 Genome-wide association studies (GWASs) have identified over 175 genetic loci that contribute to lipid levels 1–6, which are heritable risk factors for cardiovascular disease, fatty liver disease, age-related macular degeneration, and type 2 diabetes7–9. However, most of the published lipid-associated variants are found in non-protein-coding regions of the genome, are without obvious biological significance, and explain only a small fraction of the heritability of lipid levels. The examination of low-frequency and potentially functional variants, which are poorly captured by standard GWAS arrays, has the potential to pinpoint causal variants and genes for follow-up and functional analyses, thereby promoting translation of the findings of genetic studies into new therapeutic targets. For example, low-frequency coding variants in \textit{PCSK9} reduce plasma levels of low-density lipoprotein cholesterol (LDL-C), reduce the risk of coronary artery disease (CAD), and have prompted the development of a new class of therapeutics10. Thus, we investigated the effect on lipid levels of the rare and low-frequency variants in the coding portion of the genome in an East Asian population, as East Asians have not been as extensively studied as the European population 11–13.

We carried out a meta-analysis of exome-wide association studies of blood lipid levels (high-density lipoprotein cholesterol (HDL-C), LDL-C, triglycerides (TGs), and total cholesterol (TC)) in a total of 47,532 East Asian samples that were genotyped by exome array. We further integrated the exome array data for plasma lipids in over 300,000 individuals, primarily of European ancestry (84%), from a study conducted by the Global Lipids Genetics Consortium (GLGC)14. We aimed to determine whether novel or population-specific variants and genes that influence lipid levels could be identified in a meta-analysis of East Asian and multi-ancestry sample groups. We also aimed to determine whether the protein-altering variants in known

Most genome-wide association studies have been of European individuals, even though most genetic variation in humans is seen only in non-European samples. To search for novel loci associated with blood lipid levels and clarify the mechanism of action at previously identified lipid loci, we used an exome array to examine protein-coding genetic variants in 47,532 East Asian individuals. We identified 255 variants at 41 loci that reached chip-wide significance, including 3 novel loci and 14 East Asian–specific coding variant associations. After a meta-analysis including >300,000 European samples, we identified an additional nine novel loci. Sixteen genes were identified by protein-altering variants in both East Asians and Europeans, and thus are likely to be functional genes. Our data demonstrate that most of the low-frequency or rare coding variants associated with lipids are population specific, and that examining genomic data across diverse ancestries may facilitate the identification of functional genes at associated loci.
lipid loci explained the association signal or were independent evidence of functional genes. Finally, we examined whether exome data implicated the same putatively functional genes at lipid loci in both European and East Asian cohorts.

RESULTS
To improve the coverage for the low-frequency variants in Asian populations and follow up on various GWAS variants, we added approximately 60,000 custom-content variants to the standard exome array. Among 319,272 variants that passed quality control, 204,408 (64.0%) were polymorphic in East Asian individuals, of which about 25% (50,126) were from the custom content. Approximately 76.1% (155,566) of the polymorphic variants were annotated as nonsynonymous or loss-of-function (stop-gain, stop-loss, and splice variants) (Supplementary Table 2).

By determining the proportion of variants observed in Exome Aggregation Consortium (ExAC) East Asian samples (n = 4,327 individuals) that were successfully genotyped by the array, we estimated that the exome array capture a large fraction of common and low-frequency coding variants (71.15% and 72.59% for variants with minor allele frequency (MAF) > 5% and MAF = 1–5%, respectively). Among rare coding variants identified in ExAC sequenced individuals, 59.91% (MAF = 0.1–1%) and 19.92% (two or more copies) were captured by the array. Therefore, the array provided good coverage for low-frequency variants and moderate coverage for rare coding variants in East Asians. In addition, we examined 76,000 polymorphic coding variants that were unavailable or monomorphic in ExAC East Asian samples.

Discovery of novel variants associated with lipid levels
Our analysis identified three variants with study-wide significance in three novel loci in East Asians, located at least 1 Mb from previously reported GWAS signals of lipid levels (Table 1). These were rs4377290 in ACVR1C (TC; $P = 4.69 \times 10^{-8}$), rs7901016 in MCU (LDL-C; $P = 5.12 \times 10^{-9}$), and the missense variant rs4883263 (encoding p.Ile342Val) in CD163 (HDL-C; $P = 5.24 \times 10^{-11}$). Each of these three variants demonstrated evidence for association ($P = 1.80 \times 10^{-3}$ to $6.68 \times 10^{-5}$) in over 300,000 GLGC individuals.

Summary of association results
We assessed the association of 110,986 polymorphic variants that had at least 20 minor alleles in 47,532 East Asian samples. Overall, we detected 255 variants (including 51 coding variants) at 41 loci with exome-wide significant association with one or more lipid traits ($P < 4.5 \times 10^{-7}$), of which 3 loci had not been previously reported (Fig. 1). Collectively, the overall variance in each lipid trait that could be explained by exome-wide significant variants in East Asian samples was 5.97% for TC, 6.20% for LDL-C, 6.93% for HDL-C, and 6.89% for TG levels, of which 3.22%, 4.77%, 3.35%, and 3.86%, respectively, could be attributed to coding variants (Fig. 2). Our results also showed that an additional seven known loci were associated with lipid levels with suggestive significance ($P < 4.46 \times 10^{-6}$, Bonferroni correction of 11,215 variants) (Supplementary Table 2), and that, taken together, they increased the percentage of trait variance explained to 6.08–7.20%.

Evaluation of known lipid signals
Among the 38 previously established lipid loci that reached significance, we identified a more significant candidate variant at 14 loci (Supplementary Table 3 and Fig. 1) where the initially reported GWAS index variants showed no significant associations or were independent of our lead variants ($r^2 < 0.02$) (APOB and APOE), demonstrating allelic heterogeneity between people of East Asian ancestry and those of European ancestry. The lead variants in the remaining 24 loci were the same as or strongly related to ($r^2 > 0.69$) the reported GWAS index variants from previous studies in primarily European samples. Sequential conditional analyses showed that 12 loci with evidence of association had two or more significant signals (Supplementary Table 4).

