# Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease

Xiangfeng Lu[1,45], Laiyuan Wang[1,2,45], Shufeng Chen[1], Lin He[3], Xueli Yang[1], Yongyong Shi[3], Jing Cheng[4,5], Liang Zhang[4], C Charles Gu[6], Jianfeng Huang[1], Tangchun Wu[7], Yitong Ma[8], Jianxin Li[1], Jie Cao[1], Jichun Chen[1], Dongliang Ge[1], Zhongjie Fan[9], Ying Li[1], Liancheng Zhao[1], Hongfan Li[2], Xiaoyang Zhou[10], Lanying Chen[1], Donghua Liu[1], Jingping Chen[1], Xiufang Duan[1], Yongchen Hao[1], Ligui Wang[7], Fanghong Lu[11], Zhendong Liu[11], Cailiang Yao[12], Chong Shen[12], Xiaodong Pu[13], Lin Yu[13], Xianghua Fang[14], Lihua Xu[15], Jianjun Mu[16], Xianping Wu[17], Runping Zheng[18], Naqiong Wu[19], Qi Zhao[1], Yun Li[20], Xiaoli Liu[21], Mengqin Wang[22], Dahai Yu[1], Dongsheng Hu[23,24], Xu Ji[25], Dongshuang Guo[26], Dongling Sun[27], Qianqian Wang[1], Ying Yang[1], Fangchao Liu[1], Qunxia Mao[1], Xiaohua Liang[1], Jingfeng Ji[1], Panpan Chen[28], Xingbo Mo[1], Dianjiang Li[1], Guoping Chai[1], Yida Tang[19], Xiangdong Li[1], Zhenhan Du[1], Xuehui Liu[2], Chenlong Dou[2], Zili Yang[29], Qingjie Meng[30], Dong Wang[31], Renping Wang[32], Jun Yang[33], Heribert Schunkert[34,35], Nilesh J Samani[36,37], Sekar Kathiresan[38–40], Muredach P Reilly[41], Jeanette Erdmann[34,35], The Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Consortium[42], Xiaozhong Peng[43], Xigui Wu[1], Depei Liu[43], Yuejin Yang[19], Runsheng Chen[44], Boqin Qiang[43] & Dongfeng Gu[1]

© 2012 Nature America, Inc. All rights reserved.

npg

**We performed a meta-analysis of 2 genome-wide association studies of coronary artery disease comprising 1,515 cases and 5,019 controls followed by replication studies in 15,460 cases and 11,472 controls, all of Chinese Han ancestry. We identify four new loci for coronary artery disease that reached the threshold of genome-wide significance ($P < 5 \times 10^{-8}$). These loci mapped in or near *TTC32-WDR35*, *GUCY1A3*, *C6orf10-BTNL2* and *ATP2B1*. We also replicated four loci previously identified in European populations (in or near *PHACTR1*, *TCF21*, *CDKN2A-CDKN2B* and *C12orf51*). These findings provide new insights into pathways contributing to the susceptibility for coronary artery disease in the Chinese Han population.**

Coronary artery disease (CAD) and its most severe complication, myocardial infarction, are leading causes of mortality and disability worldwide[1,2]. Recent genome-wide association studies (GWAS) of CAD have identified multiple chromosomal regions associated with this disease[3]. However, most of these GWAS have focused on individuals of European ancestry, and the identified loci altogether explain only a small fraction of the risk for CAD. The associated variants identified in populations of European ancestry might not be associated with disease in other ancestry groups because of underlying genetic heterogeneity. Therefore, larger scale studies in Chinese and other non-European populations are needed to identify additional susceptibility loci and to improve understanding of the mechanisms underlying susceptibility to CAD.

Herein, we report the results of a two-stage GWAS of CAD in a sample including ~33,000 Han Chinese individuals (**Supplementary Fig. 1**). In the discovery stage (stage 1), we performed two GWAS of CAD in two independent Chinese studies: the Beijing Atherosclerosis Study (BAS) and the China Atherosclerosis Study (CAS). BAS consisted of 509 cases of myocardial infarction and 1,034 controls genotyped using the Affymetrix Human Mapping 500K Array set of 500,568 SNPs. CAS consisted of 1,034 cases of CAD and 4,245 controls genotyped using the Affymetrix Axiom Genome-Wide CHB 1 Array Plate of 657,124 SNPs. After a series of quality control procedures (Online Methods), we retained 367,129 autosomal SNPs in 505 cases and 1,021 controls in BAS and 613,724 autosomal SNPs in 1,010 cases and 3,998 controls in CAS (**Supplementary Table 1**). To facilitate combining GWAS results from the two genotyping platforms, we imputed missing genotypes using reference haplotypes from the phased Han Chinese in Beijing (CHB) and Japanese in Tokyo, Japan (JPT) HapMap (release 22) reference data set[4]. We performed a meta-analysis of the two GWAS, generating association data for approximately 2.2 million genotyped or imputed autosomal SNPs.

Principal-component analysis showed minimal evidence for population stratification in the study populations (**Supplementary Fig. 2**). A quantile-quantile plot showed that the distribution of observed *P* values deviated from expected *P* values only in the extreme tail (**Supplementary Fig. 3**). The genomic inflation factor ($\lambda$) was 1.04 for both BAS and CAS individually and 1.05 for the two studies combined, indicating that population stratification effects were negligible in our study samples.

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

Received 13 March; accepted 5 June; published online 1 July 2012; doi:10.1038/ng.2337

© 2012 Nature America, Inc. All rights reserved.

