Trans-ethnic genome-wide association study identifies 12 new genetic loci influencing blood pressure traits, and implicates a role for DNA methylation: the International Genetics of Blood Pressure (iGEN-BP) Study.
Supplementary Tables

**Supplementary Table 1.** Characteristics of participants in the GWA cohorts. \( N_{\text{max(QTs)}} \) is the maximum sample size for the blood pressure quantitative traits (SBP, DBP, MAP, PP); \( N_{\text{max(HT)}} \) is maximum sample size for HT.

**Supplementary Table 2.** Genotyping, imputation and association testing in the GWA cohorts.

**Supplementary Table 3.** Genomic control inflation factors in the GWA study.

**Supplementary Table 4.** Sentinel SNPs reaching \( P<1\times10^{-7} \) after combined analysis of results from the trans-ethnic GWA with results from ICBP. Effects are given as beta co-efficients per allele copy from linear regression (mmHg, SBP, DBP, MAP, PP) or logistic regression (log-Odds, HT). \( P_{\text{hetero}} \) is the P value for heterogeneity between the three ethnic groups.

**Supplementary Table 5.** Characteristics of participants and genotyping methods in the replication cohorts. \( N_{\text{max(QTs)}} \) is the maximum sample size for the blood pressure quantitative traits (SBP, DBP, MAP, PP); \( N_{\text{max(HT)}} \) is maximum sample size for HT.

**Supplementary Table 6.** Results for the sentinel SNPs carried forward for further testing in the replication cohorts. Nineteen SNPs were carried forward from the trans-ethnic GWA, two from the secondary ethnic-specific GWA.

**Supplementary Table 7.** Association of the 12 novel sentinel SNPs with the five blood pressure phenotypes. Effects are given as beta co-efficients per allele copy from linear regression (mmHg, SBP, DBP, MAP, PP) or logistic regression (log-Odds, HT).

**Supplementary Table 8.** Results in the trans-ethnic GWA for SNPs previously reported to be associated with blood pressure phenotypes in GWA studies. \( P_{\text{het1}} \) is the P value for heterogeneity between the ICBP and iGEN-BP. \( P_{\text{het2}} \) is the P value for heterogeneity between the three ethnic groups.

**Supplementary Table 9.** 99% credible SNP sets by trans-ethnic meta-analysis using MANTRA and varLD.

**Supplementary Table 10.** Potential coding or regulatory variants at the 12 newly identified loci. \( R^2 \) is between the sentinel SNP and the potential coding / regulatory variant at the respective loci. 'Comments' column provides detailed information corresponding to the 'Annotation column': i) codon change ii) number of motif instances based on position weight matrices (PWMs) discovered from ENCODE experiments, iii) number of proteins or cell types with evidence from the Roadmap Epigenome Mapping Consortium for regulatory protein binding, chromatin structure, chromatin state of the region and putative transcription factor binding motifs, as well as iv) eQTLs from the GTex eQTL browser.
Supplementary Table 11. Expression QTLs associated with the sentinel SNPs at the 12 newly identified loci. Results are for the best available proxy of the sentinel SNP in the Zeller et al eQTL database

Supplementary Table 12. Summary of current knowledge for candidate genes at the newly identified loci.

Supplementary Table 13. Characteristics of participants in the cohorts included in the SNP-methylation association analysis.

Supplementary Table 14. Methylation QTLs associated with the sentinel SNPs. For each of the sentinel SNPs, we examined association of the SNP with methylation at CpG sites assayed by the 450K methylation array and located within 1MB, using linear regression and an additive genetic model. Methylation levels were z-transformed for all analyses to facilitate comparison of effect sizes between CpG sites. Results are shown for SNP-CpG combination reaching \( P<4\times10^{-6} \) (ie \( P<0.05 \) after Bonferroni correction) in the discovery analysis amongst 1,904 South Asians. In addition for each sentinel SNP we identify the leading CpG site (lowest \( P \) value for association with the respective SNP; these leading CpG sites were carried forward for replication testing. 'Effect (SNP-CpG)' is the effect of SNP on methylation per allele copy. Correlation co-efficients between methylation at the sentinel CpGs and other CpG sites in the locus are provided. 'Annotated Gene' and 'Relation to Gene' are as annotated by the Illumina 450K array manifest.

