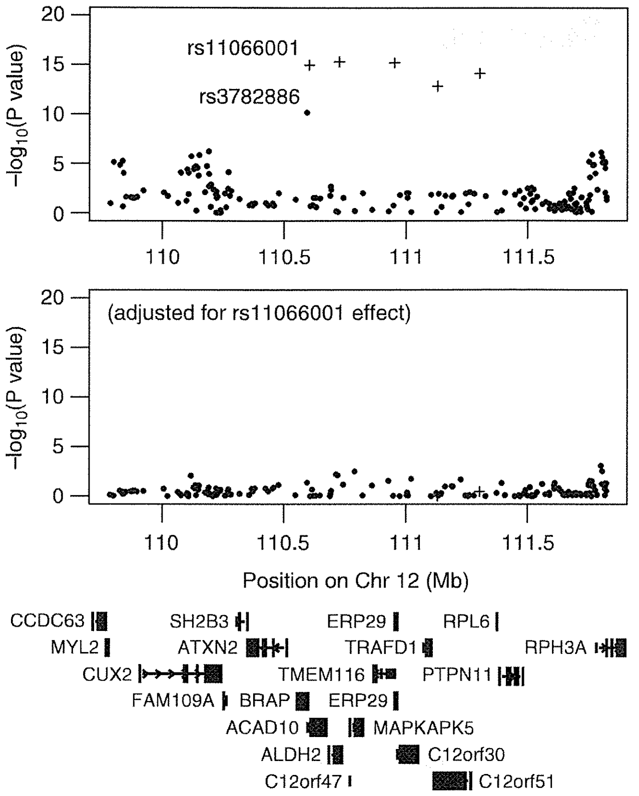


Figure 3 Plots of CAD association for the 12q24 region. In the top panel, all genotyped SNPs in the current Japanese GWA scan (dot) that passed the quality control in stage 1a, and four additionally genotyped SNPs (cross) that are in strong LD with *BRAP* rs11066001 are plotted with their $-\log_{10}(P)$ for CAD against chromosome positions (in Mb). In the middle panel, the $-\log_{10}(P)$ for CAD associations with statistical adjustment for rs11066001 are shown. The bottom panel shows the genomic location of RefSeq genes with intron and exon structures (NCBI Build 36).

across a ∼0.7-Mb interval encompassing the *BRAP* and *ALDH2* genes (Figure 3 and Supplementary Table VI). On 6p21, significant association was found across an ∼8-Mb interval encompassing the MHC region (Supplementary Figure II and Supplementary Table VI). In analyzing the joint sample of stages 1a+1b, we detected another CAD association near *CDKN2A/B* on 9p21, whereas SNPs from the other chromosomal regions did not remain after the multistage GWA scan.

In the Japanese, the strongest evidence of CAD association (P=6.9×10^{-30} at rs3782886 in *BRAP* and P=1.6×10^{-34} at rs671 in *ALDH2*; summary statistics shown in Table 2) was identified via GWA scan in a long interval on 12q24 (∼0.7 Mb), within which previously reported candidate genes, *BRAP* and *ALDH2*,[20,26] were located (Figure 3). This 12q24 region was near the locus recently identified to associate with CAD and/or MI in Europeans;[6,9] a long-range haplotype (1.6 Mb) was hypothesized to have arisen from a positive selection specific to Europeans.[9] Notably, a lead SNP in the Japanese GWA study– rs3782886 in *BRAP* – is polymorphic only in East Asians (Supplementary Figure III) and, conversely, a lead SNP in the European GWA study – rs3184504 in *SH2B3* – is polymorphic only in Europeans.[9] We retrieved a total of seven SNPs that were in strong LD (r^2≥0.8) with rs3782886 and constituted the evolutionarily derived haplotype from the HapMap data (footnote to Supplementary Table II). This haplotype turned out to show CAD association and to have arisen in East Asians independently of Europeans from the

Table 2 Association results for CAD susceptibility loci identified in the Japanese GWA study

						Discovery[d]			*Replication*					*Combined*		
												Coded allele freq.				
SNP	*Chr*	*Position (build 36)*	*Nearby gene(s)*	*Alleles (coded/other)*[b]	*Coded allele freq. in JPN*[c]	*OR (95% CI)*	*P-value*	*n*	*OR (95% CI)*	*P-value*	*n*	*Case*	*Control*	*OR (95% CI)*	*P-value*	*n, total*
rs11752643	6	32 777 351	*HLA, DRB-DQB*	T/C	0.06	2.08 (1.65–2.62)	2.0E–09	2681	1.27 (1.14–1.40)	3.9E–06	9386	0.105	0.083	1.26 (1.15–1.38)	4.7E–07	12 067
rs944797[a]	9	22 105 286	*CDKN2A/B*	C/T	0.46	1.28 (1.15–1.43)	3.8E–05	2672	1.26 (1.18–1.34)	4.4E–13	9383	0.533	0.478	1.25 (1.18–1.31)	6.1E–16	12 055
rs3782886	12	110 594 872	*BRAP*	C/T	0.31	1.54 (1.37–1.73)	2.8E–11	2681	1.38 (1.29–1.47)	1.2E–22	9384	0.373	0.301	1.38 (1.31–1.46)	6.9E–30	12 065
rs11066001	12	110 603 554	*BRAP*	C/T	0.28	1.68 (1.49–1.90)	2.1E–17	2651	1.40 (1.31–1.49)	4.2E–24	9372	0.361	0.284	1.42 (1.34–1.50)	9.1E–34	12 023
rs671	12	110 726 149	*ALDH2*	A/G	0.23	1.69 (1.50–1.91)	7.7E–18	2658	1.40 (1.31–1.49)	1.9E–24	9383	0.362	0.284	1.43 (1.35–1.51)	1.6E–34	12 041

Study design is summarized in Figure 1. Association results in the combined panel were stratified by geographical regions and combined using the Mantel–Haenszel statistics, whereas those in the discovery and replication phase panels were calculated by pooling samples within each phase. The different sex ratios between the two groups for stage 1a did not cause substantial difference in OR at the five associated loci when they were adjusted for by logistic regression (data not shown).

[a] rs1333049 was genotyped in replacement for rs944797 in the replication phase (r^2=0.92 in HapMap JPT).

[b] Alleles are nominated as those in dbSNP Build 130 mapped on the positive strand of Human Genome Build 36.3.

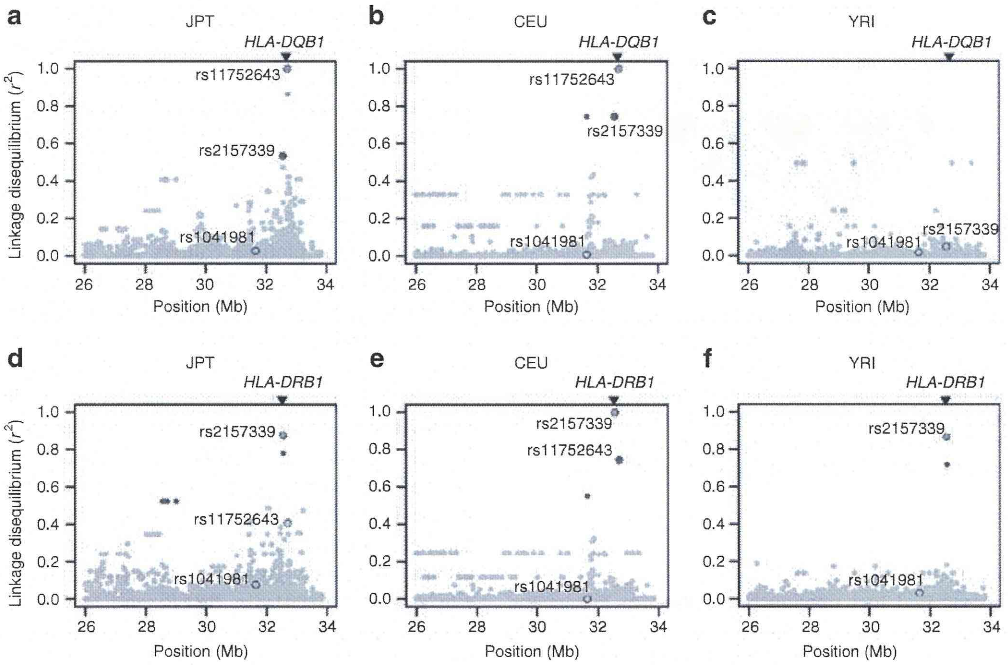
[c] Allele frequencies in the Japanese general population from GeMDBJ (n=964) or HapMap JPT (n=90; rs671).

[d] Discovery panel in this table includes stage 1a and 1b samples alone.

phylogenetic viewpoint, in a manner similar to the one that we recently reported for a blood pressure association on 12q24 in East Asians.[27]

On 6p21, we investigated CAD association across an ∼8-Mb interval encompassing the MHC region with regard to HLA types (Figure 4 and Supplementary Note). The use of high-resolution HLA and SNP haplotype map[24] enabled us to identify a significantly associated HLA haplotype, DRB1*1302–DQB1*0604 (Supplementary Tables VII and VIII). We then found that rs11752643 and rs2157339 could tag DQB1*0604 (r^2=1.0 in JPT) and DRB1*1302 (r^2=0.88 in JPT), respectively, and that these two SNPs were in LD with a candidate SNP rs1041981 in *LTA*.[21] Furthermore, our construction of phylogeny involving the three SNPs (rs11752643, rs2157339, and rs1041981) and association analysis with the resultant haplotypes revealed that an HLA allele tagged by rs11752643 would constitute a principal association signal on 6p21 (Figure 4). Independently, we also found some effects driven by population structure on 6p21 (Supplementary Figure II and Supplementary Table V). This population structure highlighted a cluster of SNPs on 6p21 (Supplementary Table V); they could be categorized into >2 subgroups with regard to LD and only one of them (represented by rs11752643) appeared to be prominent in the test of CAD association (Figure 5). Indeed, with statistical adjustment for rs11752643, most of the association signals on 6p21 were noticeably weakened (Supplementary Figure II and Supplementary Table VI). We estimated the average effect size of rs11752643 to have an odds ratio (OR) of 1.26 (95% confidence interval (95% CI) 1.15–1.38, $P=4.7\times10^{-7}$; Table 2 and Supplementary Note).

We made additional assessment of population structure in our GWA scan as indicated by a relatively high inflation factor, λ, of 1.15 in stage 1a. We found high reproducibility of association results between two different approaches, the genomic control method (PLINK)[17] and the variance component model (EMMAX),[18] indicating appropriate correction for the residual inflation of test statistic so as to constrain the risk of false positives and also to prevent the overcorrection that would remove true positives (data not shown).



g

Haplotype class	rs1041981	rs2157339	rs11752643	Samples from Nagoya in combined panel (3745 cases vs 6335 controls)			
	Tested SNPs			Haplotype frequency		OR (95% CI)	*P*
				case	control		
H1	C	A	C	0.569	0.589	0.92 (0.87, 0.98)	0.006
H2	A	A	C	0.314	0.313	1.01 (0.95, 1.07)	0.798
H3	A	T	C	0.008	0.012	0.68 (0.50, 0.92)	0.011
H4	A	T	T	0.109	0.086	1.29 (1.17, 1.42)	2.6E-07

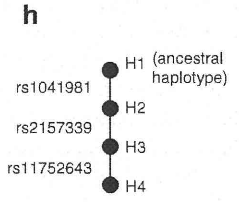


Figure 4 Overview of SNPs and haplotypes explaining CAD association on 6p21. Genotypic association between SNPs across the 8 Mb extended major histocompatibility complex region and two HLA alleles of interest – DQB1*0604 at HLA-DQB1 (top panels; **a–c**) and DRB1*1302 at HLA-DRB1 (middle panels; **d–f**) – is shown for three HapMap populations: JPT (**a**, **d**), CEU (**b**, **e**), and YRI (**c**, **f**). The SNP positions across the 8 Mb region showing weak ($r^2<0.5$; gray), moderate ($0.5\leq r^2<0.8$; blue), and strong ($r^2\geq0.8$; red) association with each HLA allele are depicted in the individual plots, above which the position of HLA-DQB1 and HLA-DRB1 is indicated by the triangle. In each plot, rs11752643 (tag SNP for DQB1*0604 in JPT (r^2=1) and CEU (r^2=1)), rs2157339 (tag SNP for DRB1*1302 in JPT (r^2=0.88), CEU (r^2=1), and YRI (r^2=0.87)), and rs1041981 (nonsynonymous SNP in *LTA*) are circled in black, except that rs11752643 is not shown for YRI because of the lack of polymorphism. Data are derived from the study by de Bakker *et al*[21] and HapMap (http://hapmap.ncbi.nlm.nih.gov/). Haplotype frequencies involving the three tested SNPs and CAD association in the largest district, Nagoya (**g**), and the phylogeny (**h**) are also shown. The color reproduction of this figure is available at the *European Journal of Human Genetics* journal online.

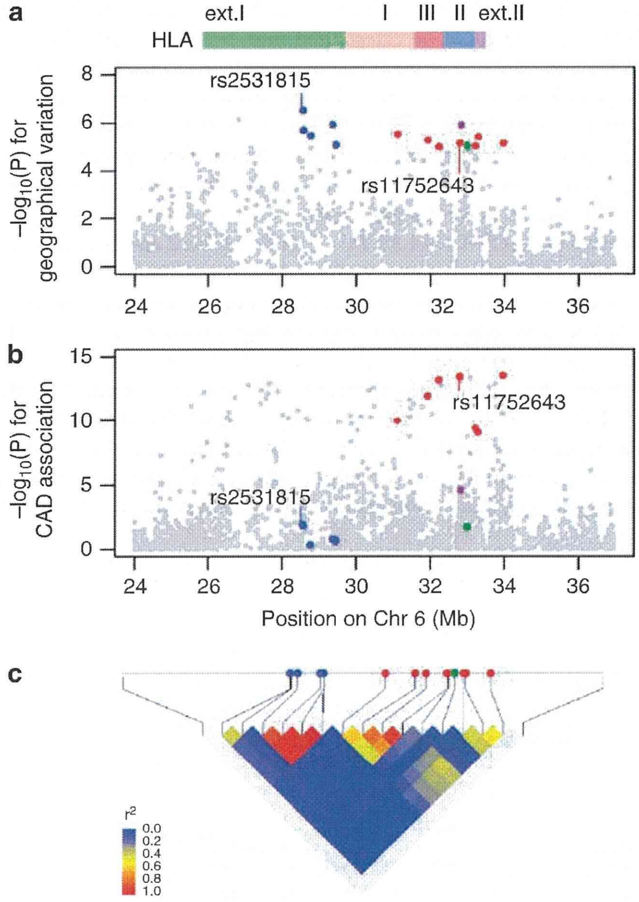


Figure 5 Relation of geographical variation (which may lead to population structure) with CAD association on 6p21. It is speculated that more than two independent subgroups of SNPs are present in the 6p21 region for geographical variation (**a**); and that one of them, including rs11752643, simultaneously accounts for CAD association and constitutes a genuine association signal (**b**). Pairwise LD coefficient r^2 among 15 SNPs that show geographical variation on 6q21 in the Japanese is demonstrated in (**c**).

Replication of previously reported SNPs

Among the CAD-associated loci previously reported in Europeans, strong association (after considering multiple testing, ie, $P<0.05/29=0.0017$) was replicated at the *CDKN2A/B* locus in the Japanese population (Table 2 and Supplementary Table IV). Despite finding significant association for rs4977574 at *CDKN2A/B* ($P=1.4\times10^{-5}$; Supplementary Table IV) in the candidate-gene-based genotyping, we proceeded with rs944797 instead of rs4977574 in the GWA scans from stages 1a+1b to stage 2; this decision was made considering the strong LD between the two SNPs ($r^2=0.94$ in stage 1a). Suggestive evidence of replication (two-tailed $P<0.1$ in the case of concordant direction of association) was observed for five other loci – *CELSR2-SORT1*, 7q22, *CNNM2-NT5C2*, *HHIPL1*, and *SMG6-SRR* – for CAD association (Supplementary Table IV).