**Table 1 Association results for four CAD susceptibility loci newly identified by GWAS in the Chinese**

| | | | | | Discovery meta-analysis (BAS + CAS) | | | | Replication 1 | | | | Replication 2 | | | | Replication 3 | | | | Combined results (discovery + replication) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Sample size | | | | Sample size | | | | Sample size | | | | Sample size | | | | Sample size | |
| SNP | Locus | Position | Risk allele (freq.) | Nearby gene | OR (95% CI) | $P$ | Cases | Controls | OR (95% CI) | $P$ | Cases | Controls | OR (95% CI) | $P$ | Cases | Controls | OR (95% CI) | $P$ | Cases | Controls | OR (95% CI) | $P$ | Cases | Controls |
| rs2123536 | 2p24.1 | 19809058 | T (0.39) | *TTC32-WDR35* | 1.25 (1.12–1.39) | $4.92 \times 10^{-5}$ | 1,008 | 3,982 | 1.09 (1.03–1.15) | $1.56 \times 10^{-3}$ | 8,732 | 5,175 | 1.16 (1.06–1.26) | $9.56 \times 10^{-4}$ | 2,373 | 1,958 | 1.11 (1.05–1.18) | $8.82 \times 10^{-4}$ | 4,103 | 4,177 | 1.12 (1.08–1.16) | $6.83 \times 10^{-11}$ | 16,216 | 15,292 |
| rs1842896 | 4q32.1 | 156730909 | T (0.76) | *GUCY1A3* | 1.23 (1.11–1.37) | $9.75 \times 10^{-5}$ | 1,503 | 4,991 | 1.13 (1.06–1.20) | $5.79 \times 10^{-5}$ | 8,644 | 5,158 | 1.17 (1.05–1.29) | $2.97 \times 10^{-3}$ | 2,324 | 1,961 | 1.11 (1.04–1.20) | $3.52 \times 10^{-3}$ | 4,229 | 4,169 | 1.14 (1.10–1.19) | $1.26 \times 10^{-11}$ | 16,700 | 16,279 |
| rs9268402 | 6p21.32 | 32449331 | G (0.59) | *C6orf10-BTNL2* | 1.17 (1.07–1.27) | $6.97 \times 10^{-4}$ | 1,510 | 5,009 | 1.19 (1.13–1.26) | $3.32 \times 10^{-11}$ | 8,708 | 5,120 | | | | | 1.11 (1.04–1.18) | $1.52 \times 10^{-3}$ | 4,208 | 4,071 | 1.16 (1.12–1.20) | $2.77 \times 10^{-15}$ | 14,426 | 14,200 |
| rs7136259 | 12q21.33 | 88605319 | T (0.39) | *ATP2B1* | 1.21 (1.11–1.33) | $1.95 \times 10^{-5}$ | 1,502 | 4,993 | 1.10 (1.04–1.16) | $4.73 \times 10^{-4}$ | 8,732 | 5,173 | 1.09 (1.00–1.18) | $5.27 \times 10^{-2}$ | 2,394 | 2,058 | 1.10 (1.03–1.17) | $3.26 \times 10^{-3}$ | 4,218 | 4,126 | 1.11 (1.08–1.15) | $5.68 \times 10^{-10}$ | 16,846 | 16,350 |

SNP IDs and chromosomal positions are based on NCBI Build 36 of the genome. The risk allele frequency (freq) in discovery studies is shown. For each SNP, ORs (with 95% CIs) and $P$ values represent the risk of CAD associated with each copy of the risk allele.

In this stage 1 discovery analysis, we used a set of criteria (Online Methods) to select 96 SNPs for genotyping in a further replication analysis in a case-control sample including 8,803 CAD cases and 5,183 controls (replication 1) (**Supplementary Table 1**). We found that two association signals reached genome-wide significance (defined as $P < 5 \times 10^{-8}$) in replication 1 alone. These included one locus (four SNPs at 9p21.3) previously identified in populations of European descent and one newly discovered locus (rs9268402 at *C6orf10-BTNL2*). In meta-analysis of the discovery and replication 1 stages, eight additional variants were found to be associated with CAD at a significance level of $P < 1 \times 10^{-5}$. These eight SNPs were further examined in an independent sample (2,408 CAD cases and 2,103 controls) (replication 2). The results of the selected SNPs in the discovery, replication 1 and replication 2 stages are summarized in **Supplementary Table 2**.

When combining the discovery data with those from replications 1 and 2, we found ten SNPs at seven regions associated with CAD at a prespecified threshold for genome-wide significance of $P < 5 \times 10^{-8}$. Data from four SNPs at 9p21.3 (rs9632884, rs10757274, rs1333042 and rs1333049), rs9349379 at 6p24.1 in *PHACTR1* and rs11066280 at 12q24.13 near *C12orf51* confirmed associations previously reported in Europeans[5–11]. rs12524865 at 6q23.2 near *TCF21* also showed consistent association evidence in the discovery and replication stages with the same direction of association previously reported in Europeans[12], with this association nearly reaching the genome-wide significance threshold in the combined analyses ($P = 1.87 \times 10^{-7}$) (**Supplementary Tables 2** and **3**). Moreover, we identified four new CAD-associated loci in the Chinese ($P$ values ranging from $4.48 \times 10^{-8}$ to $9.92 \times 10^{-14}$): (i) rs2123536 at 2p24.1 near *TTC32-WDR35*; (ii) rs1842896 at 4q32.1 near *GUCY1A3*; (iii) rs9268402 at 6p21.32 near *C6orf10-BTNL2*; and (iv) rs7136259 at 12q21.33 near *ATP2B1*.