For example, we detected a novel missense variant (rs2075260, encoding p.Val2141lle) at ACACB that was largely independent of the originally reported GWAS index variant rs7134594 at MYK ($r^2 = 0.01$), and thus represented a previously unreported association. The GWAS index variant rs7134594 could be explained by another missense variant (rs9593, encoding p.Met239Lys) at MMAB (conditional $P = 0.73$).

In gene-based analysis, nine genes (PCSK9, EIV5, HMGCR, CD36, APOA1, PCSK7, CETP, LDLR, and PPARA) reached gene-based significance ($P < 2.8 \times 10^{-6}$) in connection with lipid levels (Supplementary Fig. 1 and Supplementary Table 5). However, our gene-based analyses did not identify any new genes that had not already been highlighted by single-variant tests.

Putative functional coding variants at known loci
The identification of coding variants in known loci has the potential to pinpoint causal genes. We observed that the protein-altering variants were more likely to have strong effect sizes with regard to lipid levels (Fig. 2 and Supplementary Table 6) compared with the non-coding variants that were significantly associated with lipid levels. Ten coding variants in eight genes showed strong effects on lipid levels ($\beta$-coefficients ranging from 0.20 to 1.17 s.d.), and eight were low-frequency or rare variants (MAF < 5%). We next sought to quantify what proportion of GWAS loci might be due to a protein-altering variant, and thus implicate a candidate functional gene. We made the reasonably well-supported assumption that a protein-altering variant that is the top signal, explains the signal, or is independent of the original signal is the most likely causal variant for each region $^{15-17}$. Among the 38 known loci for which association evidence attained study-wide significance, 12 loci harbored a protein-altering variant that showed the strongest association with lipid levels, and 4 loci had a protein-altering variant that was not the top signal but could explain the association of the reported index variant (Supplementary Table 7 and Fig. 1). In 8 of these 16 loci (PCSK9, EIV5, CD36, MMAB, ALDH2, SLC12A4, LDLR, and PPARA), the previously identified lead variants in European populations did not reach exome-wide significance.

| Gene       | rsID           | Position     | Alleles | Variants   | Trait | East Asian | GLGC | Combined |
|------------|---------------|--------------|---------|------------|-------|------------|------|----------|
| ACRVR1C    | rs4377290     | 2:158437683  | C/T     |            | TC    | 0.33       | 0.007| 4.69 \times 10^{-8} | 46,025 |
|            |               |              |         |            |       | 0.46       | 1.59 \times 10^{-4} | 0.44 | 6.06 \times 10^{-8} | 16.2  |
| MCU        | rs7901016     | 10:7463732   | C/T     |            | LDL-C | 0.27       | 0.008| 5.12 \times 10^{-9} | 44,985 |
|            |               |              |         |            |       | 0.09       | 1.80 \times 10^{-3} | 0.12 | 2.21 \times 10^{-9} | 18.2  |
| CD163      | rs4883263     | 12:7649484   | C/T     | p.Ile342Val| HDL-C | 0.69       | 0.007| 6.24 \times 10^{-11} | 47,456 |
|            |               |              |         |            |       | 0.94       | 6.68 \times 10^{-5} | 0.90 | 6.30 \times 10^{-13} | 2.38  |

AAF, alternative allele frequency. Positions are reported in human genome build hg19. Alleles are listed as alternative/reference alleles on the forward strand of the reference genome.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
In the remaining eight loci (GCKR, MLXIPL, HNF1A, LPL, ABO, GPAM, PMFBP1, and TM6SF2), the GWAS index variant in each locus (P values ranged from \(4.86 \times 10^{-8}\) to \(1.26 \times 10^{-62}\)) was in strong linkage disequilibrium (LD) with the corresponding protein-altering variant \(\left(r^2 > 0.67\right)\) and did not remain significant after the effect of the protein-altering variant was accounted for (conditional P values > 0.01), which suggested that the index variant might act as a proxy for the functional protein-altering variant. Together, 42.1% (16/38) of loci seemed to have a protein-altering variant that could account for the original association signal. In addition, we identified 15 protein-altering variants in nine genes (APOB, HMGCR, ABCA1, APOA1–APOA5, ACACB, CETP, PKD1L3, LIPC, and APOE) that were independent of
the original signal but may highlight functional genes in the region. All of these putative functional variants may point to functional candidate genes—either well-established causal genes (such as the genes that cause Mendelian dyslipidemias (Supplementary Table 8)) or potential new candidate genes (MMAB, ACACB, SLC12A4, and PMFBP1). In total, the 31 protein-altering variants in the known loci may point to 25 candidate functional lipid genes.

**Association with coronary artery disease**

To further evaluate whether the novel variants and putative functional variants in known regions identified in our samples also influenced CAD risk, we tested for association in 28,899 Chinese individuals with and without coronary disease (9,661 CAD cases and 18,558 controls) and in the largest publicly available CAD GWAS analysis (CARDIoGRAMplusC4D), which includes ~185,000 CAD cases and controls (Supplementary Table 9). For the novel noncoding variant near MCU (rs7901016), the C allele associated with lower levels of LDL-C was similarly associated with reduced risk for CAD in Chinese samples (odds ratio (OR) = 0.94; 95% confidence interval (CI) = 0.90–0.98; P = 2.8 × 10^{-5}) and CARDIoGRAMplusC4D samples (OR = 0.94; 95% CI = 0.91–0.98; P = 4.35 × 10^{-5}). Among the 31 putative functional coding variants in the known regions, 20 non-HDL-C-related variants showed a consistent direction of effect between lipid traits and CAD. Fifteen out of 20 showed nominal significance (P < 0.05) in Chinese or CARDIoGRAMplusC4D CAD data, whereas 7 variants in PCSK9, APOB, LDLR, APOE, HNF1A, and APOA5 showed significant associations even after multiple testing was accounted for (P values ranged from 5.95 × 10^{-4} to 8.17 × 10^{-11} < 0.05/31). In particular, nearly all of the LDL-associated coding variants demonstrated association with CAD, and the strengths of effect on CAD risk and LDL-C level were strongly correlated (r^2 = 0.78; P = 3.3 × 10^{-4}; Supplementary Fig. 2).

