawa, T., … Teramoto, T. (2016). Genome-wide association study of serum lipids confirms previously reported associations as well as new associations of common SNPs within *PCSK7* gene with triglyceride. *Journal of Human Genetics*, 61(5), 427–433. https://doi.org/10.1038/jhg.2015.170

Lange, L. A., Hu, Y., Zhang, H. E., Xue, C., Schmidt, E. M., Tang, Z.-Z., … Sturcke, A. (2014). Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *American Journal of Human Genetics*, 94(2), 233–245. https://doi.org/10.1016/j.ajhg.2014.01.010

Lawrence, J. T. R., & Birnbaum, M. J. (2003). ADP-ribosylation factor 6 regulates insulin secretion through plasma membrane phosphatidylinositol 4,5-bisphosphate. *Proceedings of the National Academy of Sciences*, 100(23), 13320–13325. https://doi.org/10.1073/pnas.2232129100

Leslie, R., O'Donnell, C. J., & Johnson, A. D. (2014). GRASP: Analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics*, 30(12), 185–194. https://doi.org/10.1093/bioinformatics/btu273

Levy, D., Ehret, G. B., Rice, K., Verwoert, G. C., Launer, L. J., Dehghan, A., … van Duijn, C. M. (2009). Genome-wide association study of blood pressure and hypertension. *Nature Genetics*, 41(6), 677–687. https://doi.org/10.1038/ng.384

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. https://doi.org/10.1038/nature10231

Li, J., Glessner, J. T., Zhang, H., Hou, C., Wei, Z., Bradfield, J. P., … Sleiman, P. M. A. (2013). GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. *Human Molecular Genetics*, 22(7), 1457–1464. https://doi.org/10.1093/hmg/dds534

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Liu, X., Hu, C., Bao, M., Li, J., Liu, X., Tan, X., … Cai, J. (2016). Genome wide association study identifies *L3MBTL4* as a novel susceptibility gene for hypertension. *Scientific Reports*, 6, 30811. https://doi.org/10.1038/srep30811

Lo, K. S., Wilson, J. G., Lange, L. A., Folsom, A. R., Galarneau, G., Ganesh, S. K., … Lettre, G. (2011). Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Human Genetics*, 129(3), 307–317. https://doi.org/10.1007/s00439-010-0925-1

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., … Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538), 197–206. https://doi.org/10.1038/nature14177

López, S., van Dorp, L., & Hellenthal, G. (2015). Human dispersal out of Africa: A lasting debate. *Evolutionary Bioinformatics Online*, 11(S2), 57–68. https://doi.org/10.4137/EBO.S33489

Lu, X., Wang, L., Chen, S., He, L., Yang, X., Shi, Y., … Gu, D. (2012). Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nature Genetics*, 44(8), 890–894. https://doi.org/10.1038/ng.2337

Luo, H. R., Wu, G. S., Pakstis, A. J., Tong, L., Oota, H., Kidd, K. K., & Zhang, Y. P. (2009). Origin and dispersal of atypical aldehyde dehydrogenase *ALDH2*487Lys*. *Gene*, 435(1–2), 96–103. https://doi.org/10.1016/j.gene.2008.12.021

Ma, R. C. W., & Chan, J. C. N. (2013). Type 2 diabetes in East Asians: Similarities and differences with populations in Europe and the United States. *Annals of the New York Academy of Sciences*, 1281(1), 64–91. https://doi.org/10.1111/nyas.12098

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., … Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45(D1), D896–D901. https://doi.org/10.1093/nar/gkw1133

Machiela, M. J., & Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, 31(21), 3555–3557. https://doi.org/10.1093/bioinformatics/btv402

Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., … Morris, A. P. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics*, 46(3), 234–244. https://doi.org/10.1038/ng.2897

Maiuolo, J., Oppedisano, F., Gratteri, S., Muscoli, C., & Mollace, V. (2016). Regulation of uric acid metabolism and excretion. *International Journal of Cardiology*, 213, 8–14. https://doi.org/10.1016/j.ijcard.2015.08.109

Matsuba, R., Imamura, M., Tanaka, Y., Iwata, M., Hirose, H., Kaku, K., … Maeda, S. (2016). Replication study in a japanese population of six susceptibility loci for type 2 diabetes originally identified by a transethnic meta-analysis of genome-wide association studies. *PLoS ONE*, 11(4), e0154093. https://doi.org/10.1371/journal.pone.0154093

Matsuo, H., Chiba, T., Nagamori, S., Nakayama, A., Domoto, H., Phetdee, K., … Shinomiya, N. (2008). Mutations in glucose transporter 9 gene *SLC2A9* cause renal hypouricemia. *American Journal of Human Genetics*, 83(6), 744–751. https://doi.org/10.1016/j.ajhg.2008.11.001