We sought to further replicate the associations of these four SNPs in an additional validation stage, using an independent sample consisting of 4,249 CAD cases and 4,186 controls (replication 3). All four SNPs showed significant association with CAD in this additional sample after adjustment for multiple testing ($P < 0.05/4 = 0.0125$), and combining results from the discovery, replication 1, replication 2 and replication 3 stages showed associations of these loci with CAD (rs2123536: $P = 6.83 \times 10^{-11}$, odds ratio (OR) = 1.12; rs1842896: $P = 1.26 \times 10^{-11}$, OR = 1.14; rs9268402: $P = 2.77 \times 10^{-15}$, OR = 1.16; rs7136259: $P = 5.68 \times 10^{-10}$, OR = 1.11) (**Table 1**). The associations of these loci are shown in the context of their genomic coordinates (**Fig. 1**).

Of the four newly reported CAD-associated loci, 4q32.1 (rs13139571) and 12q21.33 (rs2681492) were also recently identified as susceptibility loci for blood pressure in European studies[13,14]. We observed that rs13139571 and rs1842896 at 4q32.1 showed very weak linkage disequilibrium (LD) ($r^2 = 0.002$, $D' = 0.062$ in HapMap CHB; $r^2 = 0.004$, $D' = 0.123$ in HapMap Utah residents of Northern and Western European ancestry (CEU)), whereas rs2681492 and rs7136259 at 12q21.33 were in strong LD ($r^2 = 0.90$, $D' = 0.95$ in HapMap CHB; $r^2 = 0.11$, $D' = 0.88$ in HapMap CEU). To shed light on the seemingly entangled relationship of 4q32.1 and 12q21.33 with CAD and hypertension, we examined the associations of the two CAD SNPs (rs1842896 and rs7136259) with blood pressure in all control samples from the discovery and replication samples. Suggestive associations with blood pressure and hypertension were observed (**Supplementary Table 4**). We further examined whether associations with hypertension could mediate the effects of these loci on CAD. After adjustment for their effects on hypertension, the SNPs retained association with CAD at genome-wide significance ($P = 1.31 \times 10^{-9}$, OR = 1.13 for rs1842896; $P = 6.63 \times 10^{-12}$, OR = 1.13 for rs7136259), indicating that these SNPs might be associated with CAD independently of their