**Novel loci identified in East Asian and GLGC samples**

An exome-wide association screen for plasma lipids in >300,000 individuals genotyped by exome array was conducted in parallel by the GLGC14. The majority (84%) of the participants were of European ancestry, and only 2.3% were of East Asian ancestry. We further carried out a large-scale trans-ancestry meta-analysis of our East Asian and GLGC samples, being careful to include overlapping samples only once, to seek both novel and population-specific genetic variants for lipid levels.

In the combined GLGC and East Asian samples, nine additional variants that were not significant in the East Asian or GLGC analyses showed significant association (P < 2.1 × 10^{-7}), Bonferroni correction of 242,289 variants analyzed by the GLGC) with at least one lipid trait. All of them were common (MAF > 0.05 in both East Asian and GLGC samples), including four coding variants (Table 2 and Supplementary Fig. 3): FAM114A2 (p.Gly122Ser; HDL-C; P = 1.74 × 10^{-7}), MGAT1 (p.Leu435Pro; HDL-C; P = 9.36 × 10^{-9}), ASCC3 (p.Leu146Phe; LDL-C; P = 5.84 × 10^{-8}; TC, P = 5.22 × 10^{-9}), and PLCE1 (p.Arg1575Pro; TC; P = 9.92 × 10^{-8}).

**Joint analysis of novel signals with additional samples**

To strengthen support for the observed associations, we carried out in silico replication of significant variants in three additional independent genome-wide data sets comprising a combined total of ~160,000 individuals from the Nord-Trøndelag Health Study19, GLGC GWAS samples2, and a Chinese lipids GWAS20. We found that the associations of 12 novel variants achieved greater significance...
### Table 2: Variants at novel loci associated with lipid levels identified from combined East Asian and GLGC samples

| Gene     | rsID   | Position | Alleles | Variants | Trait | AAF (% | β  | S.e.m. | Combined | GLGC | East Asian |
|----------|--------|----------|---------|----------|-------|--------|-----|--------|----------|-------|------------|
| PDGFC    | rs4691380 | 4:157720124 | T/C     | p.Gly122Ser | HDL-C | 0.35   | 0.014 | 0.003 | 1.07 × 10⁻⁷ | 335,481 | 0.54 |
| FAM114A2 | rs2578377 | 5:153413390 | T/c     | p.Leu435Pro | HDL-C | 0.67   | -0.014 | 0.003 | 1.74 × 10⁻⁷ | 335,484 | 4.25 |
| MGA71    | rs634501  | 5:180218668 | G/A     | p.Leu146Phe | HDL-C | 0.72   | -0.015 | 0.003 | 9.36 × 10⁻⁸ | 337,027 | 1.70 |
| ASCC3    | rs9390698 | 6:101296389 | A/G     | p.Leu146Phe | HDL-C | 0.39   | 0.014  | 0.003 | 5.84 × 10⁻⁸ | 331,991 | 0.40 |
| C6orf183 | rs8844366 | 6:109574095 | A/G     | HDL-C     |       | 0.31   | -0.015 | 0.003 | 1.45 × 10⁻⁸ | 327,673 | 1.70 |
| EEPD1    | rs4302748 | 7:36191699  | A/G     | LDL-C     |       | 0.18   | 0.018  | 0.003 | 2.10 × 10⁻⁸ | 333,359 | 0.43 |
| PLCE1    | rs7306523 | 10:96039597 | C/G     | p.Arg1575Pro | TC    | 0.44   | -0.020 | 0.004 | 9.92 × 10⁻⁸ | 150,798 | 1.73 |
| EIF4B    | rs7965082 | 12:100800193| T/C     | LDL-C     |       | 0.52   | -0.013 | 0.002 | 9.21 × 10⁻⁸ | 333,359 | 0.00 |

AAF, alternative allele frequency. Positions are reported in human genome build hg19. Alleles are listed as alternative/reference alleles on the forward strand of the reference genome.

### Table 3: Inter-ancestry allelic heterogeneity at lipid genes

| Gene     | Study | rsID   | Note | Position | Variants | Alleles | Trait | β  | S.e.m. | Combined | GLGC | East Asian |
|----------|-------|--------|------|----------|----------|---------|-------|-----|--------|----------|-------|------------|
| PCSK9    | GLGC  | rs11591147 | Protein-altering is top | 1:55505647 | p.Arg46Leu | T/G | -0.475 | 0.011 | 0.00 | 7.62 × 10⁻⁷ | 28.44 | 0.43 |
| APOB     | GLGC  | rs1367117 | Explaining index | 2:21263900 | p.Thr98Ile | A/G | 0.105 | 0.003 | 3.61 × 10⁻²⁷⁸ | 42.49 | 0.41 |
| CD36     | GLGC  | rs3211938 | Protein-altering is top | 7:80300449 | p.Tyr325* | G/T | 0.181 | 0.021 | 1.43 × 10⁻¹⁸ | 0.41 | 0.01 |
| ABCA1    | GLGC  | rs146292819 | Independent of index | 9:107556776 | p.Asn1800His | G/T | -0.843 | 0.059 | 3.99 × 10⁻⁴⁶ | 60.97 | 0.10 |
| CETP     | GLGC  | rs5880  | Independent of index | 16:72162966 | p.Arg368Trp | G/T | 0.342 | 0.058 | 3.17 × 10⁻⁹ | 31.06 | 0.08 |
| LIPG     | GLGC  | rs4832584 | Independent of index | 18:4709955 | p.Arg593Gly | G/A | 0.407 | 0.025 | 7.53 × 10⁻⁶ | 2.23 | 0.02 |
| LDLR     | GLGC  | rs19043155 | Independent of index | 19:11217344 | p.Asp257Glu | A/T | 0.644 | 0.024 | 1.62 × 10⁻⁸ | 15.93 | 0.01 |
| PPARA    | GLGC  | rs1042311 | Protein-altering is top | 22:46627780 | p.Val227Ala | C/T | 0.123 | 0.018 | 7.40 × 10⁻¹² | 0.50 | 0.01 |