From known variants in the LTA cascade, we tested association of rs7291467 (in *LGALS2*) and rs11066001 (in *BRAP*) with CAD and MI, as previously reported in the Japanese.[20,22] The strength of association of rs11066001 was almost equivalent to that of rs671 (in *ALDH2*), which showed the strongest association signal on 12q24, and more prominent with MI than non-MI CAD (OR=1.53, 95% CI 1.44–1.63 and $P=6.9\times10^{-40}$ for MI; OR=1.19, 95% CI 1.09–1.31 and $P=9.0\times10^{-5}$ for non-MI CAD). No significant association was found for rs7291467 regardless of the MI status (Supplementary Table IX).

DISCUSSION

We conducted a GWA study on the Japanese population with greater coverage of common SNPs (87% of all phase I+II HapMap variants (minor allele frequency ≥0. 05) in CHB+JPT) and a larger number of cases (806 subjects) and unaffected controls (1337 subjects) in the initial screen than a previous study,[21] although our sample size is rather modest by the current standard of GWA studies (and meta-analyses) of CAD or MI in populations of European descent.[2–8] Three loci – 12q24, 6p21, and 9p21 – were successfully identified through the multistage scan; notably, the 6p21 locus has not been claimed to show significant CAD association in the GWA meta-analyses of Europeans.[5,7,8] The 12q24 locus appeared to overlap with the one reported in Europeans, while there may be haplotypic heterogeneity.[6,9] In addition, the present genome-wide exploration identified significant association near *CDKN2A/B* in East Asians, as previously reported in European-descent populations.[1,2,5,7,8] Of the three loci, which we claim to be associated with CAD in the Japanese GWA study (Table 2), the disease association at two loci – 12q24 and 9p21 – have been reported in several Asian populations[10,11,26] and could be regarded confirmatory. Nevertheless, two novel findings are noted in our study: (1) the other locus within an HLA gene, *HLA-DQB1*, which is considered to explain the previous descriptions of CAD (in particular, MI) association signals on 6p21 in the Japanese[21] and (2) a pronounced association with MI, as compared with CAD, for the variants on 12q24.

The 12q24 haplotype confers risk alleles for CAD much more significantly in the Japanese than in Europeans.[6,9] A previous Japanese study reported significant association with MI risk at SNPs (including rs3782886 and rs11066001) in *BRAP*, located on 12q24.[20] The authors investigated *BRAP* (*BRCA1-associated protein*) as a candidate gene because of its potential involvement in the cytokine LTA cascade; that is, BRAP is a possible binding partner of galectin-2, which is encoded by LGALS2 that can bind to LTA. The same study group originally claimed significant association with MI risk at the *LTA* locus on 6p21 in the Japanese,[21] as discussed below. Whether a single variant or multiple variants can be present on the relevant 12q24 haplotype remains unknown. In our study, association signals at two SNPs on 12q24 – rs11066001 (in *BRAP*) and rs671 (in *ALDH2*) – were in almost complete LD ($r^2=0.99$ in the whole sample) and could not be distinguished from each other. It is possible that some molecular variant(s) in either of the two genes or other genes that are contained in the region covered by the long-range haplotype (1.5 Mb)[27] underlie susceptibility to CAD. Besides *BRAP*, one such candidate is *ALDH2*, encoding the aldehyde dehydrogenase 2; the active and inactive subunits of the enzyme are encoded by two alleles of rs671, which showed the most significant evidence of association ($P=1.6\times10^{-34}$). Using a proteomic search, Chen *et al*[28] found that enzyme activation of *ALDH2* correlated with reduced ischemic heart damage in rodent models, in accordance with the results of CAD association. With the overlapping association signals being detected, further in-depth analysis of the 12q24 region will be required to dissect the phylogenetic relationship of CAD causality between the two ethnic groups.

The second strongest association was identified via GWA scan for a cluster of SNPs in the extended MHC region on 6p21, where an HLA allele, DQB1*0604 (tagged by rs11752643), could show one of the most prominent association signals (Figure 5 and Supplementary Figure II). These findings have brought up an issue of disease association at *LTA*[21] and candidate genes in LTA cascade, such as *LGALS2*.[22] On 6p21, the association signal of rs1041981 (in *LTA*) has proven to come from that of rs11752643 via LD (D'=1.0, r^2=0.03) in the present study (Figure 4). Also, we could not detect significant association at a SNP, rs7291467, previously reported in *LGALS2* on 22q13 (OR 1.02, 95% CI 0.96–1.07, P=0.61). This is in good accordance with a pooled OR of 1.02 (95% CI, 0.99–1.06, P=0.23) that we estimated by meta-analysis involving all studies reported to date,[29] except for the original one,[22] with consideration of the winner's curse effect[30] (Supplementary Figure IV). Together, the disease association for SNPs in LTA cascade genes may be less outstanding than what was originally expected, although further investigation is warranted to address this issue.

Because of the low frequency of DQB1*0604 allele (or a proxy SNP, rs11752643; 0.02 in HapMap CEU; Supplementary Table VIII and Supplementary Figure V), a large number of samples (n >9600) would be required to verify the equivalent strength of CAD association on 6p21 in Europeans. Despite the sufficient sample size (>14 000 cases in the discovery phase),[7,8] neither of two recent large-scale meta-analyses in Europeans could show significant association on 6p21, suggesting the absence of susceptibility locus, at least, with the equivalent effect size. Thus far, positive disease association with the DQB1*0604 allele has been reported only for myasthenia gravis.[31] It is noteworthy that suggestive evidence of association has been recently observed between HLA-DRB1 and DQA1 loci and acute MI in a population-based Swedish cohort,[32] whereas the reported risk alleles are not in LD with DQB1*0604.

Moreover, this study identified CAD association at another locus, near the *CDKN2A/B* gene, originally claimed by GWA scan in Europeans. The effect size for CAD was almost comparable between the populations: OR=1.25 and 1.29 for the Japanese and Europeans,[5,7,8] respectively. Because of the limitation of discovery phase sample size (stages 1a+1b), statistical power is insufficient (power <0.7) to refute the disease association at 19 of 23 candidate loci previously identified in GWA studies of Europeans, except *CDKN2A/B* and 5 other loci showing suggestive evidence of replication (Supplementary Table IV).

Although GWA studies have thus far made major steps in unraveling the genetics of cardiovascular disease, biological explanations for the identified associations remain largely unknown. In this context, it has been hypothesized that different genetic risk factors may contribute to either CAD or MI given the complex nature of transition from a normal coronary artery to MI. Indeed, a recent GWA meta-analysis[33] has addressed this issue and has indicated the presence of genetic predispositions leading to MI (in the presence of coronary atherosclerosis) as well as those shared between CAD and MI (promoting coronary atherosclerosis). The present findings of SNP–trait associations can support this notion; that is, the strength of association is prominent for MI at *BRAP/ALDH2* on 12q24, whereas it is comparable between the disease entities at *CDKN2A/B* on 9q21 (Supplementary Table IX).

Our GWA study results have called our attention to sample admixture with geographical variation, a potential source of population structure. Because the bias introduced by population structure may not be as significant as once feared, performing GWA studies in admixed populations was suggested to be a reasonable strategy in increasing sample sizes with the necessity of independent replication.[34] Nevertheless, as has been stated by previous GWA studies in Europeans,[2] apparent disease associations in the few genomic regions identified as showing geographical differentiation need to be interpreted with caution. This seems to be the case with the 6p21 region, where more than two independent subgroups of SNPs are present for geographical variation and one of them, rs11752643, appears to simultaneously account for CAD association even after adjustment for geographical variation (Figure 5, Supplementary Note, Supplementary Figure II and Supplementary Table IX).

Just like most other GWA studies on CAD conducted to date, the present study has all the potential biases of cross-sectional analysis, most importantly a survival bias, as recently pointed out by prospective studies.[35,36] Also, although we attempted to explore CAD susceptibility loci with rather prominent effect sizes, the power cannot be necessarily sufficient; for example, the power to detect OR=1.2 is very low (Supplementary Table X). We should keep these limitations in mind when we interpret the genetic association results.

In conclusion, our GWA study has confirmed that three loci in three chromosomal regions are associated with CAD in the Japanese. These loci highlight the likely presence of risk alleles with largely two types of genetic effects – population specific and cosmopolitan – on susceptibility to CAD. The integration of multiethnic results will promote a better understanding of the global genetic architecture of CAD.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all the people who have continuously supported the Hospital-based Cohort Study at the National Center for Global Health and Medicine, the Amagasaki Study and the Kyushu University Fukuoka Cohort Study, and the KING Study. We also thank Drs Suminori Kono, Ken Sugimoto, Kei Kamide, and Chikanori Makibayashi, and the many physicians of the participating hospitals and medical institutions for their assistance in collecting the DNA samples and accompanying clinical information. This work was supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; the Program for Promotion of Fundamental Studies in Health Sciences, National Institute of Biomedical Innovation Organization (NIBIO); and KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

1 Schunkert H, Erdmann J, Samani NJ: Genetics of myocardial infarction: a progress report. *Eur Heart J* 2010; **31**: 918–925.
2 Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–978.
3 Trégouët DA, König IR, Erdmann J *et al*: Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* 2009; **41**: 283–285.
4 Erdmann J, Grosshennig A, Braund PS *et al*: New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* 2009; **41**: 280–282.
5 Myocardial Infarction Genetics Consortium: Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009; **41**: 334–341.
6 Gudbjartsson DF, Bjornsdottir US, Halapi E *et al*: Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009; **41**: 342–347.
7 Schunkert H, König IR, Kathiresan S *et al*: Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011; **43**: 333–338.
8 Coronary Artery Disease (C4D) Genetics Consortium: A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* 2011; **43**: 339–344.
9 Soranzo N, Spector TD, Mangino M *et al*: A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet* 2009; **41**: 1182–1190.
10 Shen GQ, Li L, Rao S *et al*: Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2008; **28**: 360–365.

11 Saleheen D, Alexander M, Rasheed A *et al*: Association of the 9p21.3 locus with risk of first-ever myocardial infarction in Pakistanis: case-control study in South Asia and updated meta-analysis of Europeans. *Arterioscler Thromb Vasc Biol* 2010; **30**: 1467–1473.

12 Moriyama Y, Okamura T, Inazu A *et al*: A low prevalence of coronary heart disease among subjects with increased high-density lipoprotein cholesterol levels, including those with plasma cholesteryl ester transfer protein deficiency. *Prev Med* 1998; **27**: 659–667.

13 Okamura T: Dyslipidemia and cardiovascular disease: a series of epidemiologic studies in Japanese populations. *J Epidemiol* 2010; **20**: 259–265.

14 Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**: 209–213.

15 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.

16 Takeuchi F, Serizawa M, Yamamoto K *et al*: Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009; **58**: 1690–1699.

17 Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

18 Kang HM, Sul JH, Service SK *et al*: Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; **42**: 348–354.

19 Browning BL, Browning SR: A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; **84**: 210–223.

20 Ozaki K, Sato H, Inoue K *et al*: SNPs in BRAP associated with risk of myocardial infarction in Asian populations. *Nat Genet* 2009; **41**: 329–333.

21 Ozaki K, Ohnishi Y, Iida A *et al*: Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; **32**: 650–654.

22 Ozaki K, Inoue K, Sato H *et al*: Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. *Nature* 2004; **429**: 72–75.

23 Pickrell JK, Coop G, Novembre J *et al*: Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 2009; **19**: 826–837.

24 de Bakker PI, McVean G, Sabeti PC *et al*: A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006; **38**: 1166–1172.

25 Iwakawa M, Goto M, Noda S *et al*: DNA repair capacity measured by high throughput alkaline comet assays in EBV-transformed cell lines and peripheral blood cells from cancer patients and healthy volunteers. *Mutat Res* 2005; **588**: 1–6.

26 Takagi S, Iwai N, Yamauchi R *et al*: Aldehyde dehydrogenase 2 gene is a risk factor for myocardial infarction in Japanese men. *Hypertens Res* 2002; **25**: 677–681.

27 Kato N, Takeuchi F, Tabara Y *et al*: Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet* 2011; **43**: 531–538.

28 Chen CH, Budas GR, Churchill EN, Disatnik MH, Hurley TD, Mochly-Rosen D: Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* 2008; **321**: 1493–1495.

29 Li W, Xu J, Wang X *et al*: Lack of association between lymphotoxin-alpha, galectin-2 polymorphisms and coronary artery disease: a meta-analysis. *Atherosclerosis* 2010; **208**: 433–436.

30 Zollner S, Pritchard JK: Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* 2007; **80**: 605–615.

31 Deitiker PR, Oshima M, Smith RG, Mosier DR, Atassi MZ: Subtle differences in HLA DQ haplotype-associated presentation of AChR alpha-chain peptides may suffice to mediate myasthenia gravis. *Autoimmunity* 2006; **39**: 277–288.

32 Björkbacka H, Lavant EH, Fredrikson GN *et al*: Weak associations between human leucocyte antigen genotype and acute myocardial infarction. *J Intern Med* 2010; **268**: 50–58.

33 Reilly MP, Li M, He J *et al*: Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* 2011; **377**: 383–392.

34 Hao K, Chudin E, Greenawalt D, Schadt EE: Magnitude of stratification in human populations and impacts on genome wide association studies. *PLoS One* 2010; **5**: e8695.

35 Karvanen J, Silander K, Kee F *et al*: The impact of newly identified loci on coronary heart disease, stroke and total mortality in the MORGAM prospective cohorts. *Genet Epidemiol* 2009; **33**: 237–246.

36 Ripatti S, Tikkanen E, Orho-Melander M *et al*: A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 2010; **376**: 1393–1400.

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

nature genetics

Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians

Norihiro Kato[1,42*], Fumihiko Takeuchi[1,42], Yasuharu Tabara[2,42], Tanika N Kelly[3,42], Min Jin Go[4,42], Xueling Sim[5,42], Wan Ting Tay[6,42], Chien-Hsiun Chen[7,8,42], Yi Zhang[9,10,42], Ken Yamamoto[11,42], Tomohiro Katsuya[12,42], Mitsuhiro Yokota[13,42], Young Jin Kim[4], Rick Twee Hee Ong[14], Toru Nabika[15], Dongfeng Gu[16,17], Li-ching Chang[7], Yoshihiro Kokubo[18], Wei Huang[19], Keizo Ohnaka[20], Yukio Yamori[21], Eitaro Nakashima[22,23], Cashell E Jaquish[24], Jong-Young Lee[4], Mark Seielstad[25], Masato Isono[1], James E Hixson[26], Yuan-Tsong Chen[7], Tetsuro Miki[27], Xueya Zhou[28], Takao Sugiyama[29], Jae-Pil Jeon[4], Jian Jun Liu[30], Ryoichi Takayanagi[31], Sung Soo Kim[4], Tin Aung[6,32], Yun Ju Sung[33], Xuegong Zhang[28], Tien Yin Wong[6,32,34], Bok-Ghee Han[4], Shotai Kobayashi[35], Toshio Ogihara[36,43], Dingliang Zhu[9,10,43], Naoharu Iwai[37,43], Jer-Yuarn Wu[7,8,43], Yik Ying Teo[5,14,29,38,39,43], E Shyong Tai[39,40,43], Yoon Shin Cho[4,43] & Jiang He[3,41,43]

We conducted a meta-analysis of genome-wide association studies of systolic (SBP) and diastolic (DBP) blood pressure in 19,608 subjects of east Asian ancestry from the AGEN-BP consortium followed up with *de novo* genotyping (n = 10,518) and further replication (n = 20,247) in east Asian samples. We identified genome-wide significant ($P < 5 \times 10^{-8}$) associations with SBP or DBP, which included variants at four new loci (*ST7L-CAPZA1*, *FIGN-GRB14*, *ENPEP* and *NPR3*) and a newly discovered variant near *TBX3*. Among the five newly discovered variants, we obtained significant replication in the independent samples for all of these loci except *NPR3*. We also confirmed seven loci previously identified in populations of European descent. Moreover, at 12q24.13 near *ALDH2*, we observed strong association signals ($P = 7.9 \times 10^{-31}$ and $P = 1.3 \times 10^{-35}$ for SBP and DBP, respectively) with ethnic specificity. These findings provide new insights into blood pressure regulation and potential targets for intervention.