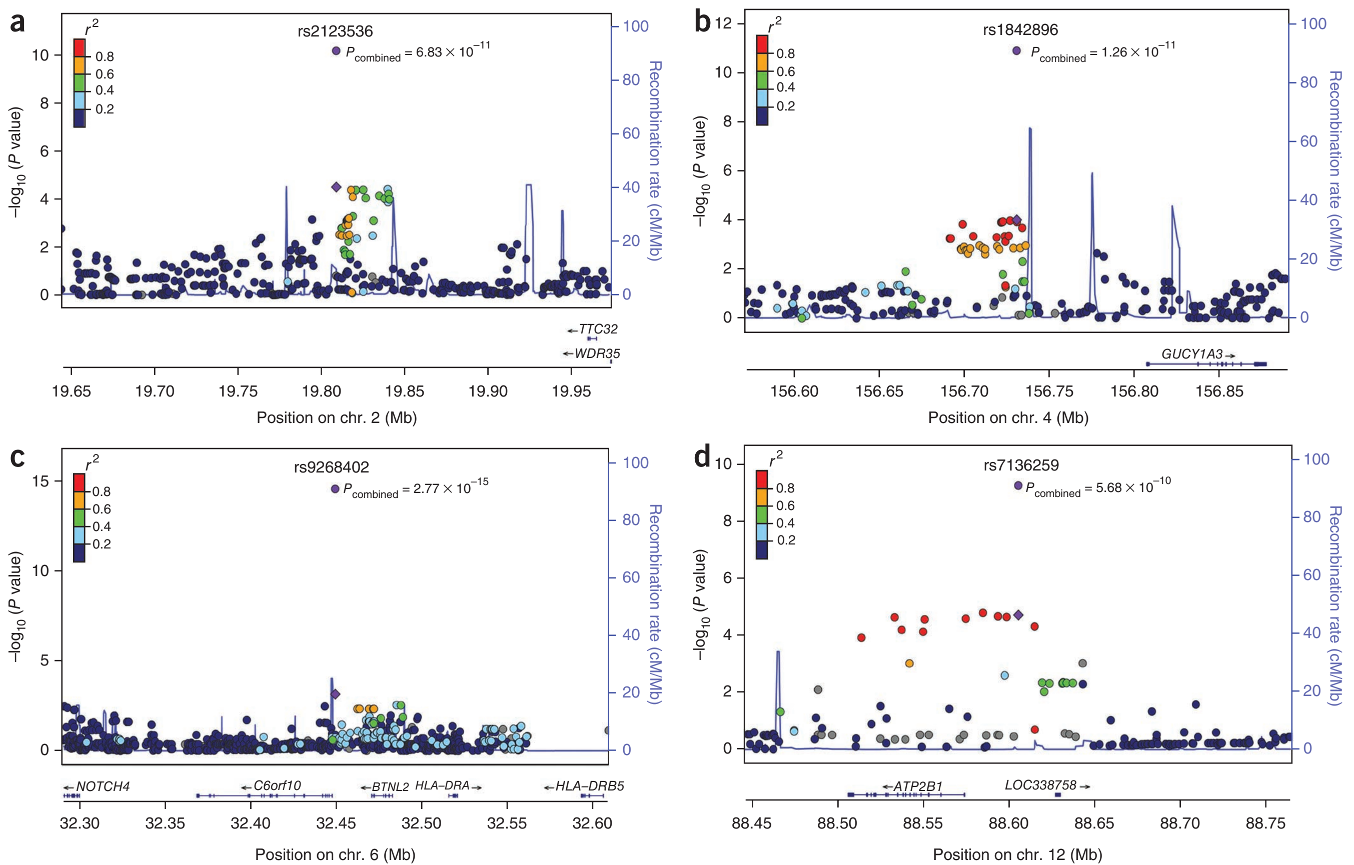


**Figure 1** Regional plots of the four loci newly associated with CAD. (**a**–**d**) Results are shown for rs2123536 (**a**), rs1842896 (**b**), rs9268402 (**c**) and rs7136259 (**d**). Top, SNPs are plotted according to their chromosomal positions (NCBI Build 36) with $-\log_{10} P$ values from the discovery meta-analysis. The lead SNP is represented by a purple diamond for the discovery meta-analysis and is represented by a purple circle for the combined analysis, denoted by the $P_{combined}$ value. LD ($r^2$ values) between the lead SNP and the other SNPs are indicated by color. The estimated recombination rates from 1000 Genomes Project (June 2010) CHB and JPT samples are plotted in cyan to reflect local LD structure. Bottom, genes within the region of interest are annotated and are shown as arrows. Plots were generated using LocusZoom.

© 2012 Nature America, Inc. All rights reserved.

effects on hypertension. rs1842896 is located 76.4 kb upstream of the *GUCY1A3* locus. The *GUCY1A3* gene encodes the α subunit of soluble guanylate cyclase (sGC), a key enzyme in the nitric oxide signaling pathway that is implicated in pathogenesis of CAD and atherosclerosis. Preclinical studies have explored the therapeutic potential of sGC stimulators[15]. rs7136259 is near *ATP2B1*, which encodes PMCA1, a plasma membrane calcium ATPase that pumps calcium ions ($Ca^{2+}$) out of the cytosol into the extracellular milieu[16].

The *C6orf10-BTNL2* locus at 6p21.32 is a region associated with immune-related diseases[17–19]. *BTNL2* encodes a member of the immunoglobulin superfamily that probably functions as a T-cell costimulatory molecule. It is noteworthy that rs2076530, a truncating splice-site G>A mutation in *BTNL2*[17], is in strong LD with rs9268402 ($r^2$ = 0.59). *BTNL2* polymorphisms have been found to be associated with susceptibility to Kawasaki disease[20], a vasculitis of young childhood that particularly affects the coronary arteries, and with increased risk of developing ischemic heart disease in the future[21]. rs2123536 at 2p24.1 is located ~150 kb downstream of *TTC32* and *WDR35*. *TTC32* encodes the protein containing the tetratricopeptide repeat motif that mediates binding to other peptides[22]. *WDR35* encodes a member of the WD repeat protein family[23], whose members are involved in cell cycle progression, signal transduction, apoptosis and gene regulation.

The association of the chromosome 12q24 region with CAD is of particular interest. The association signal at 12q24 spans ~0.7 Mb, and rs11066280 is in almost perfect LD ($r^2$ = 0.95–0.97) with rs3782886, rs4646776, rs671, rs2074356 and rs77768175 (**Supplementary Fig. 4**). Previous studies have shown significant evidence supporting signatures of natural selection[11,24] and pleiotropic effects for this region (in CAD[10,11,25] and the regulation of plasma lipid levels[24,26,27] and blood pressure[14,24,28]). All variants at 12q24 that are associated with CAD in Europeans[10,11] are not polymorphic in the Chinese, whereas all CAD-associated variants at 12q24 in the Chinese are monomorphic in Europeans (**Supplementary Table 5**). In the present study, rs11066280 also showed significant or suggestive evidence of association with the levels of high-density lipoprotein cholesterol (HDL-C), triglycerides and total cholesterol and with blood pressure (**Supplementary Table 4**). Because there is not substantial evidence for functional variants at this locus, further in-depth analysis is needed to explain the long-range LD and uncover causal mechanisms for CAD in this region.

Although the association between the *CDKN2A-CDKN2B* locus at 9p21 and CAD was replicated in our study, we note that the LD structure of the 9p21 region is different in populations of European and Asian ancestry (**Supplementary Fig. 5**). The four SNPs showing significant association with CAD in the 9p21.3 region were in almost perfect LD in individuals of European descent (pairwise $r^2$ values from 0.84–0.90). In the Chinese, however, two of the SNPs (rs10757274 and rs1333049) were in strong LD with each other ($r^2$ = 0.78) but were in only moderate LD with the other two SNPs (rs9632884 and rs1333042;

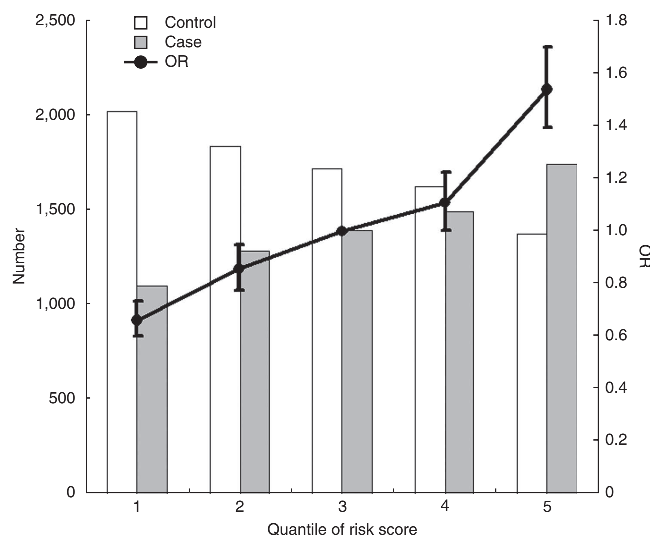


**Figure 2** CAD risk score categories and risk for CAD. ORs for CAD of the different quantiles compared with those from quintile 3 are shown as solid dots with whiskers showing 95% CIs. White and gray bars represent the number of controls and cases in each quintile, respectively. Analysis was performed in a total of 15,591 individuals (7,012 cases, 8,579 controls).

pairwise $r^2$ values from 0.27–0.43; **Supplementary Table 6**). Conditional logistic analyses showed that the individual association evidence for rs1333042 and rs10757274 remained significant after controlling for the genetic effect at any of the other three SNPs ($P$ values ranging from $3.71 \times 10^{-8}$ to $7.96 \times 10^{-9}$ and from $1.44 \times 10^{-6}$ to $1.18 \times 10^{-18}$, respectively) (**Supplementary Table 7**). Therefore, the two SNPs (rs1333042 and rs10757274) seem to represent independent association signals.