Meschia, J. F., Nalls, M., Matarin, M., Brott, T. G., Brown, R. D. Jr, Hardy, J., … … Siblings With Ischemic Stroke Study Investigators. (2011). Siblings with ischemic stroke study: Results of a genome-wide scan for stroke loci. *Stroke*, 42(10), 2726–2732. https://doi.org/10.1161/STROKEAHA.111.620484

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., … McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*, *44*(9), 981–990. https://doi.org/10.1038/ng.2383

Mousas, A., Ntritsos, G., Chen, M.-H., Song, C. I., Huffman, J. E., Tzoulaki, I., … Reiner, A. P. (2017). Rare coding variants pinpoint genes that control human hematological traits. *PLoS Genetics*, *13*(8), 1–19. https://doi.org/10.1371/journal.pgen.1006925

Mozzi, A., Forni, D., Cagliani, R., Pozzoli, U., Vertemara, J., Bresolin, N., & Sironi, M. (2014). Albuminoid genes: Evolving at the interface of dispensability and selection. *Genome Biology and Evolution*, *6*(11), 2983–2997. https://doi.org/10.1093/gbe/evu235

Nakayama, A., Nakaoka, H., Yamamoto, K., Sakiyama, M., Shaukat, A., Toyoda, Y., … … Eurogout Consortium. (2017). GWAS of clinically defined gout and subtypes identifies multiple susceptibility loci that include urate transporter genes. *Annals of the Rheumatic Diseases*, *76*(5), 869–877. https://doi.org/10.1136/annrheumdis-2016-209632

Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M. D., Bochud, M., Coin, L., … Munroe, P. B. (2009). Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics*, *41*(6), 666–676. https://doi.org/10.1038/ng.361

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*(1), 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nikpay, M., Goel, A., Won, H.-H., Hall, L. M., Willenborg, C., Kanoni, S., … … CARDIoGRAMplusC4D Consortium. (2015). A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, *47*(10), 1121–1130. https://doi.org/10.1038/ng.3396

Oguri, M., Fujimaki, T., Horibe, H., Kato, K., Matsui, K., Takeuchi, I., & Yamada, Y. (2017). Obesity-related changes in clinical parameters and conditions in a longitudinal population-based epidemiological study. *Obesity Research and Clinical Practice*, *11*(3), 299–314. https://doi.org/10.1016/j.orcp.2016.08.008

Okada, Y., & Kamatani, Y. (2012). Common genetic factors for hematological traits in Humans. *Journal of Human Genetics*, *57*(3), 161–169. https://doi.org/10.1038/jhg.2012.2

Okada, Y., Kubo, M., Ohmiya, H., Takahashi, A., Kumasaka, N., Hosono, N., … Tanaka, T. (2012). Common variants at *CDKAL1* and *KLF9* are associated with body mass index in east Asian populations. *Nature Genetics*, *44*(3), 302–306. https://doi.org/10.1038/ng.1086

Okada, Y., Sim, X., Go, M. J., Wu, J.-Y., Gu, D., Takeuchi, F., … Tanaka, T. (2012). Meta-analysis identifies multiple loci associated with kidney function–related traits in east Asian populations. *Nature Genetics*, *44*(8), 904–909. https://doi.org/10.1038/ng.2352

Okumiya, K., Sakamoto, R., Kimura, Y., Ishimoto, Y., Wada, T., Ishine, M., … Matsubayashi, K. (2010). Diabetes mellitus and hypertension in elderly highlanders in Asia. *Journal of the American Geriatrics Society*, *58*(6), 1193–1195. https://doi.org/10.1111/j.1532-5415.2010.02862.x

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., … Torroni, A. (2006). The mtDNA legacy of the levantine early upper palaeolithic in Africa. *Science*, *314*(5806), 1767–1770. https://doi.org/10.1126/science.1135566

Oota, H., Pakstis, A. J., Bonne-Tamir, B., Goldman, D., Grigorenko, E., Kajuna, S. L. B., … Kidd, K. K. (2004). The evolution and population genetics of the *ALDH2* locus: Random genetic drift, selection, and low levels of recombination. *Annals of Human Genetics*, *68*(2), 93–109. https://doi.org/10.1046/j.1529-8817.2003.00060.x

Parachoniak, C. A., Luo, Y., Abella, J. V., Keen, J. H., & Park, M. (2011). GGA3 functions as a switch to promote met receptor recycling, essential for sustained ERK and cell migration. *Developmental Cell*, *20*(6), 751–763. https://doi.org/10.1016/j.devcel.2011.05.007

Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., … Fox, C. S. (2016). Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature Communications*, *7*, 10023. https://doi.org/10.1038/ncomms10023