Hypertension is a leading risk factor of cardiovascular disease and premature death globally[1–4]. It is especially common in Asian populations, contributing to a high incidence of and mortality from stroke[5,6]. Genetic and environmental factors and their interaction determine an individual's risk for hypertension. Efforts to elucidate the genetic determinants of hypertension, or elevated blood pressure levels[7,8], yielded little success until 2009 when two large-scale meta-analyses of genome-wide association studies (GWAS) from the Global Blood Pressure Genetics (Global BPgen) and Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) consortia identified a total of 13 independent loci significantly associated with blood pressure variation[9,10]. Although the results from the two consortia represented an important advance in hypertension research[11], these studies were conducted almost exclusively in populations of European descent. Studies in non-European populations will allow us to assess the relevance of these findings to other ethnic groups and potentially discover new variants. The latter is important because some variants may be more common in specific ethnic groups, thereby providing greater power, or the effects of genetic variants on blood pressure may be larger in specific ethnic groups.

The Asian Genetic Epidemiology Network Blood Pressure (AGEN-BP) working group was established to facilitate the identification of genetic variants influencing blood pressure among populations of east Asian ancestry. AGEN-BP includes 19,608 east Asian participants from eight population- and family-based GWAS with standardized blood pressure measurements. Here we report the findings from our three-stage study that included a meta-analysis of blood pressure GWAS from AGEN-BP (stage 1; n = 19,608), *de novo* genotyping of top loci in additional individuals (stage 2; n = 10,518) and replication of findings in independent east Asian samples (stage 3; n = 20,247) (**Supplementary Fig. 1**).

RESULTS

Meta-analysis and follow up of new association signals

The stage 1 meta-analysis included 19,608 individuals from eight GWAS in east Asian populations (**Table 1** and **Supplementary Table 1**). The $-\log_{10} P$ values by chromosome location for SBP and DBP are shown in **Figure 1a** and **b**, respectively. Quantile-quantile plots for SBP and DBP are presented in **Supplementary Figure 2**. The overall inflation after meta-analysis was modest (λ_{GC} = 1.03 for SBP and λ_{GC} = 1.02 for DBP). The meta-analysis

*A full list of author affiliations appears at the end of the paper.

Received 20 December 2010; accepted 20 April 2011; published online 15 May 2011; doi:10.1038/ng.834

Table 1 Study design and sample characteristics

Study	*n*	Ancestry	Blood pressure measurement (device, number of measures)	Genotyping platform	Women[a]	Age (s.d.)[b]	SBP (s.d.)[c]	DBP (s.d.)[c]	BMI (s.d.)[d]	HTN[a,e]	Antihypertensive therapy[a]
Stage 1: AGEN-BP GWAS meta-analysis (*n* = 19,608)											
CAGE	1,547	Japanese	Standard mercury sphygmomanometer, 2–3/ digital, 2–3	Illumina 550/610K	42.8	66.1 (8.0)	134.1 (20.3)	76.8 (11.9)	23.5 (3.3)	56.1	37.9
GenSalt	1,881	Han Chinese	Random zero sphygmomanometer, 9	Affymetrix 6.0	47.2	38.7 (9.5)	116.9 (14.2)	73.7 (10.3)	23.3 (3.2)	9.8	0.37
KARE	8,842	Korean	Standard mercury sphygmomanometer, 3	Affymetrix 5.0	52.7	52.2 (8.9)	118.7 (19.4)	75.6 (12.0)	24.6 (3.1)	22.3	10.9
Shanghai-Ruijin	455	Han Chinese	Standard mercury sphygmomanometer, 2–3	Illumina 610K	49.5	54.2 (7.0)	109.1 (8,6)	72.7 (8.6)	21.9 (1.9)	0	0
SiMES	2,519	Malay	Digital, 2–3	Illumina 610K	50.5	59.0 (11.0)	147.9 (23.9)	80.1 (11.3)	26.4 (5.1)	66.6	22.9
SP2	2,431	Han Chinese	Digital, 2–3	Illumina 550K/610K/1M	46.5	48.1 (11.2)	129.4 (19.8)	76.9 (10.8)	22.9 (3.7)	34.6	14.3
Suita (1)	933	Japanese	Random zero sphygmomanometer, 3	Illumina 550K	56.3	59.0 (7.0)	119.7 (17.5)	75.0 (10.6)	22.7 (2.9)	14.8	0
Taiwan	1,000	Han Chinese	Digital, 2–3	Illumina 550K	49.8	51.2 (17.8)	121.9 (18.5)	76.2 (10.8)	23.8 (3.5)	10.9	6.8
Stage 2: *de novo* genotyping follow-up study (*n* = 10,518)											
CAGE-Amagasaki	5,331	Japanese	Digital, 2–3	TaqMan	39.8	47.8 (12.3)	124.3 (17.3)	75.9 (11.0)	23.0 (3.2)	21.5	9.0
Ehime	2,895	Japanese	Digital, 3	TaqMan	56.6	61.1 (14.0)	137.7 (22.2)	81.0 (11.8)	23.4 (3.2)	53.4	25.7
Suita (2)	2,292	Japanese	Random zero sphygmomanometer, 3	TaqMan	53.8	67.2 (11.0)	127.4 (19.0)	76.5 (10.3)	22.9 (3.2)	48.1	36.6
Stage 3: replication study (*n* = 20,247)											
CAGE-Fukuoka	12,569	Japanese	Digital, 2	TaqMan	54.9	62.6 (6.8)	138.9 (21.2)	83.9 (11.7)	23.1 (3.0)	57.9	23.9
CAGE-KING	3,975	Japanese	Digital, 2	TaqMan	56.9	63.6 (6.6)	132.3 (19.8)	76.9 (11.2)	22.9 (3.0)	48.4	24.3
HEXA-shared control	3,703	Korean	Standard mercury sphygmomanometer, 3	Affymetrix 6.0	55.4	53.2 (8.3)	121.7 (14.4)	77.1 (9.9)	24.0 (2.9)	18.0	0

AGEN-BP, Asian Genetic Epidemiology Network Blood Pressure; CAGE, Cardio-metabolic Genome Epidemiology Network; GenSalt, Genetic Epidemiology Network of Salt-Sensitivity; KARE, Korean Association Resource Project; Shanghai-Ruijin, Shanghai Hypertension Study; SiMES, Singapore Malay Eye Survey; SP2, Singapore Prospective Study; Suita, The Suita Study; Taiwan, Taiwan Type 2 Diabetes Study; *n*, sample size; s.d., standard deviation; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; HTN, hypertension; GWAS, genome-wide association study.

[a] Data are percentages. [b] Age in years. [c] Measurements in mm Hg. [d] Measurements in kg/m^2. [e] Hypertension is defined as SBP ≥ 140 mm Hg and/or DBP ≥ 90 mm Hg or taking antihypertensive medication.

identified two independent association signals reaching genome-wide significance (defined as $P < 5 \times 10^{-8}$). These included one locus (*ATP2B1*) known to harbor variants associated with blood pressure and one newly discovered locus (*FIGN-GRB14*). Eleven additional variants that have not been previously implicated in the pathogenesis of hypertension were associated with SBP and/or DBP at a significance level of $P < 1 \times 10^{-5}$ (**Supplementary Table 2**).

To increase statistical power and strengthen support for the stage 1 findings, we undertook stage 2 follow up *de novo* genotyping for 13 SNPs in an additional 10,518 Japanese individuals. These included 12 SNPs that showed associations with $P < 1 \times 10^{-5}$ and one additional SNP with $P < 1 \times 10^{-3}$ that was close to a biological candidate gene (*NPR3*)[12]. In the joint analysis of stages 1 and 2, one additional SNP (rs11066280 at the *RPL6-PTPN11* locus) reached genome-wide significance. We carried this SNP and five additional SNPs attaining borderline significance (defined as $P < 5 \times 10^{-6}$) in our joint analysis of stages 1 and 2 forward to a stage 3 replication study involving 20,247 individuals. All six SNPs showed some evidence of replication in the stage 3 study, although the SBP association for rs1173766 ($P = 0.016$) did not reach a significance level after adjustment for multiple testing ($P = 0.05/6 \approx 0.008$). All six SNPs reached genome-wide significance in the joint analysis of stages 1, 2 and 3 (**Table 2** and **Fig. 2**). Moreover, two additional variants (rs880315 at the *CASZ1* locus and rs155524 at the *ITGA9* locus) had previously been genotyped in our stage 2 and stage 3 samples[13]. When combined with data from stage 1, we found that rs880315 at *CASZ1* showed an association with DBP that reached genome-wide significance as previously reported[13].

Finally, on 12q24.21 near *TBX3*, we detected an association between blood pressure and rs35444, which is approximately 200 kb away from the reported lead SNP (rs2384550) identified in populations of European descent and its proxies. Because there is no linkage disequilibrium (LD)

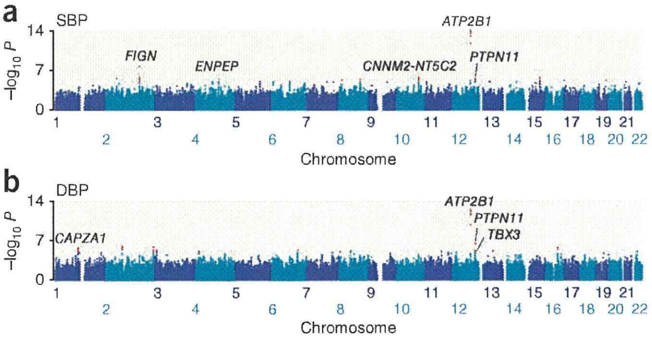


Figure 1 Genome-wide association results for the AGEN-BP meta-analysis for blood pressure. Manhattan plots show the significance of association between all SNPs and SBP (**a**) and DBP (**b**) in the stage 1 meta-analysis. Signals with suggestive levels of significance ($P < 10^{-5}$) are highlighted in red. Seven named loci showed genome-wide significance ($P < 5 \times 10^{-8}$) in the joint analysis (stages 1, 2 and 3; two reported loci, *CNNM2-NT5C2* and *ATP2B1*, were followed up in part of stage 2 and stage 3). The genes used to name signals have been chosen on the basis of proximity to the lead SNP and should not be presumed to indicate causality.

Table 2 Top genome-wide association results for SBP and DBP

								SBP		DBP	
Chr.	SNP ID	Position in NCBI build 36.3	Coded/other allele	Nearby gene(s)	Stage	*n*	Coded allele freq.	β (s.e.)[a]	*P*	β (s.e.)[a]	*P*
Loci newly identified or unique to east Asians											
1	rs17030613	112,971,190	C/A	*ST7L*	1	19,251	0.47	0.52 (0.17)	0.002	0.50 (0.11)	3.2×10^{-6}
				CAPZA1	2	10,465	0.50	0.43 (0.24)	0.073	0.25 (0.14)	0.078
					1+2	29,716	0.48	0.49 (0.14)	4.0×10^{-4}	0.41 (0.09)	1.8×10^{-6}
					3	20,236	0.50	0.49 (0.18)	0.007	0.34 (0.11)	0.002
					1+2+3	49,952	0.49	0.49 (0.11)	8.4×10^{-6}	**0.38 (0.07)**	**1.2×10^{-8}**
2	rs16849225	164,615,066	C/T	*FIGN*	1	18,867	0.60	**0.97 (0.17)**	**2.2×10^{-8}**	0.40 (0.11)	3.4×10^{-4}
				GRB14	2	10,465	0.62	0.56 (0.25)	0.022	0.26 (0.15)	0.068
					1+2	29,332	0.61	**0.84 (0.14)**	**3.8×10^{-9}**	0.35 (0.09)	7.7×10^{-5}
					3	20,179	0.62	0.60 (0.19)	0.001	0.20 (0.11)	0.076
					1+2+3	49,511	0.61	**0.75 (0.11)**	**3.5×10^{-11}**	0.29 (0.07)	2.7×10^{-5}
4	rs6825911	111,601,087	C/T	*ENPEP*	1	19,033	0.48	0.79 (0.17)	3.3×10^{-6}	0.40 (0.11)	1.7×10^{-4}
					2	10,463	0.54	0.66 (0.24)	0.007	0.40 (0.14)	0.005
					1+2	29,496	0.50	0.75 (0.14)	7.8×10^{-8}	0.40 (0.09)	2.5×10^{-6}
					3	20,019	0.52	0.33 (0.18)	0.070	0.36 (0.11)	9.1×10^{-4}
					1+2+3	49,515	0.51	0.60 (0.11)	7.3×10^{-8}	**0.39 (0.07)**	**9.0×10^{-9}**
5	rs1173766	32,840,285	C/T	*NPR3*	1	19,414	0.62	0.64 (0.17)	2.2×10^{-4}	0.38 (0.11)	6.1×10^{-4}
					2	10,461	0.58	0.94 (0.24)	1.3×10^{-4}	0.54 (0.14)	1.8×10^{-4}
					1+2	29,875	0.61	0.74 (0.14)	1.7×10^{-7}	0.44 (0.09)	5.9×10^{-7}
					3	20,095	0.58	0.45 (0.19)	0.016	0.25 (0.11)	0.026
					1+2+3	49,970	0.60	**0.63 (0.11)**	**1.9×10^{-8}**	0.36 (0.07)	1.2×10^{-7}
12	rs11066280	111,302,166	T/A	*RPL6*	1	16,268	0.78	1.03 (0.22)	3.8×10^{-6}	0.73 (0.14)	3.1×10^{-7}
				PTPN11	2	10,453	0.75	1.41 (0.28)	3.2×10^{-7}	0.83 (0.16)	3.5×10^{-7}
				ALDH2	1+2	26,721	0.77	**1.18 (0.17)**	**1.0×10^{-11}**	**0.77 (0.11)**	**5.9×10^{-13}**
					3	20,236	0.74	**2.13 (0.21)**	**2.6×10^{-23}**	**1.34 (0.13)**	**6.5×10^{-27}**
					1+2+3	46,957	0.75	**1.56 (0.13)**	**7.9×10^{-31}**	**1.01 (0.08)**	**1.3×10^{-35}**
12	rs35444	114,036,820	A/G	*TBX3*	1	19,286	0.75	0.69 (0.19)	3.7×10^{-4}	0.54 (0.12)	8.1×10^{-6}
					2	10,460	0.75	0.49 (0.28)	0.077	0.48 (0.16)	0.003
					1+2	29,746	0.75	0.63 (0.16)	8.6×10^{-5}	0.52 (0.10)	9.6×10^{-8}
					3	20,238	0.75	0.64 (0.21)	0.003	0.46 (0.13)	3.0×10^{-4}
					1+2+3	49,984	0.75	0.63 (0.13)	7.5×10^{-7}	**0.50 (0.08)**	**1.3×10^{-10}**
Loci previously identified in Europeans											
1	rs880315	10,719,453	C/T	*CASZ1*	1	10,765	0.61	0.29 (0.24)	0.226	0.26 (0.14)	0.073
					Follow-up[b]	21,846	0.67	1.03 (0.19)	8.1×10^{-8}	**0.72 (0.11)**	**5.9×10^{-11}**
					Joint analysis	32,611	0.65	0.74 (0.15)	7.3×10^{-7}	**0.56 (0.09)**	**3.1×10^{-10}**
4	rs16998073	81,541,520	T/A	*FGF5*	1		N/A	N/A		N/A	
					Follow-up[b]	21,864	0.30	**1.43 (0.20)**	**3.9×10^{-13}**	**0.76 (0.11)**	**2.0×10^{-11}**
10	rs11191548	104,836,168	T/C	*CNNM2*	1	19,457	0.74	0.91 (0.19)	2.1×10^{-6}	0.49 (0.12)	5.6×10^{-5}
				NT5C2	Follow-up[b]	21,858	0.73	**1.47 (0.20)**	**4.9×10^{-13}**	**0.66 (0.12)**	**1.6×10^{-8}**
				CYP17A1	Joint analysis	41,315	0.74	**1.18 (0.14)**	**3.9×10^{-17}**	**0.58 (0.08)**	**6.6×10^{-12}**
12	rs17249754	88,584,717	G/A	*ATP2B1*	1	18,856	0.65	**1.38 (0.18)**	**7.6×10^{-15}**	**0.83 (0.11)**	**3.2×10^{-13}**
					Follow-up[b]	21,863	0.63	0.94 (0.19)	4.2×10^{-7}	0.35 (0.11)	1.2×10^{-3}
					Joint analysis	40,719	0.64	**1.17 (0.13)**	**7.7×10^{-20}**	**0.58 (0.08)**	**1.9×10^{-13}**

β is the effect size on blood pressure in mm Hg per coded allele based on an additive genetic model. Shown is the top SNP for each independent locus significantly ($P < 5 \times 10^{-8}$) associated with systolic and/or diastolic blood pressure on joint analysis in up to 50,373 individuals of east Asian ancestry. Detailed results for the individual loci separately by study are presented in **Supplementary Table 2**. For the Fukuoka study and KING study in stage 3, we used genotype data of rs12413409 for rs11191548 ($r^2 = 0.98$) at *CNNM2-NT5C2* and a proxy (rs2681472, $r^2 = 0.98$) for rs17249754 at *ATP2B1*; the linkage disequilibrium coefficient (r^2) was estimated based on 5,331 Japanese samples (CAGE-Amagasaki study). Four loci (*CASZ1*, *FGF5*, *CNNM2-NT5C2* and *ATP2B1*) in the table were previously reported to associate with blood pressure in a Japanese replication study[13], the samples of which constitute the participants in the present GWAS meta-analysis. Imputed data were unavailable for rs16998073 at *FGF5*, and only the results for the follow-up analysis are shown in the table. [a]Measurements in mm Hg. [b]Lead SNPs at the loci previously identified in Europeans were directly genotyped for follow-up in part of stage 2 and stage 3 samples. Chr., chromosome.

between the two clusters of SNPs on 12q24.21 in either ethnic group ($r^2 = 0.000$, $D' = 0.018$ in HapMap JPT+CHB; $r^2 = 0.003$, $D' = 0.063$ in HapMap CEU between rs35444 and rs2384550), we consider the association at rs35444 independent and new, presumably indicating the presence of allelic heterogeneity between the ethnic groups.