Supplementary Table 15. DNA methylation as a potential mediator of the relationships between sentinel SNPs and blood pressure at the loci reaching genome-wide significance in our study. Results are shown for the 28 sentinel SNPs that are associated with methylation at \( P<0.05 \) after Bonferroni correction for multiple tests. The effect of SNP on methylation at the leading CpG site is given as change in methylation per allele copy of SNP in discovery samples, replication samples and in combined analysis. 'CpG-eQTL' is for the association of methylation with expression of nearest gene. 'CpG-Primary phenotype' is for the association of methylation with the discovery phenotype for the respective locus. Predicted effect on blood pressure is based on the relationship of sentinel SNPs with methylation, and methylation with blood pressure; observed is the direct relationship between the sentinel SNPs and blood pressure (discovery phenotype). Methylation levels were z-transformed for all analyses to facilitate comparison of effect sizes between CpG sites.

Supplementary Table 16. Association of sentinel blood pressure SNPs with local DNA methylation in adults (\( N=6,757 \)) and in umbilical cord blood (\( N=237 \) samples). Direction for directional consistency between adult and cord blood samples. 'Effect' is the effect of SNP on methylation per allele copy from linear regression (additive genetic model). Methylation levels were z-transformed for all analyses to facilitate comparison of effect sizes between CpG sites.

Supplementary Table 17. Characteristics of participants in the cohorts included in the Methylation-BP association analysis.

Nature Genetics: doi:10.1038/ng.3405
Supplementary Table 18. Association of sentinel SNPs with adiposity, cardiovascular disease, kidney function and type-2 diabetes as single marker tests in published GWA studies.

Supplementary Table 19. Association of weighted blood pressure genetic risk scores for the 12 novel ('Novel') and all 35 loci identified in the present study with i. Cardiovascular mortality, ii. All-cause mortality, iii. Prevalent coronary heart disease (CHD), iv. Sokolow-Lyon electrocardiographic criteria for left ventricular hypertrophy, v. Serum levels of NT-proBNP a marker of heart function, vi. Serum levels of creatinine, vii. Prevalent Type 2 Diabetes (T2D), viii. Body Mass Index (BMI), ix. Height. Effects are given per unit increase in genetic risk score, and are calculated using individual participant data in each contributing cohort followed by meta-analysis across cohorts. For comparison the association of GRS with phenotypes is also provided using summary statistics from published GWA studies71-74. 

Nature Genetics: doi:10.1038/ng.3405
Supplementary Figure 1. Overview of study design

- **GWA in East Asians** (N=31,516; 23 cohorts)
  - Global meta-analysis
  - 630 loci
    - Lead SNP $P<1\times10^{-4}$ with any BP phenotype
  - Combined meta-analysis with ICBP (N=87,205)
  - 19 novel loci
    - Sentinel SNP $P<1\times10^{-7}$
  - Further testing
    - N=133,052 (EA, EUR & SA)

- **GWA in Europeans** (N=35,352; 13 cohorts)
  - Ethnic-specific meta-analysis
  - 1 novel locus in EA
    - rs9425586
    - $P=3.0\times10^{-6}$ (HT)
  - Further testing
    - N=19,047 (EA)

- **GWA in South Asians** (N=33,126; 13 cohorts)
  - 1 novel locus EUR
    - rs13395018
    - $P=4.4\times10^{-9}$ (PP)
  - Further testing
    - N=36,023 (EUR)

**SNPs at 12 independent novel loci achieve $P<1\times10^{-9}$**
Supplementary Figure 2. Forest plots showing the results for the 19 SNPs taken forward for further testing in the replication samples. Heterogeneity tests are shown for within and between ethnic groups.
Supplementary Figure 3. Manhattan plot showing results of the GWA study.

GWA results for genetic variants at the novel loci in the GWA are shown in green. Results for the sentinel SNP at each locus are shown: Green – GWA result; Red – GWA, ICBP & replication results combined.
Supplementary Figure 4. Regional plots for the 12 newly identified loci, showing association of SNPs with phenotype (top panel) and association of methylation at CpG sites with the sentinel SNP from the GWA study (lower panel).
**Supplementary Figure 5.** Pairwise associations with blood pressure phenotypes for the 35 sentinel SNPs reaching GWA significance in the present study (Green = novel loci, red = known loci). The linear relationship between the pairwise phenotypes is shown, along with 99.9% confidence intervals (i.e., P<0.05 after Bonferroni correction for 35 tests).
Supplementary Figure 6. Effect sizes in iGEN-BP and ICBP at SNPs associated with a blood pressure phenotype in GWA studies. Novel genetic loci are shown as blue squares. Results for known genetic loci are colour coded according to the presence of heterogeneity between iGEN-BP and ICBP: red – no heterogeneity; green – $P_{het} < 0.05$; orange (N=1) - $P_{het} < 5 \times 10^{-4}$ (correction for multiple testing).
Supplementary Figure 7. Trans-ethnic fine mapping of the 5q23.2 and 15q26.1 loci for SBP.