We evaluated whether the CAD-associated variants identified in our samples from the Chinese population were associated with CAD in Europeans, using the results from the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) consortium study[12], a meta-analysis of 22,233 cases and 64,762 controls (**Supplementary Table 8**). Of the four SNPs, rs2123536 ($P = 0.0038$, OR = 1.10) and rs7136259 ($P = 0.035$, OR = 1.03) showed nominal association with CAD in the population of European ancestry, and the direction of effect was consistent with that seen in our findings. Associations for the other two SNPs were not detected in data from the CARDIoGRAM consortium study.

Conversely, we also investigated whether the 35 CAD-associated SNPs (in 29 loci) identified by previous GWAS in European populations were associated with CAD in our sample (**Supplementary Table 9**). In addition to the four loci that were confirmed by our discovery and replication studies (**Supplementary Table 3**), seven loci at 1p32.2, 1q41, 10q23.31, 10q24.32, 11q22.3, 15q25.1 and 17p13.3 showed directionally consistent and nominally significant associations in the discovery study ($P < 0.05$) (**Supplementary Table 9**, group 1). We observed that 11 SNPs in 10 loci were monomorphic or had low minor allele frequency (MAF ≤ 0.1) in the Chinese Han population, whereas these SNPs were quite polymorphic in European populations (**Supplementary Table 9**, group 2). We examined the associations of other correlated SNPs in strong LD with these 11 SNPs in HapMap CEU data with CAD in the Chinese. Of interest, association of the proxy SNPs in three loci, 3q22.3, 6q26 and 17p11.2, was also supported by our discovery analysis, albeit with only suggestive evidence ($P < 0.05$). These data suggest that differences in LD structure may partially explain the discrepancies in association between the European and Chinese populations. No associations were observed for the remaining eight SNPs with common MAF (MAF > 0.1) in the Chinese Han population (**Supplementary Table 9**, group 3). The observed differences between the results from the Chinese and European populations might be due to differences in genetic architecture and environmental factors or might result from insufficient power in the present study. A recent GWAS of CAD in the Chinese Han population[29] identified rs6903956 at 6p24.1 (*C6orf105*) as a susceptibility locus. We could not replicate this association with rs6903956 in our data, although our discovery analyses had >90% power to detect a SNP with an OR of 1.51, even when we used a $P$-value threshold of $1.0 \times 10^{-5}$.

To examine the effect of nine associated SNPs given in **Table 1** and **Supplementary Table 3** (rs2123536, rs1842896, rs9349379, rs9268402, rs12524865, rs10757274, rs1333042, rs7136259 and rs11066280) in aggregate on the risk for CAD, a CAD risk score was calculated using the weighted sum across the SNPs, combining effect sizes and doses of risk alleles. The CAD risk score could explain ~1.92% of the variance in risk for CAD. The mean CAD risk score of cases was significantly higher than that of controls ($P < 1 \times 10^{-74}$). Logistic regression was applied to test the association of risk score categories with CAD. Compared with individuals in the bottom quintile, individuals in the top quintile of CAD risk score had greater than twofold increased risk for CAD (OR = 2.34, 95% confidence interval (CI) = 2.11–2.59). The risk for CAD across quintiles of CAD risk score is shown (**Fig. 2**).

In conclusion, we identified four new loci associated with CAD (in or near *TTC32-WDR35*, *GUCY1A3*, *C6orf10-BTNL2* and *ATP2B1*) in the Chinese and replicated four previously reported loci (*PHACTR1*, *TCF21*, *CDKN2A-CDKN2B* and *C12orf51*). These results suggest that both shared and unique genetic backgrounds of CAD susceptibility are present in different ancestry groups and highlight the importance of fine-mapping efforts to pinpoint causal variants and mechanisms. Further study and integration of GWAS findings in multiple ancestry groups will surely promote a better understanding of the global genetic architecture of CAD risk.

**URLs.** PLINK v1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; R, http://www.r-project.org/; METAL, http://www.sph.umich.edu/csg/abecasis/metal/; The International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; LocusZoom, http://csg.sph.umich.edu/locuszoom/.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary information is available in the online version of the paper.*

### ACKNOWLEDGMENTS

This study was funded by the National Basic Research Program of China (973 Plan) (2011CB503901 and 2006CB503805) from the Ministry of Science and Technology of China, by the National Science Foundation of China (30930047), by the High-Tech Research and Development Program of China (863 Plan) (2009AA022703, 2012AA02A516 and 2006AA02A406) and by a grant from the Ministry of Science and Technology of China (2006BAI01A01). This study was also supported by Biomedical Project from the Council of Science and Technology, Beijing (H020220030130).