East Asian–specific associations at 12q24.13

We detected one of the most prominent blood pressure associations at 12q24.13 (rs11066280, effect allele frequency = 0.75, per-allele effect = 1.56 mm Hg for SBP ($P = 7.9 \times 10^{-31}$) and 1.01 mm Hg for DBP ($P = 1.3 \times 10^{-35}$) in the joint analysis of stages 1, 2 and 3; **Table 2**). This association is likely driven by a known functional variant rs671 at *ALDH2*, which is common in east Asians and has been reported to associate with hypertension principally through modification of alcohol consumption[14–17]. Among the proxy SNPs in the surrounding region, we found strong LD ($r^2 = 0.87$) between rs11066280 and rs671 (**Supplementary Table 3**). In addition, both SNPs showed an association with blood pressure in the

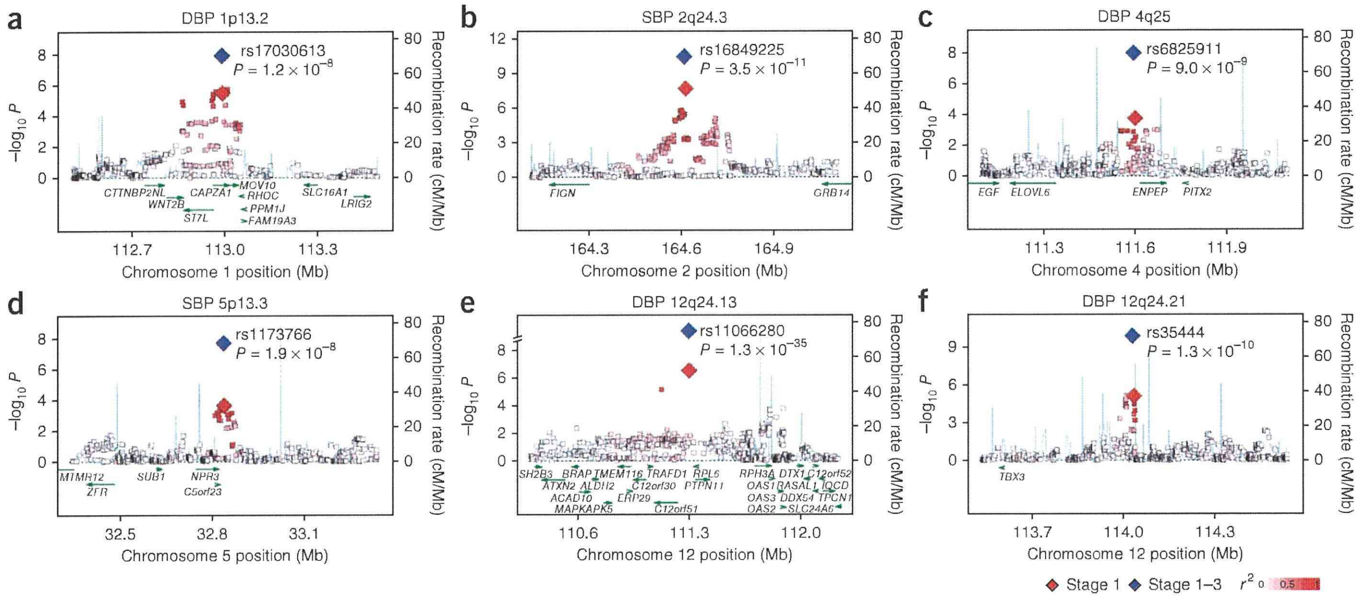


Figure 2 Regional association plots of six blood pressure loci. (**a**–**f**) Genotyped and imputed SNPs passing quality control measures in stage 1 are plotted with their meta-analysis *P* values (as $-\log_{10}$ values) as a function of genomic position (build 36). In each panel, the lead SNP is represented by a diamond, with stage 1 meta-analysis results denoted by a red diamond and the joint analysis (combined *P*) results denoted by a blue diamond. The correlation of the lead SNP to other SNPs at the locus is shown on a scale from minimal (white) to maximal (red). Superimposed on the plot are gene locations (green) and recombination rates (light blue). The regional plots were drawn using the SNAP software[35]. At 12q24.13, rs671 was not included in the construction of the regional association plots (**e**) because the genotype data for this SNP were unavailable from two GWAS in stage 1, KARE and Suita (1), which resulted in an effective sample size of 9,828.

CAGE-Amagasaki, Fukuoka and KING Studies (n = 21,875; **Supplementary Table 4**). When both SNPs were simultaneously included in the regression model, statistical significance remained for rs671 ($P = 0.018$ for SBP and $P = 8.9 \times 10^{-4}$ for DBP) but not for rs11066280 ($P > 0.05$). However, conclusive evidence for the primary source of association at 12q24.13 remains to be provided. The two SNPs are close to *SH2B3* (**Fig. 3a**), one of the blood pressure loci identified in populations of European descent[9,10].

The SNP at 12q24.13 associated with blood pressure in populations of European descent (rs3184504 at the *SH2B3* locus) is not polymorphic in east Asians, and the SNP associated with blood pressure in our study (rs671 at the *ALDH2* locus) is not polymorphic in Europeans. Also, we found modest signatures of recent selection, which was supported by several types of population genetic evidence, for example, reduction of haplotype diversity in east Asians, and suggestive evidence of a selective sweep in east Asians at 110.7–111.4 Mb (empirical $P < 0.01$) by two haplotype-based tests: iHS scores[18] and XP-EHH[19] (**Fig. 3a,b** and **Supplementary Fig. 3**). These could be further supported by data from the HapMap Project (release 22) and the 1000 Genomes Project (pilot 1; see URLs); that is, eight common SNPs (minor allele frequency (MAF) = 0.23–0.24 in the Japanese population) appeared to identify a common haplotype (H5), which arose on a haplotype (H4) that is common in east Asians but is absent in Europeans and is rare in Africans (**Supplementary Fig. 4**), spanning almost the same interval length (0.7 Mb) as that of a suggestive selective sweep. Moreover, in the haplotype analysis involving five (of eight) SNPs genotyped at 12q24.13, rs671 and rs11066132 (r^2 = 0.99 between the two SNPs) clearly differentiated a group of blood-pressure–increasing haplotypes (H1 to H3) from that of blood-pressure–decreasing haplotypes (H4 to H6; **Supplementary Fig. 5**).

These SNPs showed substantial pleiotropic effects on risk factors for cardiovascular disease[17] as well as on susceptibility to coronary artery disease (CAD) in the Japanese population (F. Takeuchi *et al.*, unpublished data), with the strongest association detected at rs671 for several traits besides blood pressure in the Japanese population studied (**Fig. 3c** and **Supplementary Table 4**). Two additional observations are of note. First, the alleles associated with increased blood pressure were associated with relatively small elevations in fasting glucose and body mass index (BMI), which would be expected to result in elevated risk of CAD. However, the effects of rs671 on low-density lipoprotein (LDL) cholesterol and high-density lipoprotein (HDL) cholesterol were larger than those on blood pressure, glucose and BMI and in a direction that would be expected to decrease CAD risk. Indeed, overall, rs671 was associated with reduced risk of CAD (OR = 0.59 for the G allele versus the A allele (95% CI 0.52–0.66), $P = 7.7 \times 10^{-18}$; **Supplementary Table 4**). A previous European GWAS[20] reported significant evidence supporting CAD association of a haplotype involving rs3184504 (at *SH2B3*) identical to that for blood pressure association as well as signatures of natural selection in an overlapping region on 12q24.13. By inspecting the phylogeny of haplotypes in the relevant region, we confirmed that these haplotypes arose independently and that each of the haplotypes was responsible for independent association signals in the two populations (**Supplementary Fig. 4**). Second, given the biological role of ALDH2 in the propensity for alcohol consumption, we further examined to what extent alcohol intake could mediate the effects of rs671 on various cardiovascular risk factors. We showed that most of the associations between rs671 and each of the cardiovascular risk factors were substantially attenuated after adjustment for alcohol intake (**Fig. 3d** and **Supplementary Table 4**).

Associations at loci previously identified in Europeans

We also sought to replicate the genome-wide significant and suggestive associations previously identified in populations

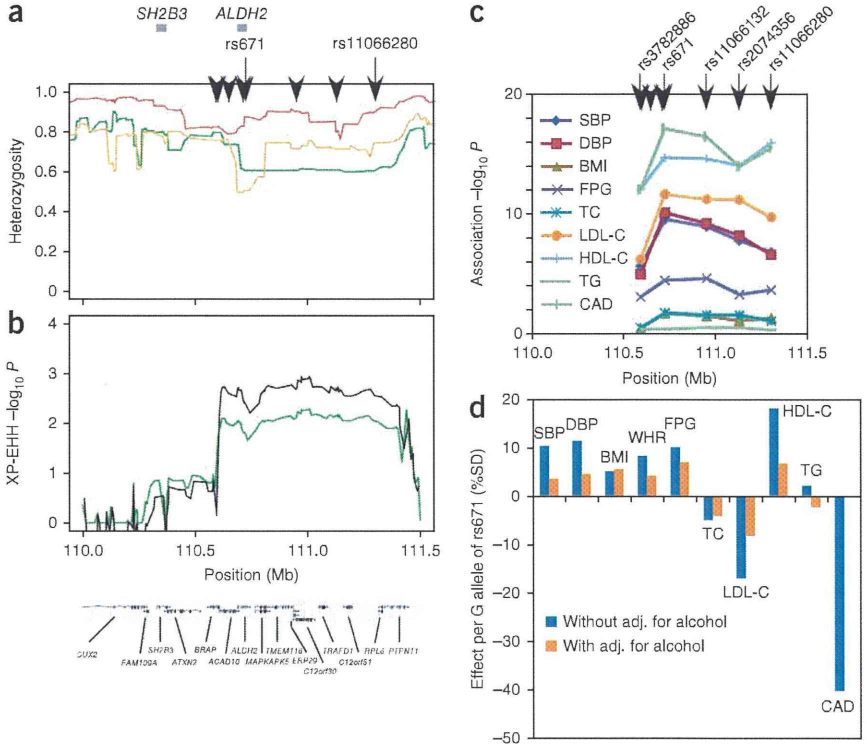


Figure 3 Evidence for positive selection and pleiotropic effects at 12q24.13. (**a**) Lines show heterozygosity calculated in a sliding window of 20 SNPs in the east Asian (green), European (orange) and African (red) populations of the Human Genome Diversity Panel[36]. Heterozygosity in east Asians dropped to 0.61 in a range from 110.7 Mb to 111.4 Mb. Black arrowheads at the top of the plot represent the positions of SNPs forming an east-Asian–specific haplotype with *ALDH2* rs671. The positions of *SH2B3* and *ALDH2* are highlighted by gray bars at the top of the plot. (**b**) Signals of selection detected using a haplotype-based test, XP-EHH[19], in the east Asian population. Because there are multiple SNP clusters showing high LD within a cluster but relatively modest LD between clusters and no ancestral haplotype in east Asians (**Supplementary Fig. 4**), XP-EHH is assumed to provide more appropriate signals of selection than iHS[18] at 12q24.13. Vertical axes represent empirical *P* values, where a suggestive level for positive selection is set at $P < 0.01$ ($-\log_{10} P > 2$). The results are shown for all east Asians combined (green line) and for Japanese (black line). (**c**) The $-\log_{10} P$ for associations with cardio-metabolic traits at five of the eight SNPs forming a common east-Asian–specific haplotype. The position of each SNP is denoted at the top of the plot. (**d**) Per-allele effects of rs671 (G versus A) on cardio-metabolic traits with and without adjustment (adj.) for alcohol intake.

of European descent[9,10] in the AGEN-BP samples (**Fig. 4** and **Supplementary Table 5**). We genotyped variants at seven loci (*CASZ1*, *MTHFR*, *ITGA9*, *FGF5*, *CNNM2-NT5C2*, *ATP2B1* and *CSK-ULK3*) in the CAGE-Amagasaki, CAGE-Fukuoka and CAGE-KING samples and at two loci (*CACNB2* and *PLEKHA7*) in the CAGE-Amagasaki samples as part of the previous replication study[13]. In the present study, we additionally genotyped in the CAGE-Amagasaki samples three (in *C10orf107*, *PLCD3* and *ZNF652*) of seven variants that showed directionally consistent and nominally significant blood pressure associations in the stage 1 study; the remaining four (of seven) variants overlapped with those previously genotyped in the CAGE study samples (**Supplementary Table 2**). When these data were combined with our stage 1 data, six of the seven previously replicated loci[13] and one additional locus (*ZNF652*) still showed directionally consistent associations with blood pressure with *P* values < 0.05.

In the six variants in which we did not detect blood pressure associations in the east Asian populations (**Supplementary Fig. 6**), we compared the effects in our study to those in the follow-up panels in Europeans (to exclude the possible winner's curse effect[21]). We detected cross-population heterogeneity in effect size for *C10orf107* and *PLEKHA7* (**Supplementary Table 5**). One possibility for our failure to detect an association between these six variants and blood pressure in our population is lack of power. To replicate blood pressure associations previously identified in populations of European descent with *P* values < 0.05 in east Asians, sample sizes were not sufficient for three (of the six) variants (*CACNB2*, *C10orf107* and *TBX3-TBX5*; **Supplementary Table 6**). Another possibility could relate to differences in LD patterns between the ethnic groups. We examined LD differences between Europeans (HapMap CEU) and east Asians (HapMap JPT+CHB) at these loci by using the varLD program[22] (**Supplementary Fig. 7**). For the seven reported variants that showed an association with blood pressure in east Asians, there was limited evidence of LD differences between European and east Asian populations at five (of seven) loci that harbor these variants. On the other hand, for the six reported variants for which we did not detect an association with blood pressure in east Asians, we found significant differences in LD between the ethnic groups at four (of six) loci harboring these variants (*CACNB2*, *C10orf107*, *PLEKHA7* and *TBX3-TBX5*).