AAF, alternative allele frequency. Positions are reported in human genome build hg19. Alleles are listed as alternative/reference alleles on the forward strand of the reference genome. Protein-altering is top: protein-altering variants are the most significant variants in the known loci. Explaining index: conditional on the coding variants; adjusted P for index variants > 0.01. Independent of index: conditional on the index variants; adjusted P for coding variants with exome-wide significance.
Table 4. Loci for which East Asian and GLGC samples identified the same putatively functional protein-altering variant

| Gene   | rsID       | Position | Variant | Alleles | Trait    | Study         | AAF      | S.e.m. | P     | β        | Variance explained (%) |
|--------|------------|----------|---------|---------|----------|---------------|----------|--------|-------|----------|------------------------|
| MLXIP1 | rs3332026  | 2:27730940 | p.Leu446Pro | C/T     | TG       | GLGC          | 0.628    | 0.003  | 0.00  | 1.26 × 10^-02 | 0.003                  |
| GCKR   | rs3126525  | 7:73012042 | p.Ala358Val | A/G     | TG       | GLGC          | 0.496    | 0.007  | 1.26 × 10^-06 | 0.011                  |
| MLXIPL | rs35332062 | 7:73012042 | p.Ala358Val | A/G     | TG       | GLGC          | 0.099    | 0.011  | 1.26 × 10^-04 | 0.011                  |
| GCKR   | rs328      | 8:19819724 | p.Ser474*  | G/C     | TG       | GLGC          | 0.039    | 0.003  | 5.22 × 10^-03 | 0.011                  |
| LDLR   | rs1169288  | 12:121416650 | p.Ile27Leu | C/A     | TC       | GLGC          | 0.706    | 0.007  | 5.99 × 10^-04 | 0.007                  |
| CETP   | rs58542926 | 19:19379549 | p.Glu167Lys | T/C     | TC       | GLGC          | 0.333    | 0.012  | 4.86 × 10^-10 | 0.007                  |
| APOE   | rs7412     | 19:45412079 | p.Arg176Gly | A/G     | TC       | GLGC          | 0.070    | 0.013  | 4.75 × 10^-07 | 0.007                  |

Coding variants point to the same genes across ancestries

We further evaluated whether the variants identified in East Asian samples were also defined as putative functional variants in GLGC samples (Supplementary Table 11). We found that East Asian and GLGC samples both pointed to the same nine functional genes, but that different associated variants were present in each ancestry (Table 3). The eight coding variants (MAF, 0.004–15.9%) at PCSK9, CD36, ABCA1, CETP, PMFBP1, LIPG, LDLR, and PPARA identified by GLGC showed lower MAFs (0–2.57%) in the East Asian samples and thus achieved no or only suggestive significance (CETP). Conversely, the coding variants at PCSK9, APOB, CD36, CETP, LDLR, and PPARA identified in East Asian samples (MAF, 0.094–12.45%) also had lower MAFs in GLGC samples (0.001–0.20%). In addition, the same putatively functional coding variants and genes at seven loci (GCKR, MLXIP1, LPL, GPAM, HNF1A, TM6SF2, and APOE) were identified in both East Asian and GLGC samples, with similar common MAFs (Table 4).

**East Asian–specific association signals**

We next attempted to identify variants that were associated with lipids in East Asian samples only. Among the known lipid loci, we identified 363 independent variants by sequential conditional analysis in GLGC exome-wide association studies (Supplementary Table 11). After conditioning on the independent variants in the corresponding loci, we identified 14 independent coding variant associations at 11 loci in East Asian samples with conditional P values < 4.5 × 10^-7 (Table 5, Figs. 1 and 3). All 14 East Asian–specific variants were included in the list of putative functional variants that we identified. Eight of these loci (EV15, APOB, HMGCR, CD36, APOA1, CETP, LDLR, and PPARA) harbored at least one low-frequency or rare independent coding variant (MAF, 4.21–0.03%). All of these variants either were monomorphic or had a frequency that was at least one order of magnitude lower in Europeans and thus showed only suggestive significance in ~300,000 GLGC individuals.

**DISCUSSION**

This study represents a large discovery effort to identify coding variation that influences lipid levels in the East Asian population, and it enabled us to systematically evaluate protein-altering variants that identify candidate functional genes. Meta-analyses of East Asian and multi-ancestry samples by exome-chip genotyping array identified 12 novel loci, 5 of which harbored nonsynonymous variants. In the 38 known loci that were replicated, we identified 31 protein-altering variants pointing to 25 functional lipid genes. Moreover, significant association with protein-altering variants identified the same 16 putative functional genes in European and East Asian samples, and 9 of those genes were identified by independent protein-altering variants in the two ancestries.

Among the novel genetic loci identified, several have been implicated in cardiovascular and metabolic phenotypes, which may provide mechanistic insight into the regulation of lipid levels and potential targets for treatment. The significant novel variant associated with both lipids and CAD is located in an intron of MCU. MCU encodes a mitochondrial inner membrane calcium uniporter that mediates calcium uptake into mitochondria. Mitochondrial calcium has an important role in the regulation of metabolism in the heart. CD163 encodes a macrophage-specific receptor involved in the clearance and endocytosis...
Table 5  East Asian–specific variants associated with blood lipids (conditional P < 4.5 × 10^{-7})