The top four panels present signal plots for the MANTRA association signal after trans-ethnic meta-analysis of three ancestry GWAS (1st panel from the top) and after meta-analysis of only the East Asian (EA, 2nd panel), European (EUR, 3rd panel), and South Asian (SA, 4th panel) ancestry GWAS, respectively. Each point represents a SNP passing QC in our MANTRA analysis, plotted with its BF (on a log$_{10}$ scale) as a function of genomic position (Build 37). In each panel, the sentinel SNP is represented by the purple circle. The color coding of all other SNPs indicates LD with the sentinel SNP (estimated by HapMap CEU for EUR and SA and HapMap JPT+CHB for EA): red, $r^2 \geq 0.8$; gold, $0.6 \leq r^2 < 0.8$; green, $0.4 \leq r^2 < 0.6$; cyan, $0.2 \leq r^2 < 0.4$; blue, $r^2 < 0.2$; and gray, $r^2$ unknown. Recombination rates are estimated from the International HapMap Project and gene annotations (5th panel from the top) are taken from the UCSC genome browser. In each panel, the gray-shaded regions correspond to the genomic interval covered by a 99% credible set of SNPs.

The bottom two panels present ethnic differences in the distribution of proxy ($r^2 \geq 0.8$) to sentinel SNP and LD structure between three ancestry groups. A series of proxy SNPs are colored in blue for EA, gray for EUR, and purple for SA (2nd panel from the bottom). To assess the extent of LD variation between a pair of ancestry groups, we use the standardized varLD score (circles colored in green (EUR vs. EA); red (SA vs. EA); and orange (EUR vs. SA), where the horizontal dotted lines indicate the 5% empirical threshold (score = 2) across the genome). Each plot illustrates the evidence for the 200-kb region surrounding the sentinel SNP (100kb on both flanks).
Supplementary Figure 8. Inter-ethnic differences in LD between three ancestry groups at novel blood pressure loci.

Inter-ethnic differences in the distribution of proxy \( r^2 \geq 0.8 \) to sentinel SNP and LD structure between three ancestry groups are shown at 13 loci apart from PRDM6 and FTO. A series of proxy SNPs are colored in blue for EA, gray for EUR, and purple for SA at each locus. See legends to Supplementary Figure 7.
Supplementary Figure 9. Permutation testing for enrichment of methylation markers.

The histogram shows the number of cis-methylation sites associated with a permuted sentinel SNP at P<0.05 after correction for the number of CpG sites within 1MB.
Supplementary Figure 10. Association of sentinel SNPs with DNA methylation at the 28 leading CpG sites in Europeans and South Asians. Results are shown as change in methylation (in SD units) per allele copy of respective SNP amongst Europeans (x axis) and South Asians (y axis).
Supplementary Figure 11. Enrichment of sentinel SNPs for association with local DNA methylation across a range of phenotypes.

Sentinel SNPs associated with 20 traits were retrieved from the NHGRI GWAS catalogue, selected as the 20 traits with the highest number of reported sentinel SNPs. We restricted the analysis to only SNPs with $P < 5 \times 10^{-8}$ and from studies with $n > 1,000$ samples. To account for biases due to linkage disequilibrium, sentinel SNPs were pruned based on a 1-Mb flanking window. We determined the proportion of the 1,197 sentinel SNPs that are observed (red arrow) to be associated with methylation at one or more CpG sites in cis ($\pm 1$ Mb) at $P < 0.05$ after multiple testing correction. We then determined expectations under the null hypotheses by permutation testing using a set of SNPs matched for MAF ($\pm 2\%$), distance to gene ($\pm 10$ kb), and number of CpGs in cis ($\pm 200$). Empirical $P$-value is for the comparison between observed (red arrow) and 1,000,000 permutations (histogram).
Supplementary Figure 12. CpG sites quantified at the AMH locus assessed by next generation sequencing.