Cumulative impact of risk alleles and association with hypertension

In total, ten variants showed associations that reached genome-wide significance after all stages of our study. These ten variants had a cumulative impact on SBP and DBP (**Supplementary Note** and **Supplementary Fig. 8**) and were also associated with risk of hypertension (as a dichotomous trait) in directions consistent with the continuous trait effect (**Supplementary Note, Supplementary Fig. 9** and **Supplementary Tables 7** and **8**).

eQTL analysis

In search of putative functional variation at the newly identified loci, we found that, among the six genes on 1p13.2, expression of *ST7L* was associated with expression-associated SNP (eSNP) rs17030613 ($P = 9.4 \times 10^{-15}$ and $P = 1.4 \times 10^{-8}$, reported in two independent studies)[23,24] in lymphoblastoid cell lines (LCLs) derived from samples of European and east Asian ancestry (**Supplementary Fig. 10**). In the current study, the rs17030613 A allele associated with lower *ST7L* transcript levels was also associated with lower blood pressure. We identified no other eSNPs or non-synonymous SNPs in LD with blood-pressure–associated variants in currently available public databases for the five newly discovered association signals.

DISCUSSION

By using a three-stage GWAS meta-analysis of up to 50,373 individuals of east Asian ancestry, we identified five new blood pressure

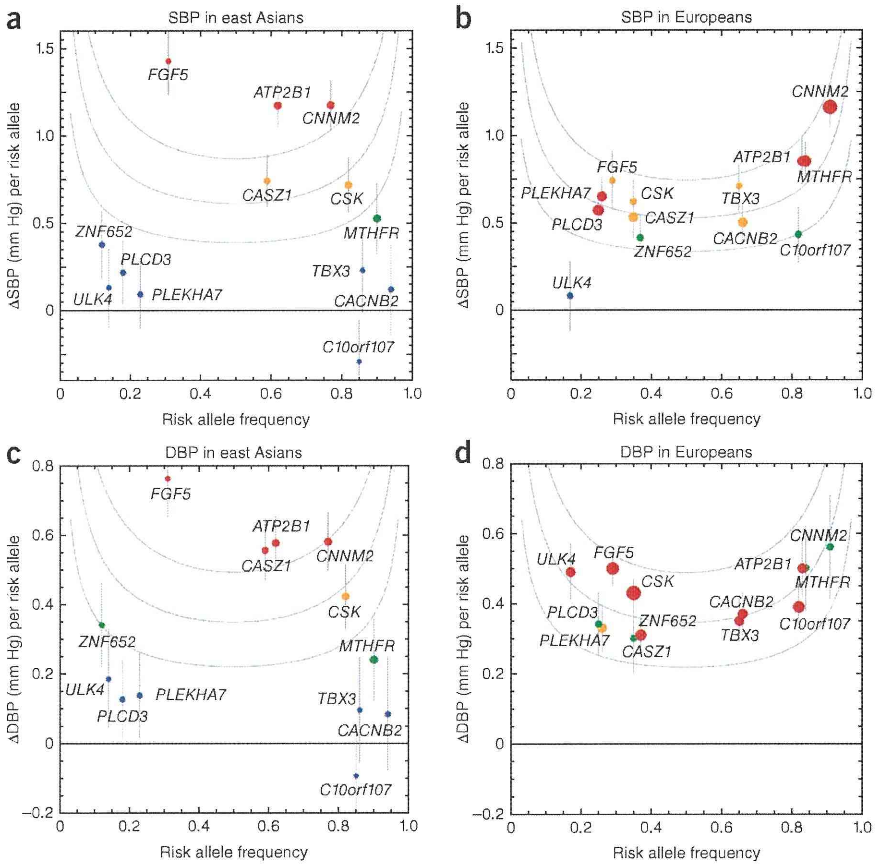


Figure 4 Plots of effect size (β) versus risk-allele frequency of 13 loci previously identified in GWAS meta-analysis of blood pressure in individuals of European descent. The list of 13 loci includes one locus (*CASZ1*) that did not reach genome-wide significance in Europeans but was found to show a significant association with blood pressure in the present AGEN-BP meta-analysis. Plots for SBP (east Asians (**a**) and Europeans (**b**)) and DBP (east Asians (**c**) and Europeans (**d**)) are depicted separately. Each point refers to a single blood pressure association signal, with colors denoting the strength of the blood pressure association (red, $P < 5 \times 10^{-8}$; orange, $5 \times 10^{-8} \leq P < 10^{-5}$; green, $10^{-5} \leq P < 0.05$; blue, $P \geq 0.05$) and with sizes being proportional to the tested sample size. Whiskers are ± standard errors. The gene names associated with each signal have been chosen on the basis of proximity to the lead SNP and should not be presumed to indicate causality. The gray curves represent the coefficient of determination (R^2) (blood pressure variance explained by a SNP (%)); those from the top to the bottom correspond to $R^2 = 0.1$, $R^2 = 0.05$ and $R^2 = 0.02$. For estimating R^2, the standard deviations of blood pressure were assumed to be 19.4 mm Hg (SBP) and 11.0 mm Hg (DBP) in east Asians and 16.6 mm Hg (SBP) and 10.9 mm Hg (DBP) in Europeans[9]. The risk alleles were designated as previously reported in European GWAS meta-analysis. For details, see **Supplementary Table 5**.

association signals with genome-wide significance. These include 1p13.2 in *ST7L-CAPZA1* ($P = 1.2 \times 10^{-8}$ for DBP), 2q24.3 in *FIGN-GRB14* ($P = 3.5 \times 10^{-11}$ for SBP), 4q25 in *ENPEP* ($P = 9.0 \times 10^{-9}$ for DBP) and 5p13.3 near *NPR3* ($P = 1.9 \times 10^{-8}$ for SBP). At 12q24.21 near *TBX3* (a known blood-pressure–associated locus in populations of European descent), we also identified a new blood-pressure–associated variant, illustrating the possible presence of allelic heterogeneity at this locus in relation to blood pressure regulation.

The glutamyl aminopeptidase encoded by *ENPEP* plays a pivotal role in blood pressure regulation by facilitating conversion of angiotensin II, the main effector protein of the renin-angiotensin-aldosterone system, to angiotensin III, and the *Enpep* knockout mouse develops hypertension[25]. The natriuretic peptide receptor C/guanylate cyclase C encoded by *NPR3* is one of three receptor subtypes that mediate specific binding of the natriuretic peptides, which are important in maintaining blood pressure and extracellular fluid volume[12]. At 2q24.3, the lead SNP (rs16849225) is located between *FIGN* and *GRB14*. *FIGN* encodes fidgetin, a member of a family of ATPases associated with diverse cellular activities[26]. *GRB14* encodes the growth factor receptor-bound protein 14 that interacts with insulin receptors and insulin-like growth factor receptors, suggesting a role for GRB14 in signaling pathways that regulate growth and metabolism[27]. A variant close to the *GRB14* locus (rs10195252) has been associated with waist-to-hip ratio (WHR) in populations of European descent[28]. The variant we identified for the blood pressure association in this study (rs16849225) is >600 kb from that identified for the WHR association, with the two SNPs being in linkage equilibrium. Although the newly discovered variant at 1p13.2 is in close proximity to six annotated genes, it is an eSNP for *ST7L*, with consistent results supporting an association between rs17030613 and *ST7L* transcript levels in the LCLs being shown in two independent studies[23,24] in three ethnic groups (HapMap CEU, YRI and JPT+CHB; $P > 0.05$ for testing inter-population per-allele effect difference[23]). *ST7L* (suppression of tumorigenicity 7 like) has been identified by its similarity to the *ST7* tumor suppressor gene found in the 7q31 region known to be deleted and rearranged in a variety of cancers[29], although the function of this gene remains to be determined.

Our study provides important information on the genetic architecture of blood pressure variation and CAD in relation to known loci. In this study, we provided evidence suggesting that the association between variants at 12q24.13 and blood pressure is likely related to a non-synonymous SNP (rs671, p.Glu504Lys) in *ALDH2*. Furthermore, the associations seem to be largely mediated by alcohol intake, a finding that is supported by previous studies showing that this variant determines an individual's tolerance of alcohol intake by altering ALDH2 enzymatic activity[30,31]. Variants at this locus also appear to have undergone natural selection; phylogenetic analysis suggests that rs671 is not responsible for the blood pressure association identified in the overlapping chromosomal region (near *SH2B3*) in populations of European descent[9,10]. Thus, the associations described in our study appear to be specific to east Asians. Moreover, our study highlights the importance of extending studies examining associations with surrogate markers of cardiovascular risk factors (such as blood pressure) to include clinically relevant events (such as CAD) when trying to identify therapeutic targets, especially if the variant of interest could have pleiotropic effects. In this instance, the allele (of rs671) associated with elevated blood pressure was associated with reduced risk of CAD, a finding that is counterintuitive based on what is known about the relationship between blood pressure and CAD. Further analysis suggests that the deleterious effect of the variant on blood pressure was balanced by protective effects on HDL cholesterol and LDL cholesterol, resulting in a net reduction in CAD risk. This is reminiscent of the effects of the CETP inhibitor torcetrapib, which increased HDL cholesterol and decreased LDL cholesterol. Despite these cardioprotective effects, its use was associated with increased risk of CAD and mortality, which may relate to an off-target effect causing blood pressure elevation[32]. Second, on examining blood pressure associations for 13 variants identified in GWAS meta-analyses in individuals of European descent (**Supplementary Table 5**), we found that 7 of the 13 loci (54%) showed nominally significant associations of the reported lead SNPs in east Asians. This is consistent with previous studies showing that findings from populations of European ancestry may be relevant to other ethnic groups[13,33,34]. We also found another locus (*TBX3-TBX5*) that showed some evidence of allelic heterogeneity in relation to blood pressure. These data suggest that, although some inter-population differences may exist in the pathways involved in the blood pressure elevation (or hypertension) between Europeans and east Asians, the majority of pathways are common.

A number of factors could have influenced our ability to detect associations at the remaining six loci in the east Asian populations. First, the sample size for genome-wide exploration (stage 1) was smaller in AGEN-BP (n = 19,608) than in two previous consortia of populations of European descent, the Global BPgen (n = 34,433) and CHARGE (n = 29,136)[9,10], resulting in limited power to detect several of these loci, even using a P value threshold of <0.05 (**Supplementary Table 6**). This limited power also means that additional loci could be identified by enlarging the sample size in the discovery stage, providing motivation for larger meta-analyses in the future. Second, we found some evidence of LD differences between European and east Asian populations at two of seven loci (29%) for which we did detect an association in our study, versus at four of six loci (67%) for which we did not detect an association (**Supplementary Fig. 7**). This could also contribute to our failure to detect associations for some of the known variants in the present study.

In conclusion, although variants identified in European populations are often relevant to other ethnic groups, by conducting a large-scale GWAS meta-analysis in east Asians, we identified previously unreported association signals for blood pressure (at *ST7L-CAPZA1*, *FIGN-GRB14*, *ENPEP* and *NPR3*) that reached genome-wide significance. Near *TBX3*, we also identified some evidence for allelic heterogeneity in east Asians compared to Europeans in relation to blood pressure associations. Further, our data have provided evidence of east-Asian–specific blood pressure association at *ALDH2*, which has pleiotropic effects on other metabolic traits and CAD, highlighting the importance of fine-mapping efforts to pinpoint causal variants and causal genes, thereby providing new insights into the physiology of complex diseases.

URLs. HapMap, http://www.hapmap.org/; 1000 Genomes Project, http://www.1000genomes.org/; METAL, http://www.sph.umich.edu/csg/abecasis/Metal; R, http://www.r-project.org/; GTEx (Genotype-Tissue Expression) eQTL Browser, http://www.ncbi.nlm.nih.gov/gtex/test/GTEX2/gtex.cgi; eQTL.Chicago.edu, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/; Gene Expression Analysis Based on Imputed Genotypes, http://www.sph.umich.edu/csg/liang/imputation/; SNPExpress, http://people.genome.duke.edu/~dg48/SNPExpress/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors acknowledge the essential role of the Asian Genetic Epidemiology Network (AGEN) in developing and supporting this manuscript. AGEN members include the Cardio-metabolic Genome Epidemiology (CAGE) Network, Genetic Epidemiology Network of Salt-Sensitivity (GenSalt), Korean Association Resource (KARE) Project, Shanghai Hypertension Study, Singapore Malay Eye Survey (SiMES), Singapore Prospective Study (SP2) Program, Suita Study and Taiwan Super Control Study.

CAGE: The CAGE Network Studies were supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; the Program for Promotion of Fundamental Studies in Health Sciences, National Institute of Biomedical Innovation Organization (NIBIO); KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas 'Applied Genomics' from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and the Grant of National Center for Global Health and Medicine (NCGM).

GenSalt: The Genetic Epidemiology Network of Salt Sensitivity is supported by research grants (U01HL072507, R01HL087263 and R01HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health (Bethesda, Maryland, USA). T.N.K. is supported partially by Award Number K12HD043451 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (Bethesda, Maryland, USA).

KARE: KARE and HEXA-shared control studies were supported by grants from Korea Centers for Disease Control & Prevention, Republic of Korea (4845-301, 4851-302, 4851-307).

Shanghai: This work was supported by the Chinese National Key Program for Basic Research (grants 973:2004CB518603, 2006CB503804 and 2009CB521905) and Chinese National High Tech Program (grants 863:2009AA022703 and 2006AA02Z179) and the Ministry of Science and Technology, National Natural Science Foundation (30871361).

SiMES: The Singapore Malay Eye Study (SiMES) was funded by the National Medical Research Council (NMRC 0796/2003 and NMRC/STaR/0003/2008) and the Biomedical Research Council (BMRC, 09/1/35/19/616).

SP2: The Singapore Prospective Study Program (SP2) was funded through grants from the Biomedical Research Council of Singapore (BMRC 05/1/36/19/413 and 03/1/27/18/216) and the National Medical Research Council of Singapore (NMRC/1174/2008). Y.Y.T. acknowledges support from the Singapore National Research Foundation (NRF-RF-2010-05). E.S.T. also receives additional support from the National Medical Research Council through a clinician scientist award (NMRC/CSA/008/2009).

Suita/Ehime Study: The Ehime Study was supported by Grants for Scientific Research (Priority Areas 'Medical Genome Science (Millennium Genome Project)' and 'Applied Genomics') from the Ministry of Education, Culture, Sports, Science and Technology, Japan; a Grants-in-Aid (H15-longevity-005, H17-longevity-003, H16-kenko-001, H18-longevity (kokusai)) from the Ministry of Health, Labor and Welfare, Health and Labor Sciences Research Grants, Japan; a Science and Technology Incubation Program in Advanced Regions, Japan Science and Technology Agency; and the Japan Atherosclerosis Prevention Fund.

Taiwan Super Control Study: This study was supported by Academia Sinica Genomic Medicine Multicenter Study; National Research Program for Genomic Medicine, National Science Council, Taiwan (National Clinical Core, NSC97-3112-B-001-014; and National Genotyping Center, NSC97-3112-B-001-015).

AUTHOR CONTRIBUTIONS

Principal investigators: N.K., J.H.

Project coordination leaders: N.K., J.H., Y.T.

Manuscript writing group: N.K., F.T., T.N.K., J.H., Y.Y.T., Y.S.C., E.S.T.

Project data management: T.N.K.

Genotyping and quality control: F.T., M.I., K.Y., Y.T., N.I., Y.K., X.S., W.T.T., Y.Y.T.

Phenotype collection, data management: CAGE: N.K., K.Y., T.K., T.N., M.Y., K.O., Y.Y., E.N., T.S., R.T., S.K., T.O.; **GenSalt:** T.N.K., D.G., J.H.; **KARE:** J.-P.J., S.S.K., Y.S.C.; **Shanghai:** Y.Z., X. Zhang, X. Zhou, D.Z.; **SiMES/SP2:** T.A., T.Y.W., E.S.T.; **Suita:** N.I., Y.K., Y.T., T.M.; **Taiwan:** C.-H.C., L.-c.C., Y.-T.C., J.-Y.W.