| Gene      | Position | miD         | Allele   | Variant    | Test | Adj AAF (%) | S.e.m. | P            | β        | S.e.m. | P          |
|-----------|----------|-------------|----------|------------|------|-------------|--------|--------------|----------|--------|------------|
| EVS       | 1:93156972 | rs17171462   | AG       | p.Arg852Cys | TC   | 0.69        | 0.040  | 1.41 × 10^{-12} | 8.44 × 10^{-10} |
| ADIPOQ    | 2:212128637 | rs35069148   | G/A      | p.Ile3768Thr | TC   | 0.15        | 0.097  | 8.44 × 10^{-12} | 9.96 × 10^{-12} |
| LDL-C     | 2:21252807  | rs2237662    | T/C      | p.Cys478Tyr | TC   | 0.09        | 0.088  | 3.35 × 10^{-9}  | 4.44 × 10^{-9}  |
| LDL-C     | 2:21252534  | rs13306194   | A/G      | p.Arg532Trp | TC   | 12.39       | 0.010  | 9.53 × 10^{-22} | 2.08 × 10^{-13} |
| TG        | 2:21252534  | rs13306194   | A/G      | p.Arg532Trp | TC   | 12.45       | 0.010  | 9.53 × 10^{-22} | 2.08 × 10^{-13} |
| LDL-C     | 5:74466766  | rs19183691    | C/A      | p.Tyr311Ser | LDL-C | 1.73    | 0.026  | 2.20 × 10^{-13} | 2.68 × 10^{-9}  |
| HMGCR     | 7:80302116  | rs148910227   | T/C      | p.Arg386Trp | HDL-C | 0.31       | 0.058  | 3.17 × 10^{-9}  | 3.60 × 10^{-9}  |
| APOA1     | 11:116707736 | rs12718465   | T/C      | p.Ala61Thr | HDL-C | 3.27       | 0.058  | 5.50 × 10^{-10} | 1.41 × 10^{-7}  |
| APOB      | 12:109696838 | rs2075260    | A/G      | p.Val2141Ile | TG    | 74.34      | 0.008  | 3.95 × 10^{-8}  | 7.64 × 10^{-8}  |
| ACACB     | 12:112241766 | rs671       | A/G      | p.Glu457Lys | HDL-C | 20.43      | 0.008  | 1.16 × 10^{-8}  | 1.85 × 10^{-8}  |
| ALDH2     | 16:57017292  | rs2303790    | G/A      | p.Asp459Gly | HDL-C | 2.23       | 0.025  | 7.53 × 10^{-62} | 1.89 × 10^{-31} |
| CETP      | 16:71967927  | rs17358402   | T/C      | p.Arg1572His | LDL-C | 5.40       | 0.015  | 2.11 × 10^{-8}  | 1.86 × 10^{-9}  |
| LDLR      | 19:111217315 | rs10009925   | T/C      | p.Arg257Trp | TC    | 0.09       | 0.032  | 6.74 × 10^{-3}  | 2.96 × 10^{-5}  |
| PPARA     | 19:111217315 | rs10009925   | T/C      | p.Arg257Trp | TC    | 0.09       | 0.032  | 6.74 × 10^{-3}  | 2.96 × 10^{-5}  |
| MGAT1     | 19:111217315 | rs10009925   | T/C      | p.Arg257Trp | TC    | 0.09       | 0.032  | 6.74 × 10^{-3}  | 2.96 × 10^{-5}  |

Genes included are listed as terminal variants among the independent variants in the corresponding loci identified by GLGC exome-wide association studies (see Supplementary Table 11). All selected loci were associated with lipid levels in an East Asian lipids GWAS meta-analysis published after our manuscript was submitted25. To further clarify the possible transcriptional mechanisms underlying the identified loci in association with lipids, we investigated the relationships of the novel variants and proxies with expression quantitative trait loci (eQTLs) by using the Genotype-Tissue Expression (GTEx) eQTL browser. We found significant cis-eQTL effects in human tissues at five loci at P < 4.5 × 10^{-7} (Supplementary Table 12). We further predicted putatively regulatory variants in seven novel non-coding regions in 81 cell lines on the basis of deltaSVM scores26, and found that the variants in PDGFC, C6orf183, and MCUR had high regulatory potential with extreme deltaSVM scores greater than 10 in absolute value (Supplementary Fig. 4).

Our data allow a more comprehensive understanding of the genetic architecture of lipid susceptibility by revealing novel lipid genes and identifying allelic heterogeneity across populations of different ancestries. We detected multiple independent association signals or new lead variants in known lipid-associated loci that frequently showed no or moderate LD with the corresponding GWAS index variants in European populations. Specifically, we identified 14 East Asian–specific variants that could not be explained by all the independent variants in the corresponding loci identified in GLGC samples. Our study demonstrated the benefits of distinct LD patterns between ancestry groups for the investigation of validated loci. We also found substantial inter-ancestry differences in the identification of rare coding variants across populations, which may have been subjected to natural selection during human evolution or genetic drift. All the low-frequency or rare functional coding variants identified in East Asians (MAF, 0.03–4.21%) appeared to be population specific, and were monomorphic or not present in European individuals who were part of the 1000 Genomes Project; this allelic heterogeneity across populations of different ancestry has been reported in part6.11. However, we observed that these rare variants were not monomorphic in more than 300,000 GLGC individuals, but had 15-fold to 160-fold lower frequencies (MAF, 0.001–0.15%) in Europeans than in East Asians (Supplementary Table 13 and Supplementary Fig. 5), with little power to indicate association in Europeans. Similarly, the low-frequency and rare coding variants identified in GLGC samples were extremely rare or monomorphic in East Asian samples (Supplementary Fig. 6 and Supplementary Table 11). Overall, our findings demonstrate that rare and low-frequency coding variants are more likely to be population specific, which underscores the value of discovering ancestry-specific rare variants in diverse populations, particularly for low-frequency variations.
As most GWAS index variants are located in noncoding regions, the identification of associated protein-coding variants may allow scientists to prioritize functional genes and variation. Among the 38 known loci that reached chip-wide significance in our data, coding variants at 16 loci (42.1%) were found to completely account for the original association signal. At an additional nine loci, an independent protein-altering variant indicated a likely functional gene. The coding variants were more likely to have consistent effect sizes across ethnic groups than noncoding variants were. For the GWAS index variants that could not be replicated in East Asian samples, the effect sizes were poorly correlated with those observed in Europeans. In contrast, the effect sizes of the putatively functional coding variants in the same loci were strongly related across ethnic groups (Supplementary Fig. 7). Trans-ancestry comparisons provided additional credible evidence to support the identification of the same 16 genes as putative functional genes. The functional genes pointed to by coding variants were either well-known genes or genes with previously unknown roles in lipid metabolism (such as GPAM and PMFBP1), which may be good candidates for functional assessment. More importantly, we found that the effects of these putative functional coding variants on levels of LDL-C, TG, and TC were highly correlated with the effect on CAD, but the effects on HDL-C levels were not correlated with CAD. Our findings are in agreement with recent genetic studies showing that both LDL-C and TG levels, but not HDL-C levels, are causally related to CAD risk27–30.