Phipps-Green, A. J., Merriman, M. E., Topless, R., Altaf, S., Montgomery, G. W., Franklin, C., … Merriman, T. R. (2016). Twenty-eight loci that influence serum urate levels: Analysis of association with gout. *Annals of the Rheumatic Diseases*, *75*(1), 124–130. https://doi.org/10.1136/annrheumdis-2014-205877

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., … Pritchard, J. K. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*, *19*(5), 826–837. https://doi.org/10.1101/gr.087577.108

Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., … Furlong, L. I.. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, *2015*, bav028. https://doi.org/10.1093/database/bav028

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, *20*(1), 110–121. https://doi.org/10.1101/gr.097857.109

Povel, C. M., Boer, J. M. A., Reiling, E., & Feskens, E. J. M. (2011). Genetic variants and the metabolic syndrome: A systematic review. *Obesity Reviews*, *12*(11), 952–967. https://doi.org/10.1111/j.1467-789X.2011.00907.x

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., … Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, *499*(7459), 471–475.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1), D886–D894. https://doi.org/10.1093/nar/gky1016

Reyes-Centeno, H., Hubbe, M., Hanihara, T., Stringer, C., & Harvati, K. (2015). Testing modern human out-of-Africa dispersal models and implications for modern human origins. *Journal of Human Evolution*, *87*, 95–106. https://doi.org/10.1016/j.jhevol.2015.06.008

RStudio Team. (2016). RStudio: Integrated development environment for R. Retrieved from http://www.rstudio.com/

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., … Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832–837. https://doi.org/10.1038/nature01140

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. https://doi.org/10.1038/nature06250

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., … Erdmann, J. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*, *43*(4), 333–338. https://doi.org/10.1038/ng.784

Scott, G. K., Fei, H., Thomas, L., Medigeshi, G. R., & Thomas, G. (2006). A PACS-1, GGA3 and CK2 complex regulates CI-MPR trafficking. *The EMBO Journal*, *25*(19), 4423–4435. https://doi.org/10.1038/sj.emboj.7601336

Scott, W. R., Zhang, W., Loh, M., Tan, S.-T., Lehne, B., Afzal, U., … Kooner, J. S. (2016). Investigation of genetic variation underlying central obesity amongst South Asians. *PLoS ONE*, *11*(5), e0155478. https://doi.org/10.1371/journal.pone.0155478

Simino, J., Shi, G., Bis, J. C., Chasman, D. I., Ehret, G. B., Gu, X., … Rao, D. C. (2014). Gene-age interactions in blood pressure regulation: A large-scale investigation with the CHARGE, global BPgen, and ICBP consortia. *American Journal of Human Genetics*, *95*(1), 24–38. https://doi.org/10.1016/j.ajhg.2014.05.010

Simonson, T. S., Huff, C. D., Witherspoon, D. J., Prchal, J. T., & Jorde, L. B. (2015). Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. *Experimental Physiology*, *100*(11), 1263–1268. https://doi.org/10.1113/EP085035

Sitlani, C. M., Rice, K. M., Lumley, T., McKnight, B., Cupples, L. A., Avery, C. L., … Psaty, B. M. (2015). Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Statistics in Medicine*, *34*(1), 118–130. https://doi.org/10.1002/sim.6323

Snijder, M. B., van Dam, R. M., Visser, M., Deeg, D. J. H., Dekker, J. M., Bouter, L. M., … Lips, P. (2005). Adiposity in relation to vitamin D status and parathyroid hormone levels: A population-based study in older men and women. *The Journal of Clinical Endocrinology & Metabolism*, *90*(7), 4119–4123. https://doi.org/10.1210/jc.2005-0216

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., … Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics*, *84*(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Spaetgens, B., de Vries, F., Driessen, J. H. M., Leufkens, H. G., Souverein, P. C., Boonen, A., … Joosten, L. A. B. (2017). Risk of infections in patients with gout: A population-based cohort study. *Scientific Reports*, *7*(1), 1429. https://doi.org/10.1038/s41598-017-01588-5

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., … Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, *466*(7307), 707–713. https://doi.org/10.1038/nature09270

The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

van Rooij, F. J. A., Qayyum, R., Smith, A. V., Zhou, Y. I., Trompet, S., Tanaka, T., … Ganesh, S. K. (2017). Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for *RBPMS* in erythropoiesis. *The American Journal of Human Genetics*, *100*(1), 51–63. https://doi.org/10.1016/j.ajhg.2016.11.016

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9. https://doi.org/10.1038/nprot.2015.123

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), e72. https://doi.org/10.1371/journal.pbio.0040072

Wainschtein, P., Jain, D. P., Yengo, L., Zheng, Z., TOPMed Anthropometry Working Group, Trans-Omics for Precision Medicine Consortium, … Visscher, P. M. (2019). Recovery of trait heritability from whole genome sequence data. *BioRxiv*. https://doi.org/10.1101/588020