Genome-wide genotyping: CAGE: N.K., M.I.; **GenSalt:** J.E.H., Y.J.S.; **KARE:** J.-Y.L., B.-G.H., Y.S.C.; **Shanghai:** W.H.; **SiMES/SP2:** M.S., J.J.L.; **Suita:** N.I.; **Taiwan:** Y.-T.C., J.-Y.W.

Data analysis and data interpretation: N.K., F.T., T.N.K., Y.T., Y.Y.T., E.S.T., M.J.G., Y.J.K., X.S., W.T.T., R.T.H.O., C.-H.C., L.-c.C., C.E.J., J.H.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprints/inxex.html.

1. Ezzati, M. *et al.* Selected major risk factors and global and regional burden of disease. *Lancet* **360**, 1347–1360 (2002).
2. Lopez, A.D. *et al.* Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006).
3. Lawes, C.M.M., Vander Hoorn, S. & Rodgers, A. International Society of Hypertension. Global burden of blood-pressure-related disease, 2001. *Lancet* **371**, 1513–1518 (2008).
4. Kearney, P.M. *et al.* Global burden of hypertension: analysis of worldwide data. *Lancet* **365**, 217–223 (2005).
5. He, J. *et al.* Premature deaths attributable to blood pressure in China: a prospective cohort study. *Lancet* **374**, 1765–1772 (2009).
6. Eastern Stroke and Coronary Heart Disease Collaborative Research Group. Blood pressure, cholesterol, and stroke in eastern Asia. *Lancet* **352**, 1801–1807 (1998).
7. Cowley, A.W. Jr. The genetic dissection of essential hypertension. *Nat. Rev. Genet.* **7**, 829–840 (2006).
8. Kurtz, T.W. Genome-wide association studies will unlock the genetic basis of hypertension: con side of the argument. *Hypertension* **56**, 1021–1025 (2010).
9. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
10. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
11. Dominiczak, A.F. & Munroe, P.B. Genome-wide association studies will unlock the genetic basis of hypertension: pro side of the argument. *Hypertension* **56**, 1017–1020 (2010).
12. Anand-Srivastava, M.B. Natriuretic peptide receptor-C signaling and regulation. *Peptides* **26**, 1044–1059 (2005).
13. Takeuchi, F. *et al.* Blood pressure and hypertension are associated with 7 loci in the Japanese population. *Circulation* **121**, 2302–2309 (2010).
14. Chen, L., Davey Smith, G., Harbord, R.M. & Lewis, S.J. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Med.* **5**, e52 (2008).
15. Tsuchihashi-Makaya, M. *et al.* Gene-environmental interaction regarding alcohol-metabolizing enzymes in the Japanese general population. *Hypertens. Res.* **32**, 207–213 (2009).
16. Li, H. *et al.* Refined geographic distribution of the oriental ALDH2*504Lys (nee 487Lys) variant. *Ann. Hum. Genet.* **73**, 335–345 (2009).
17. Takeuchi, F. *et al.* Confirmation of *ALDH2* as a major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in Japanese. *Circ. J.* **75**, 911–918 (2011).
18. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
19. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
20. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182–1190 (2009).
21. Zollner, S. & Pritchard, J.K. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* **80**, 605–615 (2007).
22. Ong, R.T. & Teo, Y.Y. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinformatics* **26**, 1269–1270 (2010).
23. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
24. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
25. Mizutani, S. *et al.* New insights into the importance of aminopeptidase A in hypertension. *Heart Fail. Rev.* **13**, 273–284 (2008).
26. Cox, G.A., Mahaffey, C.L., Nystuen, A., Letts, V.A. & Frankel, W.N. The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development. *Nat. Genet.* **26**, 198–202 (2000).
27. Goenaga, D. *et al.* Molecular determinants of Grb14-mediated inhibition of insulin signaling. *Mol. Endocrinol.* **23**, 1043–1051 (2009).
28. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
29. Katoh, M. Molecular cloning and characterization of *ST7R* (*ST7*-like, *ST7L*) on human chromosome 1p13, a novel gene homologous to tumor suppressor gene *ST7* on human chromosome 7q31. *Int. J. Oncol.* **20**, 1247–1253 (2002).
30. Brooks, P.J., Enoch, M.A., Goldman, D., Li, T.K. & Yokoyama, A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. *PLoS Med.* **6**, e50 (2009).
31. Harada, S., Misawa, S., Agarwal, D.P. & Goedde, H.W. Liver alcohol dehydrogenase and aldehyde dehydrogenase in the Japanese: isozyme variation and its possible role in alcohol intoxication. *Am. J. Hum. Genet.* **32**, 8–15 (1980).
32. Barter, P.J. *et al.* Effects of torcetrapib in patients at high risk for coronary events. *N. Engl. J. Med.* **357**, 2109–2122 (2007).
33. Tabara, Y. *et al.* Common variants in the *ATP2B1* gene are associated with susceptibility to hypertension: the Japanese Millennium Genome Project. *Hypertension* **56**, 973–980 (2010).
34. Hong, K.W. *et al.* Recapitulation of two genomewide association studies on blood pressure and essential hypertension in the Korean population. *J. Hum. Genet.* **55**, 336–341 (2010).
35. Johnson, A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
36. Pickrell, J.K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).

[1]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan. [2]Department of Basic Medical Research and Education, Ehime University Graduate School of Medicine, Toon, Japan. [3]Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana, USA. [4]Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, Korea. [5]Centre for Molecular Epidemiology, National University of Singapore, Singapore. [6]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore. [7]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [8]School of Chinese Medicine, China Medical University, Taichung, Taiwan. [9]State Key Laboratory of Medical Genetics, Shanghai Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [10]Shanghai Institute of Hypertension, Shanghai, China. [11]Division of Genome Analysis, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan. [12]Department of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, Japan. [13]Department of Genome Science, Aichi-Gakuin University, School of Dentistry, Nagoya, Japan. [14]National University of Singapore Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore. [15]Department of Functional Pathology, Shimane University School of Medicine, Izumo, Japan. [16]Cardiovascular Institute and Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [17]Chinese National Center for Cardiovascular Diseases, Beijing, China. [18]Department of Preventive Cardiology, National Cerebral and Cardiovascular Center, Suita, Japan. [19]Department of Genetics, Chinese National Human Genomic Center, Shanghai, China. [20]Department of Geriatric Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [21]Mukogawa Women's University Institute for World Health Development, Nishinomiya, Japan. [22]Division of Endocrinology and Diabetes, Department of Internal Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. [23]Department of Diabetes and Endocrinology, Chubu Rosai Hospital, Nagoya, Japan. [24]Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA. [25]Institute of Human Genetics, University of California, San Francisco, California, USA. [26]Department of Epidemiology, University of Texas School of Public Health, Houston, Texas, USA. [27]Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, Japan. [28]Key Laboratory of Bioinformatics, Department of Automation, Tsinghua University, Beijing, China. [29]Institute for Adult Diseases, Asahi Life Foundation, Tokyo, Japan. [30]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore. [31]Department of Medicine and Bioregulatory Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [32]Department of Ophthalmology, National University of Singapore, Singapore. [33]Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. [34]Center for Eye Research Australia, University of Melbourne, Melbourne, Australia. [35]Director, Shimane University Hospital, Izumo, Japan. [36]Department of Geriatric Medicine and Nephrology, Osaka University Graduate School of Medicine, Suita, Japan. [37]Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Suita, Japan. [38]Department of Statistics and Applied Probability, National University of Singapore, Singapore. [39]Department of Epidemiology and Public Health, National University of Singapore, Singapore. [40]Department of Medicine, National University of Singapore, Singapore. [41]Department of Medicine, Tulane University School of Medicine, New Orleans, Louisiana, USA. [42]These authors contributed equally to this work. [43]These authors jointly directed this work. Correspondence should be addressed to N.K. (nokato@ri.ncgm.go.jp).

ONLINE METHODS

Stage 1 samples. The Asian Genetic Epidemiology Network (AGEN) is a consortium of genetic epidemiology studies of cardiovascular disease-related phenotypes, such as blood pressure (or hypertension), diabetes and obesity, conducted among Asian populations. AGEN-BP consists of 19,608 east Asian participants who underwent standardized collection of blood pressure measurements in eight population- and family-based GWAS including the Cardio-metabolic Genome Epidemiology (CAGE) Network, Genetic Epidemiology Network of Salt-Sensitivity (GenSalt), Korean Association Resource (KARE) Project, Shanghai Hypertension Study, Singapore Malay Eye Survey (SiMES), Singapore Prospective Study (SP2) Program, Suita Study and Taiwan Super Control Study. Each study established a consensus on phenotype harmonization and analytical plan for the within-study GWAS and meta-analysis of results across studies. Each study received approval from the institutional review boards of local research institutions, and all participants in each study provided written informed consent for participation in the study. A detailed description of the study design and phenotype measurement for each study (or cohort) is provided in the **Supplementary Note** and in **Supplementary Table 1**. The overall design of the three-staged GWAS meta-analysis is depicted in **Supplementary Figure 1**. Stage 1 was a meta-analysis of directly genotyped and imputed SNPs from individuals of east Asian descent drawn from population-based or control samples in case-control studies in AGEN-BP.

Genome-wide genotyping and quality control. Genotyping arrays and quality control filters applied to the individual studies are provided in **Supplementary Table 1**.

Genotype imputation. Imputation of genotypes to the HapMap Phase 2 (JPT+CHB except for SiMES, which used HapMap JPT+CHB+CEU+YRI) set was carried out using MACH[37], IMPUTE[38] or BEAGLE[39] with preimputation filters as specified in **Supplementary Table 1**. Imputation results are summarized as an 'allele dosage' defined as the expected number of copies of the coded allele at that SNP (a fractional value between 0 and 2) for each genotype. In total, up to 2.4 million genotyped or imputed autosomal SNPs were analyzed.

Phenotype modeling and SNP association analysis. For participants taking antihypertensive therapies, blood pressure was imputed by adding 10 mm Hg and 5 mm Hg to SBP and DBP values, respectively[40]. Within each study, continuous SBP and DBP were adjusted for age, age squared, sex, BMI and any study-specific covariates in linear regression models.

In secondary and confirmatory analyses of hypertension, hypertensive cases were defined as follows: (i) SBP ≥ 160 mm Hg and/or DBP ≥ 100 mm Hg for untreated subjects; (ii) individuals receiving chronic antihypertensive treatments; and (iii) age of onset ≤ 65 years. Normotensive controls were defined as follows: (i) SBP < 130 mm Hg and DBP < 85 mm Hg without antihypertensive treatments and (ii) age ≥ 50 years. Within each study, a dichotomous trait of hypertension status was adjusted for age, age squared, sex and BMI in logistic regression models, except for the CAGE samples used for the follow-up and replication (3,294 cases and 6,831 controls; **Supplementary Table 7**), in which age was not adjusted for because of its relation to case-control status (the mean age was younger in cases than in controls).

Stage 1 meta-analysis. All study-specific effect estimates and coded alleles were oriented to the forward strand of the NCBI36 reference sequence of the human genome. If a SNP from a study did not meet quality standards, it was reported as missing from that study for the purpose of meta-analysis. Results for this SNP were pooled among the other contributing studies. SNPs were excluded if they had study-specific imputation quality $R^2 < 0.5$. Genomic control[41] was carried out on study-specific test statistics: genomic control lambda (λ_{GC}) estimates are given in **Supplementary Table 1**. We used an inverse-variance–weighted meta-analysis to combine association results for stage 1 with METAL software (see URLs). Evidence for the heterogeneity of the effect sizes was investigated using Cochran's Q statistic[42]. We retained 1.7 million SNPs with MAF > 0.05 and an effective sample size of >10,000. In **Figure 1**, we show the Manhattan plots drawn using the WGAViewer software[43].

SNP prioritization for stage 2. We selected 13 SNPs for follow up in stage 2 using two methods in parallel. Using the first method, the most strongly associated SNP was chosen from each of 12 distinct regions containing ≥1 SNP with $P < 1 \times 10^{-5}$ for SBP and/or DBP based on the stage 1 data, in which samples from ≥7 (of 8) GWAS were available for the meta-analysis, in principle, to decrease the potential noise of false positive associations caused by the small sample size. We excluded SNPs within 50 kb of the lead SNPs for blood pressure associations previously identified in Europeans[9,10] at *CNNM2-NT5C2*, *ATP2B1* and *CSK-ULK3*. Although attaining a significance level of $P < 1 \times 10^{-5}$ for SBP, we did not subject two SNPs (rs4671977 and rs12547784) to follow up in stage 2 because their association with the DBP trait was not significant ($P > 0.05$) in stage 1. Using the second method, we generated a list of the next most significant SNPs (21 unique loci showed $P \le 1 \times 10^{-3}$ for both SBP and DBP and consistent association signals across the studies). We selected a SNP at 5p13.3 *ad hoc* because of the physiological candidacy of the nearby gene (*NPR3*)[12].

Stage 2 samples and genotyping. We genotyped 13 SNPs in 10,518 individuals of Japanese descent from three studies (CAGE-Amagasaki Study, Suita (2) Study and Ehime Study). Summary characteristics are shown in **Table 1**. Study information and genotyping methods are provided in the **Supplementary Note**.

Stage 3 samples, genotyping and joint analysis. For SNP selection in the stage 3 study, we set a threshold of $P < 5 \times 10^{-6}$ for joint analysis of stages 1 and 2. Six of 13 SNPs exceeded this significance level and were followed up in 20,247 individuals of east Asian descent from three studies by *de novo* genotyping (for Fukuoka Study and Kita-Nagoya Genomic Epidemiology (KING) Study of the CAGE Network) and *in silico* replication (for HEXA-shared control).

We carried out meta-analysis of stages 1, 2 and 3 results and considered associations genome-wide significant if they attained $P < 5 \times 10^{-8}$. Association results for multiple stages (1 and 2 or 1, 2 and 3) were combined using inverse variance weighting with the rmeta package of the R software (see URLs).

Replication of previously reported loci. Along with genome-wide exploration of new loci, we examined blood pressure associations at 27 loci reported by the European GWAS meta-analyses[9,10]: 13 genome-wide significant loci and 14 loci with suggestive association (**Supplementary Table 5**). A one-tailed $P < 0.05$ (two-tailed $P < 0.1$) was considered statistically significant for the 13 loci previously shown to have genome-wide significant ($P < 5 \times 10^{-8}$) associations in Europeans; for an association to be considered significant, it had to involve the same risk allele as that reported in Europeans and was accordingly assessed with the one-tailed test. However, two-tailed P values are presented throughout the text for readability. For the 14 loci with suggestive association in the original European studies[9,10], a one-tailed $P < 0.05/14 = 0.0036$ was considered statistically significant after Bonferroni correction.

Haplotype phylogeny and positive selection at 12q24.13. Previous studies in populations of European descent claimed significant evidence supporting blood pressure associations[9,10] and signatures of natural selection[20] in a 1.6-Mb interval at 12q24.13, near which the present study in east Asians also identified strong blood pressure associations. To clarify phylogenic differences in susceptibility variants between the two populations, we inferred the haplotypes and constructed their phylogeny across three HapMap panels (JPT, CEU and YRI) in the 12q24.13 region of interest. Further, using haplotype-based tests, iHS[18] and XP-EHH[19], we tested the hypothesis that a long-range, evolutionarily derived haplotype, upon which blood-pressure–decreasing alleles could lie, arose from a positive selection in east Asians independently of Europeans.