This large-scale exome-wide association study allowed us to detect a greater number of low-frequency and rare variants than previously identified, 30% of which were not polymorphic in an earlier exome-wide study involving 12,685 Chinese individuals11. Nonetheless, the exome array offered moderate coverage for rare variants observed in ExAC East Asian samples. Power calculations indicated that the available sample size provided 80% power to detect variants with an effect size of 0.27 s.d. and MAFs as low as 0.5% at a sample size provided 80% power to detect variants with an effect size of 0.27 s.d. and MAFs as low as 0.5% at 80% power to detect variants with an effect size larger than 5 × 10−7. However, we had considerably less power to evaluate extremely rare variants (MAF < 0.1%). Studies with larger sample sizes and of sequenced genomes are therefore needed to fully investigate associations of rare variants with lipid levels.

In conclusion, we identified 12 new loci associated with lipid levels. We also identified coding variants that highlight 25 likely functional genes at previously known loci, including several with previously undiscovered roles in lipids. We also found an abundance of population-specific coding variant associations that underlie lipid traits, highlighting the importance of including individuals of diverse ancestral backgrounds. At the same time, our data demonstrate that the integration of genomic data across diverse ancestral groups may enable researchers to identify functional variants and genes for further functional study.

URLs. Genotype-Tissue Expression (GTEx) Portal, http://www.gtexportal.org/home; Genezoom, http://genome.sph.umich.edu/wiki/Genezoom; ExAC, http://exac.broadinstitute.org; RareMETALS, http://genome.sph.umich.edu/wiki/RareMETALS; RVTESTS, http://genome.sph.umich.edu/wiki/RvTests; RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all the participants of this study for their contributions. X. Lu is supported by the CAMS Innovation Fund for Medical Sciences (grants 2016-12M-1-009, 2017-12M-1-004, and 2016-12M-1-011) and the National Science Foundation of China (grants 81422043, 91439202, 81770002, 81773537, and 81230069). C.J.W. is supported by the National Institutes of Health (grant HL135824). S.K. and C.J.W. are supported by the National Institutes of Health (grant HL127564). P.C.S. was supported by the Hong Kong Research Grants Council (grant TRS T12/705/11 and GFR 1712851). G.M.P. is supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (award K01HL125751). We thank P. Marshall for professional editing. Additional acknowledgments of funding sources for the primary studies are provided in Supplementary Note 1.

AUTHOR CONTRIBUTIONS

X. Lu, C.J.W., G.M.P., D.J.L., D.G., and K.L.M. drafted the manuscript. C.J.W., D.G., X. Lu, P.C.S., S.K., K.L.M., and Y.E.C. coordinated the project. X. Lu, D.J.L., G.M.P., and H.Z. served as the central meta-analysis group. X. Lu and J.B.N. carried out eQTL analysis. X. Lu and W. Zhou carried out DeltaSV analysis. X. Lu, G.M.P., D.J.L., Y. Wu, H.Z., J. Li, C.S.T., R.D., J. Long, X.G., C.N.S., Y.C., Y. Wang, C.Y.C., Q.L., Y. Xu, Y. Zhao, M.H., and J.B.N. carried out exome-sequence data analysis. W. Zhou, H.L., C.C.K., J. Li, L.W., F.W., and W. carried out cohort genotyping. H.L., M.X., X. Liu, Y.Z., L.S., Y.G., Y. Hu, K.Y., J.H., Q.C., S.C., A.B.F., L.S.A., P.L., S.D., K.H., and L.G.F. carried out cohort phenotyping. X. Lu, W.H.-H.S., S.S.C., A.B.F., L.S.A., P.L., S.D., R.Y., Y.-D.I.C., X.-O.S., K.S.L.L., T.Y.W., S.K.G., Z.M., K.H., L.G.F., H.T., Y. Huo, C.Y.C., E.W.C., Zheng, E.S.T., WG., X. Lin, W.H., S.K., K.L.M., T.W., P.C.S., D.G., and C.J.W. were the principal investigators for the cohort.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Teslovich, T.M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713 (2010).
2. Willer, C.J. et al. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013).
3. Surakka, I. et al. The impact of low-frequency and rare variants on lipid levels. Nat. Genet. 47, 589–597 (2015).
4. Asselbergs, T.F. et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. Am. J. Hum. Genet. 91, 823–838 (2012).
5. Kim, Y.J. et al. Large-scale genome-wide association studies in East Asians identify novel lipid loci influencing body mass index. Nat. Genet. 43, 990–995 (2011).
6. Wu, K. et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. 9, e1003379 (2013).
7. Lambert, N.G. et al. Risk factors and biomarkers of age-related macular degeneration. Prog. Retin. Eye Res. 54, 64–102 (2016).
8. Herder, C., Kowail, B., Tabak, A.G. & Rathmann, W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. Diabetologia 57, 16–29 (2014).
9. Wierzbicki, A.S. & Oben, J. Nonalcoholic fatty liver disease and lipids. Curr. Opin. Lipidol. 23, 345–352 (2012).
10. Kathiresan, S. ACSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. N. Engl. J. Med. 358, 2299–2300 (2008).
11. Tang, C.S. et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. Nat. Commun. 6, 10206 (2015).
12. Peloso, G.M. et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. Am. J. Hum. Genet. 94, 223–232 (2014).
13. Lange, L.A. et al. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. Am. J. Hum. Genet. 94, 233–245 (2014).
14. Liu, D.J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nat. Genet. http://dx.doi.org/10.1038/ng.3977 (2017).
15. Holmen, O.L. et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat. Genet. 46, 345–351 (2014).
16. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J.A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324, 387–389 (2009).
Correspondence should be addressed to C.W. (cristen@umich.edu), D.G. (gudongfeng@vip.sina.com) or P.C.S. (pcsham@hku.hk).