### AUTHOR CONTRIBUTIONS

D. Gu conceived of and designed the study and supervised all the sample selection, genotyping, data analysis and interpretation. B.Q., Depei Liu and X. Peng participated in the study design and interpretation. Genotyping experiments were

© 2012 Nature America, Inc. All rights reserved.
npg

performed by S.C., Y.S., L. Zhang and H.L. under the supervision of D. Gu, L.H., J. Cheng and L.C. DNA sample preparation was carried out by Laiyuan Wang, H.L., T.W., Y.M., Q.Z., Yun Li, D.Y., Q.W., Ying Yang, F. Liu, Q. Mao, X. Liang, J.J., X.M., D. Li, Xuehui Liu and C.D. Phenotype collection and data management were conducted by J.H., S.C., J.L., J. Cao, Jichun Chen, Donghua Liu, Jingping Chen, X.D., T.W., Ligui Wang, Y.M., Z.F., Ying Li, L. Zhao, X.Z., F. Lu, Z.L., C.Y., C.S., X. Pu, L.Y., X.F., L.X., J.M., Xianping Wu, R.Z., N.W., Xiaoli Liu, M.W., D.H., X.J., D. Guo, D.S., P.C., G.C., Xigui Wu, L.C., Yuejin Yang, Y.T., X. Li, Z.D., Z.Y., Q. Meng, D.W., R.W. and J.Y. H.S., N.J.S., S.K., M.P.R. and J.E. provided the data from the CARDIoGRAM Consortium. Statistical analysis was performed by X. Lu, X.Y., Y.H., D. Ge, C.C.G. and R.C. The manuscript was written by D. Gu, X. Lu and Laiyuan Wang. All authors reviewed the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2337.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Roger, V.L. *et al.* Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation* **123**, e18–e209 (2011).
2. He, J. *et al.* Major causes of death among men and women in China. *N. Engl. J. Med.* **353**, 1124–1134 (2005).
3. Peden, J.F. & Farrall, M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum. Mol. Genet.* **20**, R198–R205 (2011).
4. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
5. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
6. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
7. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
8. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
9. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
10. Gudbjartsson, D.F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* **41**, 342–347 (2009).
11. 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).
12. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
13. Ehret, G.B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
14. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
15. Stasch, J.P. Soluble guanylate cyclase as an emerging therapeutic target in cardiopulmonary disease. *Circulation* **123**, 2263–2273 (2011).
16. Tempel, B.L. & Shilling, D.J. The plasma membrane calcium ATPase and disease. *Subcell. Biochem.* **45**, 365–383 (2007).
17. Valentonyte, R. *et al.* Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*. *Nat. Genet.* **37**, 357–364 (2005).
18. Clancy, R.M. *et al.* Identification of candidate loci at 6p21 and 21q22 in a genome-wide association study of cardiac manifestations of neonatal lupus. *Arthritis Rheum.* **62**, 3415–3424 (2010).
19. Jin, Y. *et al.* Genome-wide analysis identifies a quantitative trait locus in the MHC class II region associated with generalized vitiligo age of onset. *J. Invest. Dermatol.* **131**, 1308–1312 (2011).
20. Hsueh, K.C., Lin, Y.J., Chang, J.S., Wan, L. & Tsai, F.J. *BTNL2* gene polymorphisms may be associated with susceptibility to Kawasaki disease and formation of coronary artery lesions in Taiwanese children. *Eur. J. Pediatr.* **169**, 713–719 (2010).
21. Kato, H. *et al.* Long-term consequences of Kawasaki disease. A 10- to 21-year follow-up study of 594 patients. *Circulation* **94**, 1379–1385 (1996).
22. Smith, D.F. Tetratricopeptide repeat cochaperones in steroid receptor complexes. *Cell Stress Chaperones* **9**, 109–121 (2004).
23. Nakajima, D. *et al.* Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones. *DNA Res.* **9**, 99–106 (2002).
24. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* **43**, 531–538 (2011).
25. Takeuchi, F. *et al.* Genome-wide association study of coronary artery disease in the Japanese. *Eur. J. Hum. Genet.* **20**, 333–340 (2012).
26. Kim, Y.J. *et al.* Large-scale genome-wide association studies in east Asians identify new genetic loci influencing metabolic traits. *Nat. Genet.* **43**, 990–995 (2011).
27. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
28. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
29. Wang, F. *et al.* Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat. Genet.* **43**, 345–349 (2011).

[1]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [2]National Human Genome Center at Beijing, Beijing, China. [3]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China. [4]National Engineering Research Center for Beijing Biochip Technology (NERCBBT), Beijing, China. [5]Medical Systems Biology Research Center, School of Medicine, Tsinghua University, Beijing, China. [6]Division of Biostatistics, Washington University School of Medicine, St Louis, Missouri, USA. [7]Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, China. [8]Department of Cardiology, First Affiliated Hospital of Xinjiang Medical University, Urumchi, China. [9]Department of Cardiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China. [10]Department of Cardiology, Renmin Hospital of Wuhan University, Wuhan, China. [11]Cardio-Cerebrovascular Control and Research Center, Institute of Basic Medicine, Shandong Academy of Medical Sciences, Jinan, China. [12]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China. [13]Department of Cardiology, Fujian Provincial People's Hospital, Fuzhou, China. [14]Department of Epidemiology and Social Medicine, Xuanwu Hospital, Capital Medical University, Beijing, China. [15]Department of Cardiology, The Affiliated Hospital of Beihua University, Beihua University, Jilin, China. [16]Department of Cardiology, First Affiliated Hospital of Medical College, Xi'an Jiaotong University, Xi'an, China. [17]Center for Chronic and Noncommunicable Disease Control and Prevention, Sichuan Center of Disease Control and Prevention, Chengdu, China. [18]Cardiovascular Disease Prevention Center, Shougang Hospital, Peking University, Beijing, China. [19]Department of Cardiology, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [20]School of Public Health, Hebei United University, Tangshan, China. [21]Department of Cardiology, Anzhen Hospital, Capital Medical University, Beijing, China. [22]Department of Emergency Medicine, Beijing Jishuitan Hospital, Beijing, China. [23]School of Public Health, Zhengzhou University, Zhengzhou, China. [24]School of Medicine, Shenzhen University, Shenzhen, China. [25]Department of Internal Medicine, Xinle Red Cross Hospital, Xinle, China. [26]Department of Internal Medicine, Yuxian Renmin Hospital, Yuxian, China. [27]Institute of Geriatrics, Chinese People's Liberation Army (PLA) General Hospital, Beijing, China. [28]Renqiu Heath Bureau, Renqiu, China. [29]Nantong Center of Disease Control and Prevention, Nantong, China. [30]Zhangqiu Heath Bureau, Zhangqiu, China. [31]Xin'an Heath Bureau, Xin'an, China. [32]Department of Internal Medicine, Shijiazhuang Greatwall Hospital, Shijiazhuang, China. [33]Department of Cardiology, Hanzhong Renmin Hospital, Hanzhong, China. [34]Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany. [35]Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK), Universität zu Lübeck, Lübeck, Germany. [36]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [37]Leicester National Institute for Health Research, Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, UK. [38]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [39]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [40]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [41]The Cardiovascular Institute, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [42]A full list of members and affiliations is provided in the **Supplementary Note**. [43]Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [44]Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. [45]These authors contributed equally to this work. Correspondence should be addressed to D. Gu (gudongfeng@vip.sina.com).