Pleiotropic effects on cardio-metabolic traits. We tested SNP-trait associations of five SNPs located in the target interval at 12q24.13 for risk factors of cardiovascular disease and CAD. The associations with SBP and DBP, BMI, fasting plasma glucose and lipids were analyzed in the CAGE-Amagasaki Study (for five SNPs) and CAGE-Fukuoka Study (for rs11066280 and rs671 alone) samples with and without adjustment for alcohol intake, where the amount of alcohol consumed was denoted in terms of servings of sake (1 gou (180 ml), the traditional Japanese unit for cup size) of Japanese rice wine is considered equal to 22 g of ethanol)[17]. The association with CAD was analyzed in the CAD case-control panel (1,347 CAD cases and 1,337 controls) derived from the CAGE Network (F. Takeuchi *et al.*, unpublished).

eQTL analysis. Using publicly available data (see URLs), we examined the *cis* associations (defined as genes within 1 Mb) between each of the five newly discovered blood pressure SNPs and expression of nearby genes in a variety of cells/tissues, for example, LCLs, monocytes, fibroblasts, liver and brain tissues. After initial screening, we narrowed the targets of eQTLs down to LCLs in two databases[23,24]. Briefly, LCLs, derived from peripheral blood lymphocytes, were available for 209 HapMap samples (comprising 60 CEU, 60 YRI and 89 JPT+CHB individuals)[23] and 378 individuals of European descent[24]. The expression of 47,294 (for 209 HapMap samples) and 54,675 (for 378 Europeans) transcripts was assessed for each individual, whose genotypes were also imputed using the corresponding HapMap dataset. SNPs were tested for *cis* associations, assuming an additive genetic model, adjusting for non-genetic effects in the gene expression value.

37. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
38. Marchini, J. *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
39. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
40. Cui, J.S., Hopper, J.L. & Harrap, S.B. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension* **41**, 207–210 (2003).
41. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
42. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
43. Ge, D. *et al.* WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.* **18**, 640–643 (2008).

Detection of common single nucleotide polymorphisms synthesizing quantitative trait association of rarer causal variants

Fumihiko Takeuchi,[1,2,6] Shotai Kobayashi,[3] Toshio Ogihara,[4] Akihiro Fujioka,[5] and Norihiro Kato[1]

[1]Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo 162-8655, Japan; [2]Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo 162-8640, Japan; [3]Shimane University Hospital, Izumo 693-8501, Japan; [4]Department of Geriatric Medicine and Nephrology, Osaka University Graduate School of Medicine, Suita 565-0871, Japan; [5]Amagasaki Health Medical Foundation, Amagasaki 661-0012, Japan

Genome-wide association (GWA) studies have identified hundreds of common (minor allele frequency ≥5%) single nucleotide polymorphisms (SNPs) associated with phenotype traits or diseases, yet causal variants accounting for the association signals have rarely been determined. A question then raised is whether a GWA signal represents an "indirect association" as a proxy of a strongly correlated causal variant with similar frequency, or a "synthetic association" of one or more rarer causal variants in linkage disequilibrium ($D' \approx 1$, but r^2 not large); answering the question generally requires extensive resequencing and association analysis. Instead, we propose to test statistically whether a quantitative trait (QT) association of an SNP represents a synthetic association or not by inspecting the QT distribution at each genotype, not requiring the causal variant(s) to be known. We devised two test statistics and assessed the power by mathematical analysis and simulation. Testing the heterogeneity of variance was powerful when low-frequency causal alleles are linked mostly to one SNP allele, while testing the skewness outperformed when the causal alleles are linked evenly to either of the SNP alleles. By testing a statistic combining these two in 5000 individuals, we could detect synthetic association of a GWA signal when causal alleles sum up to 3% in frequency. Such signal only partially explains the heritability contributed by the whole locus. The proposed test is useful for designing fine mapping after studying association of common SNPs exhaustively; we can prioritize which GWA signal and which individuals to be resequenced, and identify the causal variants efficiently.

[Supplemental material is available for this article. The synthetic association test software is freely available at http://www.fumihiko.takeuchi.name/PUBLICATIONS/synthetic.R.]

Genome-wide association (GWA) studies have identified hundreds of common (minor allele frequency [MAF] ≥5%) single nucleotide polymorphisms (SNPs) associated with a few hundred traits or diseases, yet the associated SNPs and their proxies mostly do not show evident function related to the target trait, and eventual identification of causal variants accounting for GWA signals has been challenging (Wellcome Trust Case Control Consortium 2007; McCarthy et al. 2008). A question that is then raised is whether a common SNP identified in a GWA study represents an "indirect association" as a proxy of a strongly correlated causal variant with similar frequency, or a "synthetic association" of one or more rarer causal variants that are in linkage disequilibrium (LD) ($D' \approx 1$, but r^2 not large) with the common SNP (Cirulli and Goldstein 2010; Dickson et al. 2010).

Synthetic association accounted for GWA signals in several studies. In a GWA study for dose of anticoagulant drug warfarin, the strongest association signal in the *CYP2C9* gene was observed at an SNP rs4917639, whose minor allele (frequency 18%) is a composite of two functional alleles *CYP2C9*2* (rs1799853, frequency 11%) and *CYP2C9*3* (rs1057910, frequency 7%) (Wadelius et al. 2007; Takeuchi et al. 2009). In a GWA study for anemia in patients treated for chronic hepatitis, the strongest signal was observed at an SNP rs6051702 in *C20orf194*, whose minor allele (frequency 19%) is almost exactly a composite of two causal variants in the neighboring *ITPA* gene (frequency 8% and 12%) (Fellay et al. 2010). When there are many rare causal variants, but no common one, as in the *HBB* gene for sickle cell anemia or the *GJB2/GJB6* locus for hearing loss, the association of common SNPs detected in GWA studies were attributable to the rare variants (Dickson et al. 2010). Using simulations, Dickson and colleagues showed that synthetic association is likely to occur when there are multiple rare variants in a locus (Dickson et al. 2010).

In general, identification of the causal variants accounting for a synthetic association requires extensive resequencing and association analysis. Instead, here we propose to test statistically whether a quantitative trait (QT) association of an SNP represents a synthetic association or not by inspecting only the QT distribution at each genotype of the SNP, without a priori knowledge about rarer causal variants. We focus on two statistics of the QT distribution: the heterogeneity of variance (i.e., heteroscedasticity) among SNP genotypes and the skewness. The statistical tests were examined in real data of the apolipoprotein E (*APOE*) gene, and in simulated data for representative models of synthetic association. Moreover, we formulated a general mathematical model of synthetic association, and assessed the test statistics theoretically. The two statistics were suitable for complementary scenarios: Heteroscedasticity was more sensitive than skewness when low-frequency

[6]**Corresponding author.**
E-mail fumihiko@takeuchi.name.

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.115832.110.

21:1122–1130 © 2011 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/11; www.genome.org

(<5%) causal alleles were linked mostly to one SNP allele, while skewness outperformed when the causal alleles were linked in balance to either of the two SNP alleles. We thus devised a test combining the two statistics, which was powerful for any of the assumed models.

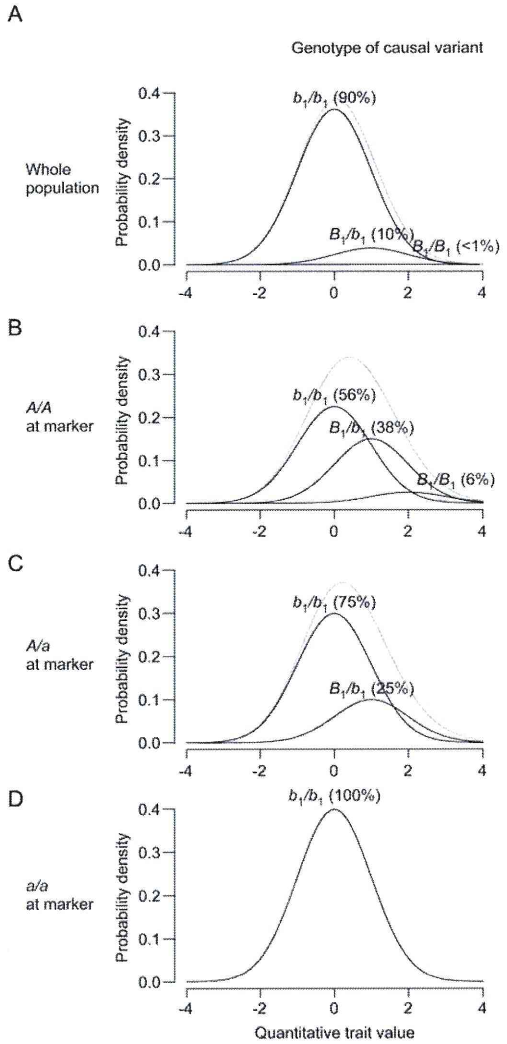


Figure 1. Probability distribution of the QT value within subgroups classified by marker SNP genotypes. (*A*) In the whole population, the total QT distribution (gray curve) comprises a mixture of normal distributions (black curves) with unit variance and the mean 0, 1, or 2, which correspond to genotypes b_1/b_1, B_1/b_1, and B_1/B_1 at the causal variant. As genotype B_1/B_1 is rare (0.25%), the corresponding curve appears flat. (*B*) QT distribution among individuals with *A/A* genotype at the marker. As B_1/B_1 and B_1/b_1 genotypes are enriched in this subgroup due to LD, the variance is enlarged, as noticeable from the lower peak and wider distribution of the gray curve. (*C*) Individuals with the *A/a* genotype have either genotypes b_1/b_1 or B_1/b_1, and the QT variance is moderately enlarged. (*D*) All individuals with *a/a* genotype at the marker have b_1/b_1 genotype at the causal variant. The QT variance is 1.10 in *A*, 1.38 in *B*, 1.19 in *C*, and 1 in *D*.

Results

Test of heteroscedasticity

We first show a schematic example of synthetic association and illustrate how QT variance can differ among individuals classified by marker SNP genotypes. We assume a common marker SNP with alleles *A* and *a*, and a single causal variant with alleles B_1 and b_1. The allele B_1 (5% in frequency) is always linked to allele *A* (20% in frequency); thus, existing haplotype classes are AB_1, Ab_1, and ab_1. We assume the QT is normally distributed with the unit variance and the mean equal to 2, 1, and 0 within a subgroup of individuals having genotype B_1/B_1, B_1/b_1, and b_1/b_1, respectively. The QT distribution in the whole population becomes a mixture of the normal distributions combined according to the frequency of genotypes B_1/B_1, B_1/b_1, and b_1/b_1 (Fig. 1A). Individuals with *A/A* genotype at the marker SNP are enriched with the genotypes of B_1/B_1 and B_1/b_1 at the causal variant, thus their QT distribution widens (Fig. 1B). On the contrary, individuals with *a/a* genotype at the marker all have b_1/b_1 genotype at the causal variant, and their QT variance equals one (Fig. 1D). The QT variance is the largest in the subgroup with *A/A* genotype, which is linked more frequently to the low-frequency causal allele B_1, and the smallest in the subgroup with *a/a* genotype. Indeed, QT variance among individuals of a specific marker genotype enlarges proportionally to two factors: the variance of the causal genotype within the subgroup, and the squared effect-size of the causal allele (equation M2). The low-frequency causal variant causes the synthetic association of the marker SNP, and the heteroscedasticity of QT distribution among the marker genotypes.

We could exemplify the detection of synthetic association using heteroscedasticity in the *APOE* gene, which is known to associate with LDL cholesterol (LDL-C) level through three classical isoforms coded by two functional (or causal) variants—rs7412 (Arg158Cys) and rs429358 (Cys112Arg). As compared with E3 (the most common isoform), E2 (coded by rs7412) and E4 (coded by rs429358) decreased and increased the LDL-C level, respectively (Weisgraber et al. 1981; Weisgraber 1994; Bennet et al. 2007). The two variants had MAF <10% (in Europeans and East Asians) and were not included in SNP chips of GWA scan (except for the recent ones containing >1 million SNPs). In a GWA study for lipids in 1210 Japanese (F Takeuchi, et al., in prep.), we initially found four SNPs near *APOE* to attain locus-wise significant *P*-values for LDL-C association, although any of these were not significant after adjustment for the two functional variants. When only the chip SNPs were analyzed, rs405509 and rs377702 showed statistically independent signals of association (Supplemental Fig. 1). In a larger panel of 4840 individuals, the association signals remained at the two chip SNPs, and heteroscedasticity was significant for rs405509 (*P* = 0.019) (Table 1). Indeed, the causal minor alleles of rs7412 (T) and rs429358 (C) were linked to alternate alleles of rs405509 (C and A, respectively), demonstrating synthetic association (Fig. 2). The two causal variants could simultaneously enlarge the QT variance at all three genotypes of rs405509, and consequently, diminish heteroscedasticity (equation M5). However, in this case, as the effect-size of rs7412 was much larger than that of rs429358, heteroscedasticity remained detectable; rs7412 enlarged the QT variance at C/C genotype of rs405509 to 1.182, whereas rs429358 kept the QT variance at A/A genotype at 0.978. On the other hand, the heteroscedasticity of rs377702 did not reach statistical significance due to its recombination with rs7412 (*D'* = 0.34). Thus, even if we identified the association signals at rs405509 and rs377702 via the GWA scan, by detecting heteroscedasticity we could notice the presence of synthetic association and the necessity to search for variants not on the chip.

We next estimated the power to detect synthetic association at an SNP that could be identified in a GWA study. We assumed

Table 1. Testing heteroscedasticity of SNPs in the *APOE* locus associated with LDL-C

SNP	Genotype	Number of individuals	Distribution of LDL-C level		Association with LDL-C level			Heteroscedasticity
			Mean	Variance	Beta	*P*-value	R^2	*P*-value
rs405509	C/C	462	−0.153	1.182	−0.117	1.0×10^{-7}	0.006	0.019
(GWAS SNP)	C/A	2035	−0.050	0.976				
	A/A	2343	0.073	0.978				
rs377702	T/T	32	−0.487	1.231	−0.191	5.1×10^{-7}	0.005	0.583
(GWAS SNP)	T/C	677	−0.149	1.025				
	C/C	4131	0.028	0.991				
rs7412	T/T	12	−1.302	1.079	−0.651	2.0×10^{-44}	0.040	0.92
(causal variant)	T/C	452	−0.584	0.981				
	C/C	4376	0.064	0.960				
rs429358	T/T	3954	−0.042	0.987	−0.212	1.4×10^{-9}	0.008	0.73
(causal variant)	T/C	850	0.185	1.023				
	C/C	36	0.214	1.104				

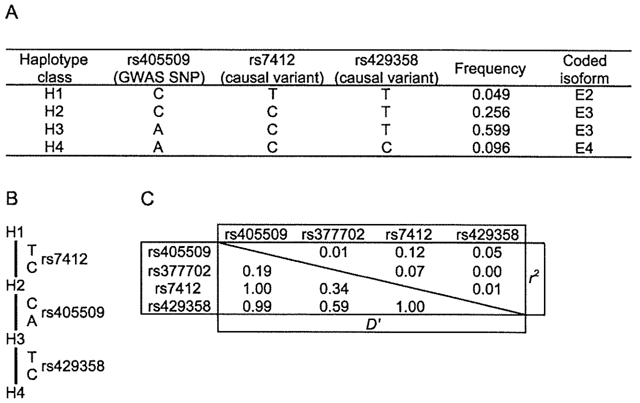
We first adjusted LDL-C level for body mass index and categories by sex and age (≤40, 41–50, 51–60, ≥61 yr) and then applied rank-based inverse normal transformation. Individuals under lipid treatment were excluded. Data are shown for 4840 individuals with complete observation from the Amagasaki study in Takeuchi et al. (2010).

that the marker SNP has MAF ≥5%, and that the proportion of QT variance explained by the marker is $R^2_{mrk} = 0.00592$, a borderline level to attain genome-wide significance (see Supplemental Notes). Figure 3 illustrates the statistical power for detecting heteroscedasticity in 5000 individuals. We examined four representative models of synthetic association by simulation. Under Model 1, there are l causal variants with alleles B_1 and b_1, B_2 and b_2, up to B_l and b_l, and the low-frequency causal alleles B_i have a uniform effect (e.g., increase QT) and are all linked to marker allele A. The QT variance enlarges for individuals with A/A genotype at the marker since they carry various numbers of the causal alleles, whereas individuals with a/a genotype at the marker carry none. Heteroscedasticity of the marker was detectable (power >0.8) in the region marked with an asterisk: For example, when the A allele frequency, $p_A \geq 45\%$, or alternatively when $p_A = 25\%$ and the cumulative frequency of causal alleles is <3%. For a fixed value of p_A, the power for detecting heteroscedasticity increases as the cumulative frequency of causal alleles decreases. When p_A becomes small, the detectable range narrows; the highest cumulative frequency in the detectable range changes proportionally to $\sqrt{p_A/(1-p_A)}$, as estimated in equation M12.