1Department of Epidemiology, State Key Laboratory of Cardiovascular Disease, Fujwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. 2Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA. 3Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. 4Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. 5Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. 6Department of Public Health Sciences, Institute of Personalized Medicine, Penn State University, University Park, Pennsylvania, USA. 7Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. 8Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA. 9MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, Hubei, China. 10Department of Surgery, Li KaShing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; Dr. Li Dak-Sum Research Centre, The University of Hong Kong–Karolinska Institutet Collaboration in Regenerative Medicine, Hong Kong, China. 11Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. 12Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and University of the Chinese Academy of Sciences, Shanghai, China. 13Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. 14Institute for Translational Genomics and Population Sciences, L&BioMed at Harbor-UCLA Medical Center, Los Angeles, California, USA. 15Department of Cardiology, Institute of Vascular Medicine, Peking University Third Hospital, Key Laboratory of Molecular Cardiovascular Sciences, Ministry of Education, Beijing, China. 16Center for Genomic and Personalized Medicine, Medical Scientific Research Center and Department of Occupational Health and Environmental Health, School of Public Health, Guangxi Medical University, Nanning, Guangxi, China. 17Department of Cardiology, Peking University First Hospital, Beijing, China. 18Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore, Singapore. 19Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. 20Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. 21Department of Medicine, the University of Hong Kong, Hong Kong, China. 22Community Health Center, The 3rd Affiliated Hospital of Shenzhen University, Shenzhen, China. 23Duke-National University of Singapore Graduate Medical School, Singapore, Singapore. 24Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China. 25Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. 26Department of Psychiatry, the University of Hong Kong, Hong Kong, China. 27Centre for Genomic Sciences, Li KaShing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. 28State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China. 29USC-Office of Population Studies Foundation, University of San Carlos, Cebu City, Philippines. 30Department of Anthropology, Sociology, and History, University of San Carlos, Cebu City, Philippines. 31Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA. 32Carolina Population Center, University of North Carolina, Chapel Hill, North Carolina, USA. 33USC Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA. 34Research Centre of Heart, Brain, Hormone and Healthy Aging, Li KaShing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. 35State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China. 36Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore. 37Department of Ophthalmology, National University of Singapore, Singapore, Singapore. 38HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway. 39K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway. 40Department of Medicine, Leverhanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. 41Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. 42Hong Kong–Guangdong Joint Laboratory on Stem Cell and Regenerative Medicine, the University of Hong Kong, Hong Kong, China. 43Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore. 44Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore, Singapore. 45A full list of members and affiliations appears in Supplementary Note 1. 46Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. 47These authors jointly supervised this work. Correspondence should be addressed to C.W. (cristen@umich.edu), D.G. (gudongfeng@vip.sina.com) or P.C.S. (pcsham@hku.hk).
ONLINE METHODS

Study cohorts. Twenty-three studies, including both population-based studies and case-control studies of CAD and type 2 diabetes, were genotyped with the Illumina HumanExome array, resulting in a total of 47,532 participants, all of whom were of East Asian ancestry (Supplementary Table 14). All participants provided written informed consent, and ethics approval for data generation and analyses was individually obtained for each contributing study. The relevant human genetic data were also approved by the Ministry of Science and Technology of China. For the GLGC exome study, 95 studies contributed association results for exome chip genotypes and plasma lipid levels (Supplementary Note 1 and Supplementary Table 15).

Phenotypes. For most East Asian subjects (86%), TC, HDL-C, and TGs were measured at >8 h of fasting. LDL-C levels were directly measured in 18 studies (88% of total study individuals) and were estimated via the Friedewald formula in the remaining studies, with missing values assigned to individuals with >400 mg/dl TGs. We adjusted the TC values for individuals on lipid-lowering medication by replacing their TC values by TC/0.8 with lipid medication status available. If measured LDL-C was available in a study, the treated LDL-C value was divided by 0.7. No adjustment for individuals using medication was made for HDL-C or TG.

Exome array genotyping and quality control. All study participants were genotyped on the HumanExome Bead-Chip (Illumina), and most samples (83%) also included the custom Asian Vanderbilt content. This custom content was added to the standard Illumina HumanExome Bead-Chip to improve the coverage of low-frequency variants in Asian populations. The variants were selected from 1,077 (581 female Chinese subjects and 496 male Singapore Chinese) whole-exome–sequenced East Asian samples generously provided by W. Zheng and J. Liu31. Approximately 29,000 additional common variants were added to the array, including previously identified GWAS variants selected from the GWAS catalog. Genotype calling was done with GenTrain version 2.0 in GenomeStudio V2011.1 (Illumina) in combination with zCall version 2.2 (ref. 32). Within each study, individuals with low genotype completion rates, individuals expressing gender mismatches or a high level of heterozygosity, related individuals, and PCA outliers were excluded from further analysis (Supplementary Table 16). In addition, variants that did not meet the 95% or 98% genotyping threshold or that showed deviation from Hardy–Weinberg equilibrium were removed.

Statistical analyses. For each cohort, HDL-C, LDL-C, TC, TG, and TC measurements were transformed via the inverse normal distribution after adjustment of each trait for age, sex, and study–specific covariates, including principal components to account for population structure. In studies on diabetes or cardiovascular disease status, cases and controls were analyzed separately.

We performed two single-variant and gene-level association tests. Single-variant analyses in each cohort were carried out with either RAREMETALWORKER or RVTESTS33, both of which generate single-variant score statistics and their covariance matrix between single-marker statistics. The joint effects of variants in a locus were approximated by the proportion of variance explained by the set of independently associated variants. Joint effects of variants in a locus were approximated by \( \hat{\beta}_{\text{joint}} = \frac{\chi^2_{\text{meta}}}{\hat{V}_{\text{meta}}} \), where \( \chi^2_{\text{meta}} \) represents single-variant score statistics and \( \hat{V}_{\text{meta}} \) is the covariance matrix between them. The covariance between single-variant genetic effects was approximated by the inverse of the variance–covariance matrix of score statistics, that is, \( \hat{V}_{\text{meta}}^{-1} \). The phenotype variance explained by independently associated variants in a locus was given by \( \hat{\beta}_{\text{joint}} \), which is equal to the joint effect size divided by the square root of the number of independent associations.