© 2012 Nature America, Inc. All rights reserved.
npg

## ONLINE METHODS

**Study populations.** We performed a two-stage case-control analysis in individuals of Chinese Han ancestry. The general characteristics of the study participants are summarized in **Supplementary Table 1**. In the discovery stage, we performed a meta-analysis of two independent GWAS studies, BAS and CAS. BAS[30] comprised 505 cases of myocardial infarction and 1,021 controls. All participants were from Beijing. All cases had a validated history of myocardial infarction, and disease status was verified by hospital records and by cardiologists according to standard protocol[31]. Controls were randomly selected from subjects participating in a community-based survey of cardiovascular risk factors in Beijing. The control subjects were judged to be free of CAD by history, clinical examination, electrocardiography and Rose questionnaire[32]. Detailed data were collected through in-person interviews with each case and control. Subjects with congenital heart disease, cardiomyopathy, valvular disease or renal or hepatic disease were excluded. CAS comprised 1,010 cases of CAD and 3,998 controls. The cases were from the northern provinces in China and were enrolled from Fuwai Hospital at the National Center of Cardiovascular Diseases. Of the cases, 83.8% had a family history of CAD. Diagnoses of cases with myocardial infarction followed strict diagnostic rules based on signs, symptoms, electrocardiograms and the activity of cardiac enzymes[31]. The subjects with CAD who did not have a known history of myocardial infarction had >70% stenosis in at least one major epicardial vessel, with the exception of the left main coronary artery where >50% stenosis was sufficient to meet the diagnosis of CAD. Controls in CAS were recruited from the International Collaborative Study of Cardiovascular Disease in Asia (InterASIA in China)[33]. InterASIA used a four-stage stratified sampling method to select a nationally representative sample of the general population aged 35 to 74 years in China. A total of 15,838 persons completed the survey and examination in 2001, and a follow-up of this survey in 2008 was conducted. There were 3,998 controls who did not develop incident CAD and had no family history of CAD during the 8-year follow-up period of the study in four northern field centers of InterASIA.

In stage 2, replication analyses were conducted in three independent samples with a total of 15,460 cases and 11,472 controls (8,803 cases and 5,183 controls for replication 1; 2,408 cases and 2,103 controls for replication 2; 4,249 cases and 4,186 controls for replication 3). All the cases in the replication stage were recruited using uniform criteria, and clinical information was collected using the same questionnaire as in CAS. For replication 1, CAD cases were recruited using a standardized protocol through collaboration among multiple hospitals in China; controls were selected from samples of the China Collaborative Study of Cardiovascular Epidemiology. For replication 2, cases were enrolled from Fuwai Hospital, Beijing; controls were selected from urban and suburban communities in Beijing. For replication 3, cases from northern China were enrolled from Fuwai Hospital, Beijing, and other medical centers; controls were selected from northern field centers of the China Cardiovascular Health Study (CCHS) project. CCHS has been a population-based investigation of risk factors for cardiovascular diseases in China since 2006. The controls for the stage 2 validation were selected using identical criteria to that used for the discovery populations.

Each study obtained approval from the institutional review boards of Fuwai Hospital, the Chinese Academy of Medical Sciences and Peking Union Medical College, and other medical institutions. All participants gave written informed consent.

**Genotyping and quality control.** For BAS in the discovery stage, a total of 509 cases of myocardial infarction and 1,034 controls were genotyped with the Affymetrix GeneChip Human Mapping 500K Array Set, including 500,568 SNPs. Principal-component analysis using EIGENSOFT[34,35] was used to compare all samples with reference samples from the HapMap Yoruba from Ibadan, Nigeria (YRI), CHB, JPT and CEU panels. We excluded SNPs with MAF of <0.01 in cases or controls ($n$ = 100,865, including 46,048 monomorphic SNPs); genotype call rate of <95% in cases or controls ($n$ = 20,030); or deviation from Hardy-Weinberg equilibrium ($P$ value of $<1 \times 10^{-4}$; $n$ = 12,544). We also excluded 17 samples because of gender discordance, high genotype missing rate (>3.0%) or cryptic relatedness (identity by descent (IBD) of >0.1875) or because the samples were population outliers. After quality control, 1,526 samples and 367,129 autosomal SNPs remained for subsequent analysis. For CAS in the discovery stage, a total of 1,034 cases of CAD and 4,245 controls were genotyped with the Axiom Genome-Wide CHB 1 Array Plate, which was designed for the Chinese population and includes 657,124 SNPs. After quality control procedures were applied, 5,008 samples and 613,724 autosomal SNPs were retained for subsequent analysis.