We next examine Models 2–4, where both of the marker alleles are loaded with low-frequency causal alleles. In addition to l causal variants with alleles B_1 and b_1, B_2 and b_2, up to B_l and b_l, there are m other causal variants with alleles C_1 and c_1, C_2 and c_2, up to C_m and c_m, and we designate the low-frequency alleles B_i and c_j as causal. The two groups of causal alleles, B_i and c_j, affect the QT in opposing directions and are linked to alternate alleles A and a of the marker, respectively, and thus synthetically generate the marker association. The QT variance at marker genotype A/A enlarges due to the causal alleles B_i, and the variance at marker genotype a/a enlarges due to the causal alleles c_j (equation M3). Indeed, the variances for all marker genotypes increase and become less heterogeneous than under Model 1. Under Model 2, there is exact balance in effect-size and cumulative frequency between the two groups of causal alleles. The heteroscedasticity disappears if $p_A = 50\%$ (equation M5), and became undetectably weak around the frequency (Fig. 3). The heteroscedasticity was detectable when p_A is close to 5% or 95%: For example, when $p_A = 15\%$ or 85% and the cumulative frequency of the causal alleles B_i, which equals the cumulative frequency of c_j, is <1%. Under Models 3 and 4, where the causal alleles B_i and c_j are not balanced, heteroscedasticity still disappeared, but around a different marker allele frequency. Under Model 3, the effect-size of the causal variants is uniform, yet the cumulative frequency of alleles c_j is half that of alleles B_i, and under Model 4, the cumulative frequencies are identical, yet the effect-size of alleles C_j is half that of alleles B_i. Heteroscedasticity was undetectable around $p_A = 65\%$ and 80% under Models 3 and 4, respectively. At $p_A = 25\%$ heteroscedasticity was detectable when the cumulative frequency of B_i alleles was <2%.

Test of skewness

As the test of heteroscedasticity could not detect synthetic association at a certain marker allele frequency around $p_A = 50\%$, when both alleles of the marker were loaded with low-frequency causal alleles (Fig. 3, Models 2–4) we introduced the test of skewness to cope with such a case. We observed that synthetic association skews the QT distribution at the marker genotypes A/A and a/a oppositely (equation M7): QT distribution among individuals with marker genotype A/A is skewed toward the effect direction of causal alleles B_i, and the QT distribution at genotype a/a is skewed toward the opposite direction, which is the effect direction of causal alleles



Haplotype class	rs405509 (GWAS SNP)	rs7412 (causal variant)	rs429358 (causal variant)	Frequency	Coded isoform
H1	C	T	T	0.049	E2
H2	C	C	T	0.256	E3
H3	A	C	T	0.599	E3
H4	A	C	C	0.096	E4

	rs405509	rs377702	rs7412	rs429358
rs405509		0.01	0.12	0.05
rs377702	0.19		0.07	0.00
rs7412	1.00	0.34		0.01
rs429358	0.99	0.59	1.00	

Figure 2. Haplotype classes (*A*), their phylogeny (*B*) for the marker SNP rs405509 showing synthetic association of functional variants rs7412 and rs429358 in the *APOE* locus. LD coefficients between the SNPs associated with LDL-C (*C*). Haplotype frequencies were calculated using the PLINK software (Purcell et al. 2007).

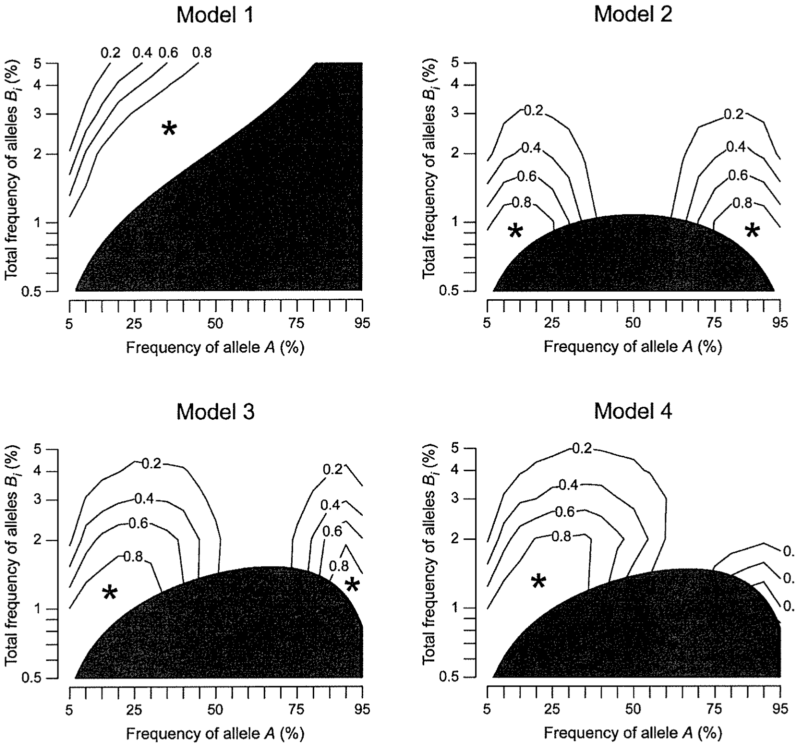


Figure 3. Power for detecting synthetic association by testing heteroscedasticity. The power was computed from simulation under four representative genetic models of synthetic association (see Methods), assuming the strength of marker association (R^2_{mrk}) of 0.00592. Horizontal and vertical axes represent the frequency of the marker allele A, and the cumulative frequency of causal alleles B_i (linked to allele A), respectively. The asterisk indicates the region where synthetic association is detectable with power >0.8. The black region of the parameter space should be neglected, as it does not include causal variants accounting for the marker association.

c_j. Thus, we added the skewness test statistics for the two genotypes, taking the direction into account (equation M8). Accordingly, under Model 2, the test of skewness could detect synthetic association around p_A = 50% (Fig. 4). The detectable range with regard to the cumulative frequency of causal alleles B_i is wide (extends up to 2%) when p_A is around 50% and is narrow when near 5% or 95%; as estimated in equation M14, the maximum detectable cumulative frequency is proportional to $(p_A(1-p_A))^{\frac{3}{4}}$. On the other hand, the skewness test was less powerful than the heteroscedasticity test when all causal alleles are linked to one marker allele (Model 1 in Fig. 4 vs. Fig. 3).

Combined test

Between the two tests to detect synthetic association, the test of heteroscedasticity was more powerful when one marker allele was loaded with the causal alleles (Model 1), and the test of skewness was more powerful when both of the marker alleles were loaded with a balanced amount of causal alleles (Model 2). By combining the two tests (equation M10), we devised the third test that was powerful under all of the models (Fig. 5). The detectable range for the cumulative frequency of B_i alleles exceeds 1% when the causal variants are exactly balanced (Model 2), and is up to 2% otherwise. Overall, we could detect synthetic association if the cumulative frequency of all causal alleles, B_i and c_j altogether, is <3%.

The power to detect synthetic association is influenced by the strength of marker SNP association and sample size. So far, we studied association at a borderline level of genome-wide significance, which is much weaker than some reported SNPs, for example, of lipid traits (Chasman et al. 2009). When the strength of association is doubled to R^2_{mrk} = 0.0118, synthetic association could be detected if the cumulative frequency of all causal alleles is <6%, in a wider range (Supplemental Fig. 2). When the sample size is halved to 2500 individuals, the detectable region narrowed (Supplemental Fig. 3), because the χ^2 statistics of the tests are proportional to the sample size (equations M5 and M9).

Discussion

As seen through mathematical analysis and simulation, we could detect an SNP representing synthetic association of rarer causal variant(s) by testing heteroscedasticity and skewness. The test only requires the genotype–phenotype data obtained in association studies, and the causal variants can be unknown. Whereas previous studies of synthetic association were based on empirical results and simulation (Dickson et al. 2010), we introduced a general mathematical formulation (see Methods) and estimated the variance and skewness of the marker SNP. We also performed computer simulations under representative models of synthetic association and obtained concordant results. The test of heteroscedasticity outperformed the test of skewness when low-frequency causal alleles were linked mostly to one SNP allele, while the test of skewness was better when the causal alleles were linked in balance to either of the two SNP alleles. The test combining the two could detect synthetic association if the cumulative frequency of causal alleles is <3% when tested in 5000 individuals for a marker SNP associated with QT at a borderline level of genome-wide significance (Fig. 5).

Genetic or environmental factors not correlated or interacting with the tested marker SNP do not skew the proposed test statistics. Thus, even when there is allelic heterogeneity, the variants not in LD with the marker SNP have no effects on the test. Although we modeled the causal variants to have an additive effect on QT, the mode of inheritance does not change the results, because homozygotes for a low-frequency allele are rare and negligible. In the power assessment by simulation, we modeled the causal variants to be in complete LD (D' = 1) with the marker. When LD decays, the heteroscedasticity or skewness at the marker becomes weaker and less detectable. However, since the marker is associated with QT, causal variant(s) of the same directional effect should be loaded mostly to one allele of the marker, thus the decay of LD would be limited.

There are a few limitations in using heteroscedasticity and skewness to detect synthetic association. False positives arise if a causal variant itself shows heteroscedasticity. This can result from a strong gene–environment interaction. Indeed, the test of heteroscedasticity has been used for detecting such interaction

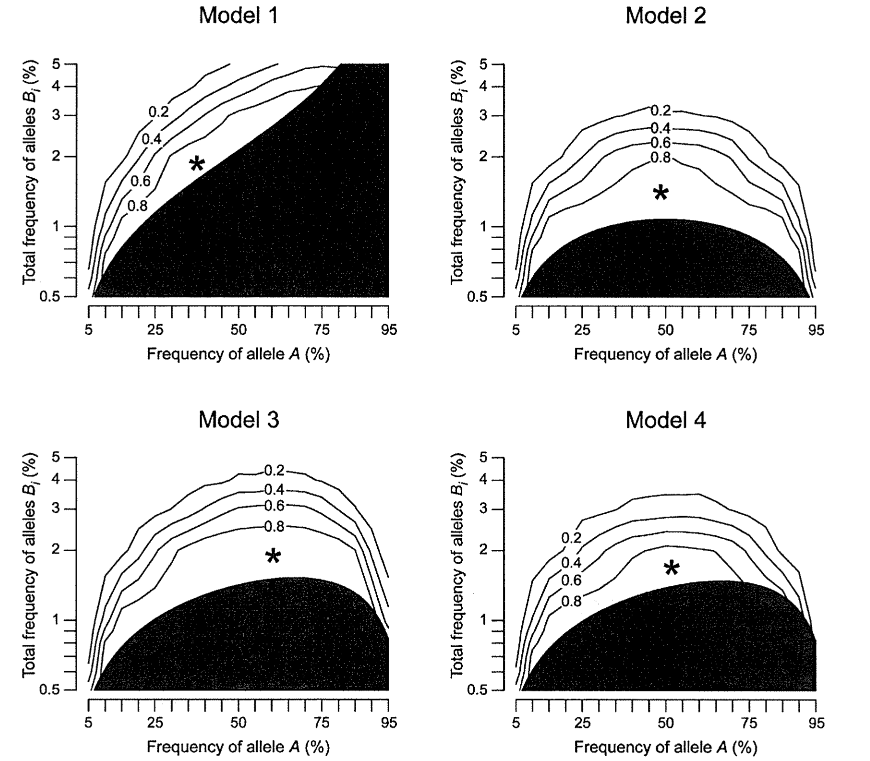


Figure 4. Power for detecting synthetic association by testing skewness. The power was computed from simulation under four representative genetic models, assuming the strength of marker association (R^2_{mrk}) of 0.00592. The format of the figure is the same as Figure 3.

(Pare et al. 2010). Another possible source of false positives is population stratification; if two subpopulations have a different mean QT at a specific causal genotype, the QT variance enlarges when the subpopulations are combined. Although a realistic level of population stratification is unlikely problematic (Supplemental Table 1), we recommend applying the test to each cohort separately.

Although the tests we proposed are limited to QTs, the idea of stratifying individuals by marker genotype leads to another test of synthetic association, which is applicable to quantitative as well as dichotomous traits. Here, we compare the association of neighboring common SNPs among the strata. If rare causal alleles are loaded onto the marker allele A, but not on the allele a, neighboring SNPs in LD with the causal variants will show association in the individuals with A/A or A/a genotypes, but not in the individuals with a/a genotype. As a result, the association P-value of neighboring SNPs would be distributed differently among the strata. Contrarily, if the marker SNP represents indirect association, no neighboring SNPs will be associated in any of the strata.

The proposed test can be helpful in understanding the "missing heritability" that GWA studies failed to account for (Maher 2008). If synthetic association is detected at a GWA signal SNP, the heritability of the SNP is likely an underestimate of the heritability of the whole locus (Dickson et al. 2010). Actually, in the *APOE* example, the explained variance was much smaller for the leading SNPs on a GWA chip (R^2 = 0.006 and 0.005) than for the two causal variants (R^2 = 0.040 and 0.008; see Table 1). In other words, our method can identify loci that contribute to a trait more than what we would expect from GWA study results.

It is unknown what proportion of GWA signals are due to synthetic association rather than indirect association. The situation is likely to differ by the function and molecular evolution of the genes. For example, several common causal variants are known for pharmacological traits that have not been under evolutionary selection (Cirulli and Goldstein 2010). On the contrary, only rare causal variants are found in genes for renal salt reabsorption, which have been under purifying selection; homozygotes of mutant alleles are susceptible to severe renal salt wasting and hypotension, although heterozygotes confer health benefits from lower blood pressure in postreproductive ages (Ji et al. 2008). Although the proposed test would help detecting synthetic association of a marker SNP, the discovery of the causal variants can require resequencing of a large number of individuals if the causal variants are rare. The aim of the proposed test is to assess potential synthetic association at a particular locus and then use the information to help design future resequencing studies.

Testing synthetic association is useful for designing fine mapping after exhaustively interrogating association of common SNPs at a locus. Exhaustive analysis of common SNPs (MAF ≥5%) is becoming accomplishable by genotyping with SNP chips of the GWA test and by imputing the unassayed SNPs using the HapMap or the 1000 genomes project data. The next focus is to explore rarer variants by resequencing and to identify the causal variants. Since resequencing is still expensive, we need to prioritize which GWA loci and which individuals are to be resequenced. Such information is obtainable by testing synthetic association of the common leading SNP(s), showing the strongest association in a locus. If synthetic association is detected for the leading SNP(s), rarer variants need to be examined in order to pinpoint the variants causing synthetic association. Moreover, if heteroscedasticity is detected, we can discover the causal variants efficiently by resequencing individuals having the homozygote genotype with larger QT variance, and especially those having extreme QT values, who are enriched with the rare causal alleles. Alternatively, if the test for synthetic association is not significant (in >5000 samples), the leading SNP(s) or their proxies are likely causal. Whereas conventional fine-mapping techniques aim to find the causal SNP(s) or haplotype(s) from a set of SNPs tested for association (McCarthy et al. 2008), our method is unique in suggesting that causal variants can be discovered if the study is extended to rarer variants.

Numerous SNP associations have been identified in recent GWA studies, yet our understanding of causal variants is very limited; it is not easy to prove functional changes, let alone the causality with the associated phenotype (Cirulli and Goldstein 2010). We proposed a simple statistical test, which helps to detect whether a common SNP associated with a QT is a noncausal marker in LD with rarer causal variant(s). The proposed test statistic can serve as a milepost in fine mapping and help understand the genetic structure of complex traits.