In silico replication samples. The in silico replication study was conducted with data from additional independent individuals of European ancestry from the Nord–Trøndelag Health Study (HUNT)39 and GLGC GWAS35, and Chinese subjects from a Chinese lipids GWAS20. HUNT encompasses a population-based cohort of 62,168 individuals with genome-wide genotypes (Illumina Human CoreExome), imputation from the Haploype Reference Consortium panel, and non-fasting lipid phenotypes. The Chinese lipids GWAS was a meta-analysis of over 13,000 Han Chinese who underwent standardized blood lipid measurements in four independent GWASs. These studies included CAS, the Beijing Atherosclerosis Study, the Genetic Epidemiology Network of Salt Sensitivity study30, and CAS phase II.

Heritability and estimated proportion of variance explained. We estimated the proportion of variance explained by the set of independently associated variants. Joint effects of variants in a locus were approximated by \( \hat{\beta}_{\text{joint}} = \frac{\chi^2_{\text{meta}}}{\hat{V}_{\text{meta}}} \), where \( \hat{V}_{\text{meta}} \) represents single-variant score statistics and \( \chi^2_{\text{meta}} \) is the covariance matrix between them. The covariance between single-variant genetic effects was approximated by the inverse of the variance–covariance matrix of score statistics, that is, \( \hat{V}_{\text{meta}}^{-1} \). The phenotype variance explained by independently associated variants in a locus was given by \( \hat{\beta}_{\text{joint}} \), which is equal to the joint effect size divided by the square root of the number of independent associations.

Annotation. We used ANNOVAR (version 2012-05-25)39 to annotate variants as missense, splice, stop-gain/loss, synonymous, or noncoding. Variant identifiers and chromosomal positions are listed with respect to the hg19 genome build.

DeltaSVM analysis. DeltaSVM uses a gapped k-mer support vector machine to estimate the effect of a variant in a cell-type-specific manner38. DeltaSVM can accurately predict variants associated with DNase I hypersensitivity. Precomputed weights were available from a total of 222 ENCODE samples—99 from the Duke University set, and 123 from the University of Washington set40. For the current study, genetic variants were scored for deltaSVM in 81 cell lines from four tissues (blood, blood vessel, heart, and liver). For each of the seven novel noncoding regions, all proxies (\( r^2 > 0.8 \)) were identified on the basis of data from 1000 Genomes.

Data availability. Summary statistics are available for download from the University of Michigan Center for Statistical Genetics (http://csg.aph.umich.edu/abecasis/public/lipids/2017EastAsian). Additional supporting data are provided in the supplementary material.

A Life Sciences Reporting Summary for this paper is available.

31. Zhang, Y. et al. Rare coding variants and breast cancer risk: evaluation of susceptibility loci identified in genome-wide association studies. Cancer Epidemiol. Biomarkers Prev. 23, 622–628 (2014).
32. Goldstein, I. et al. ZCall: a rare variant caller for array-based genotyping; genetics and population analysis. Bioinformatics 28, 2543–2545 (2012).
33. Liu, D.J. et al. Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
34. Feng, S., Liu, D., Zhan, X., Wing, M.K. & Abecasis, G.R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828–2829 (2014).
35. Lu, X. et al. Coding-sequence variants are associated with blood lipid levels in 14,473 Chinese. *Hum. Mol. Genet.* **25**, 4107–4116 (2016).
36. Lu, X. et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat. Genet.* **44**, 890–894 (2012).
37. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
38. GenSalt Collaborative Research Group. GenSalt: rationale, design, methods and baseline characteristics of study participants. *J. Hum. Hypertens.* **21**, 639–646 (2007).
39. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
40. Thurman, R.E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   Describe how sample size was determined.
   We identified the largest sample size possible of East Asian individuals with exome chip genotyping (N=47,532).

2. Data exclusions
   Describe any data exclusions.
   At the meta-analysis level, no studies were excluded. At the primary study level, exclusions were made using standard quality control criteria: exclusion of samples were made based on sex mismatch, high missing data rate, etc. and exclusion of variants was made based on high missing data rate, Hardy-Weinberg disequilibrium, etc.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   All variants that reached study-wide significance were replicated in three additional cohorts and the statistical evidence for all variants became stronger after including replication data (Supplementary Table 10).

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   All East Asian exome chip samples were analyzed together in the discovery phase. Replication was obtained after the discovery analysis was complete using other datasets (HUNT Europeans GWAS, GLGC European GWAS, Chinese GWAS) as listed in Supplementary Table 10.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Blinding was not relevant to this study.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
| ☑   |           |

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.).
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- A statement indicating how many times each experiment was replicated.
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section).
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons.
- The test results (e.g., p values) given as exact values whenever possible and with confidence intervals noted.
- A summary of the descriptive statistics, including central tendency (e.g., median, mean) and variation (e.g., standard deviation, interquartile range).
- Clearly defined error bars.

See the web collection on statistics for biologists for further resources and guidance.

Software

7. Software
Describe the software used to analyze the data in this study. All software used is publicly available.

- Genezoom, http://genome.sph.umich.edu/wiki/Genezoom
- RareMETALS, http://genome.sph.umich.edu/wiki/RareMETALS
- RVTESTS, http://genome.sph.umich.edu/wiki/RvTests
- RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER

For all studies, we encourage code deposition in a community repository (e.g., GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

8. Materials availability
Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Primary summary statistics were provided by individual study PIs. No companies were involved.

9. Antibodies
Describe the antibodies used and how they were validated for use in the system under study (i.e., assay and species).

N/A

10. Eukaryotic cell lines
a. State the source of each eukaryotic cell line used.

N/A

b. Describe the method of cell line authentication used.

N/A
c. Report whether the cell lines were tested for mycoplasma contamination.

N/A
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

N/A
Animals and human research participants

Policy information about studies involving animals: when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
Provide details on animals and/or animal-derived materials used in the study.

N/A

Policy information about studies involving human research participants

12. Description of human research participants
Describe the covariate-relevant population characteristics of the human research participants.

Summaries of research participants for each primary study are given in Supplementary Table 14. Since only summary statistics by variants were shared, and no individual-level genetic or phenotypic data were shared, this study is not considered to be Human Subjects research.