In stage 2, 96 SNPs were selected and genotyped using TaqMan SNP Genotyping Assays on the Fludigm EP1 platform for replication 1. Of the 96 SNPs genotyped, 5 with Hardy-Weinberg equilibrium $P$ values of <0.001 were removed from further association analysis. To assess genotyping reproducibility, 48 duplicate samples were genotyped, and the concordance rate was determined to be >99.4%. We selected and genotyped nine SNPs not at 9p21 with $P$ of $<1 \times 10^{-5}$ in the combined discovery and replication 1 analysis using the iPLEX MassARRAY platform (Sequenom) in replication 2. rs9268402 was not taken forward into replication 2 because of difficulty in the design of the replication array, leaving eight SNPs for replication. The concordance rate for 96 replicate samples was 99.7%. To evaluate the quality of the genotype data between different genotyping platforms, 8 SNPs in 48 random replication samples genotyped on the Fludigm EP1 platform were also genotyped on the iPLEX Sequenom MassARRAY platform, and the concordance rate between the genotypes from the two platforms was found to be 99.5%. In replication 3 of the four newly associated SNPs, samples were genotyped using a TaqMan genotyping platform (ABI 7900HT Real Time PCR system, Applied Biosystems). Cluster patterns of genotyping data from the Fludigm EP1, Sequenom and TaqMan analyses were examined to confirm high quality.

**Genotype imputation.** In the discovery stage, imputation of allele dosage for ungenotyped SNPs was carried out using MACH[36,37] with data from HapMap Phase 2 (JPT and CHB). After excluding imputed SNPs with an imputation quality score below a set threshold ($R^2$ of <0.30), call rate of <0.90 in either cases or controls, MAF of <0.01 in either cases or controls, Hardy-Weinberg equilibrium $P$ of $<1 \times 10^{-5}$ in controls or significantly different missing genotype rates in cases and controls ($P < 1 \times 10^{-5}$), we retained a total of 1,532,051 genotyped and imputed autosomal SNPs from BAS, 2,042,781 from CAS and 2,228,999 from the combined GWAS samples for subsequent association analysis.

**Selection of replication SNPs.** After genome-wide association analyses for each of the two discovery studies and meta-analysis in the combined sample, SNPs were taken forward to replication if they (i) showed potential association ($P < 1.0 \times 10^{-4}$) in the meta-analysis of the two studies; (ii) had a consistent association at $P \leq 1.0 \times 10^{-2}$ in both discovery populations; and (iii) associated at $P < 1.0 \times 10^{-3}$ in the meta-analysis discovery stage within 500 kb of a locus previously reported to show association with genome-wide or suggestive significance. All SNPs for replication had at least a correlated SNP (within 25 kb) that also showed an association signal ($P < 0.01$). With the exception of four SNPs at the 9p21.3 locus, SNPs in strong LD ($r^2 > 0.5$) with the most significantly associated SNP at each locus were removed from analysis. If a SNP could not be genotyped, alternative tagging SNPs (with $r^2$ of >0.8) were considered.

**Association analysis.** Association of imputed and genotyped SNPs with CAD was tested with multiple logistic regression analysis in an additive genetic model (with 1 degree of freedom) after adjusting for age (onset of the first event for cases or time of recruitment for controls) and sex. We used allele dosages from imputation to account for uncertainty in imputed genotypes. Association analyses were performed using PLINK[38] (see URLs). A fixed-effects inverse variance–weighted meta-analysis implemented in METAL[39] (see URLs) was used to combine the two discovery studies and to obtain results for each SNP across all replication studies. A quantile-quantile plot generated using R was used to evaluate the overall significance of the GWAS results and the potential impact of population stratification. The genomic inflation factor[40] ($\lambda$) was estimated from the median of the $\chi^2$ statistic divided by 0.456.

Association of loci with established cardiovascular risk factors was examined in all control samples from the discovery and replication samples. For quantitative traits (the levels of high-density lipoprotein, low-density lipoprotein, total cholesterol, triglycerides and fasting plasma glucose, blood pressure and body mass index), linear regressions were used, whereas, for the binary

© 2012 Nature America, Inc. All rights reserved.

npg

trait (hypertension), a logistic regression model was applied. We combined the regression estimates from each stage in a meta-analysis using inverse variance weighting.

Conditional analyses were performed to test the independence of significant SNPs in each region, conditioning on the genotype of the SNP chosen for replication. These analyses were carried out using PLINK with the *–logistic* and *–condition* options.


30. Hou, L. *et al.* Associations of *PLA2G7* gene polymorphisms with plasma lipoprotein-associated phospholipase A2 activity and coronary heart disease in a Chinese Han population: the Beijing atherosclerosis study. *Hum. Genet.* **125**, 11–20 (2009).
31. Alpert, J.S., Thygesen, K., Antman, E. & Bassand, J.P. Myocardial infarction redefined—a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J. Am. Coll. Cardiol.* **36**, 959–969 (2000).
32. Rose, G.A. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bull. World Health Organ.* **27**, 645–658 (1962).
33. Gu, D. *et al.* Prevalence of the metabolic syndrome and overweight among adults in China. *Lancet* **365**, 1398–1405 (2005).
34. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
35. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
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. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
39. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
40. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).


© 2012 Nature America, Inc. All rights reserved.

npg

 doi:10.1038/ng.2337