Wakil, S. M., Ram, R., Muiya, N. P., Mehta, M., Andres, E., Mazhar, N., … Dzimiri, N. (2016). A genome-wide association study reveals susceptibility loci for myocardial infarction/coronary artery disease in Saudi Arabs. *Atherosclerosis*, *245*, 62–70. https://doi.org/10.1016/j.atherosclerosis.2015.11.019

Wall, J. D. (2017). Inferring human demographic histories of non-African populations from patterns of allele sharing. *American Journal of Human Genetics*, *100*(5), 766–772. https://doi.org/10.1016/j.ajhg.2017.04.002

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., … Burger, J. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences*, *111*(13), 4832–4837. https://doi.org/10.1073/pnas.1316513111

Wu, J. Y., & Kuo, C. C. (2012). Pivotal role of ADP-ribosylation factor 6 in toll-like receptor 9-mediated immune signaling. *Journal of Biological Chemistry*, *287*(6), 4323–4334. https://doi.org/10.1074/jbc.M111.295113

Yabe, D., Seino, Y., Fukushima, M., & Seino, S. (2015). β cell dysfunction versus insulin resistance in the pathogenesis of type 2 diabetes in East Asians. *Current Diabetes Reports*, *15*(36). https://doi.org/10.1007/s11892-015-0602-9

Yamada, Y., Fuku, N., Tanaka, M., Aoyagi, Y., Sawabe, M., Metoki, N., … Kojima, T. (2009). Identification of *CELSR*1 as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. *Atherosclerosis*, *207*(1), 144–149. https://doi.org/10.1016/j.atherosclerosis.2009.03.038

Yamada, Y., Matsui, K., Takeuchi, I., & Fujimaki, T. (2015a). Association of genetic variants with coronary artery disease and ischemic stroke in a longitudinal population-based genetic epidemiological study. *Biomedical Reports*, *3*(3), 413–419. https://doi.org/10.3892/br.2015.440

Yamada, Y., Matsui, K., Takeuchi, I., & Fujimaki, T. (2015b). Association of genetic variants with dyslipidemia and chronic kidney disease in a longitudinal population-based genetic epidemiological study. *International Journal of Molecular Medicine*, *35*(5), 1290–1300. https://doi.org/10.3892/ijmm.2015.2152

Yamada, Y., Matsui, K., Takeuchi, I., Oguri, M., & Fujimaki, T. (2015a). Association of genetic variants of the α-kinase 1 gene with type 2 diabetes mellitus in a longitudinal population-based genetic epidemiological study. *Biomedical Reports*, *3*(3), 347–354. https://doi.org/10.3892/br.2015.439

Yamada, Y., Matsui, K., Takeuchi, I., Oguri, M., & Fujimaki, T. (2015b). Association of genetic variants with hypertension in a longitudinal population-based genetic epidemiological study. *International Journal of Molecular Medicine*, *35*(5), 1189–1198. https://doi.org/10.3892/ijmm.2015.2151

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2017). Longitudinal exome-wide association study to identify genetic susceptibility loci for hypertension in a

Japanese population. *Experimental & Molecular Medicine*, *49*(12), e409. https://doi.org/10.1038/emm.2017.209

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2018a). Identification of *CDC42BPG* as a novel susceptibility locus for hyperuricemia in a Japanese population. *Molecular Genetics and Genomics*, *293*(2), 371–379. https://doi.org/10.1007/s00438-017-1394-1

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2018b). Identification of nine novel loci related to hematological traits in a Japanese population. *Physiological Genomics*, *50*(9), 758–769. https://doi.org/10.1152/physiolgenomics.00088.2017

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2018d). Identification of three genetic variants as novel susceptibility loci for body mass index in a Japanese population. *Physiological Genomics*, *50*(3), 179–189. https://doi.org/10.1152/physiolgenomics.00117.2017

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2018e). Six novel susceptibility loci for coronary artery disease and cerebral infarction identified by longitudinal exome-wide association studies in a Japanese population. *Biomedical Reports*, *9*(2), 123–134. https://doi.org/10.3892/br.2018.1109

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., … Yamada, Y. (2019). Two novel susceptibility loci for type 2 diabetes mellitus identified by longitudinal exome-wide association studies in a Japanese population. *Genomics*, *111*(1), 34–42. https://doi.org/10.1016/j.ygeno.2017.12.010

Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., & Yamada, Y. (2018c). Identification of six novel susceptibility loci for dyslipidemia using longitudinal exome-wide association studies in a Japanese population. *Genomics*. (Published online, https://doi.org/10.1016/j.ygeno.2018.05.015).

Zabaneh, D., & Balding, D. J. (2010). A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS ONE*, *5*(8), e11961. https://doi.org/10.1371/journal.pone.0011961

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., … Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761. https://doi.org/10.1093/nar/gkx1098

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.