Correlation between markers is shown (0: no correlation – white; 1: perfect correlation – red). Bars present mean methylation at the CpG sites evaluated; the sentinel marker is shown as a green bar. The location and structure of the AMH gene and CpG sites in the genome are shown.

+: CpG-BP P<0.05
Supplementary Figure 13. Association of sentinel SNP (rs740406) with methylation, and of methylation with blood pressure, amongst 34 CpG sites at AMH locus assayed by next generation sequencing in 168 samples.

Effect of SNP on methylation is change in methylation per copy of effect allele. Association of methylation with blood pressure is change in blood pressure (mmHg) per unit change in methylation. The 34 CpG sites are colour coded: red – associated with BP at P<0.05; green– associated with SNP at P<0.05; black – associated with both BP and SNP at P<0.05. The P value shown is for the directional consistency across the 34 loci between the association of SNP with methylation, and methylation with blood pressure (sign test).
Supplementary Figure 14. Pairwise correlation of mean methylation levels across tissues.

Results are shown for the 28 CpG sites most strongly associated with the blood pressure sentinel SNPs across 7 tissue types (blood: n=10; liver, muscle, omentum, pancreas, subcutaneous (SC) fat, spleen: n=6). The lower panel displays pairwise scatterplot (trendline in red), while the upper panel shows the Pearson correlation coefficient. The data presented here is a subset of Gene Expression Omnibus (GEO) data series GSE48472\(^9\).
Supplementary Figure 15. Methylation profiles of 28 CpG sites most strongly associated with the 28 sentinel SNPs across 7 tissue types (n=41). Samples (rows) are ordered by tissue type, and columns (CpG sites) by hierarchical clustering of methylation levels with Euclidean distance as similarity measure. Methylation values range from 0 (red) to 1 (green). The data presented in this heatmap is a subset of Gene Expression Omnibus (GEO) data series GSE48472. doi:10.1038/ng.3405
Supplementary Note

Cohort specific methods

**AASC: The Anti-aging study cohort.** The study subjects are middle-aged to elderly persons who were consecutive participants in the medical check-up program at Ehime University Hospital Anti-aging Center. This medical check-up program is provided to general residents of Ehime Prefecture, and is specifically designed to evaluate aging-related disorders, including arteriosclerosis, cardiovascular diseases, physical function, and cognitive function. Clinical data used in this study were obtained from the personal medical check-up records of the subjects. All study procedures were approved by the ethics committees of Ehime University Graduate School of Medicine, and informed consent was obtained from all participants. DNA was extracted from peripheral blood using a QIAamp DNA blood kit (Qiagen GmbH, Hilden, North Rhine-Westphalia, Germany). Genome-wide SNP genotyping was performed using a HumanOmni 2.5 BeadChip array.

**AIDHS/SDS: Asian Indian Diabetic Heart Study/Sikh Diabetes Study.** AIDHS/SDS is a case-control study designed to investigate the association between genetic and environmental factors and their risk on type 2 diabetes and cardiovascular disease in a population of Punjabi ancestry from India. A total of 1534 participants from AIDHS/SDS included in this investigation were available with genome-wide genotyping data (Illumina’s 660W-Quad BeadChip). Both men and women aged 20-90 years participated. Normoglycemic control subjects were random individuals recruited from the same Asian Indian community as the patients, and matched for ethnicity and geographic location. All blood samples were obtained at the baseline visits. All participants signed a written informed consent for the investigations. The study was reviewed and approved by the University of Oklahoma Health Sciences Center’s Institutional Review Board, as well as the Human Subject Protection Committees at the participating hospitals and institutes in India.

**BIOS Consortium: Biobank-based Integrative Omics Study Consortium.** The mission of the BIOS Consortium is to create a large-scale data infrastructure and to bring together researchers focusing on integrative omics studies in Dutch Biobanks. The advent of the GWAS led to the successful identification of thousands of variants that are robustly associated with complex disease phenotypes. For most of these variants, however, the mechanisms through which they contribute to these phenotypes remain unknown. The BIOS Consortium applies a functional genomics approach that integrates genome-wide genetic data with data on the epigenome and transcriptome to elucidate these mechanisms. Over 4000 samples from Dutch biobanks with in-depth information on disease phenotypes and GWAS data are being enriched with RNA-sequencing and genome-wide DNA methylation data. The same is true for samples with whole-genome sequencing data that are part of the Genome of The Netherlands project. This data infrastructure provides a powerful platform to evaluate key questions in integrative omics from establishing comprehensive eQTL and meQTL catalogues to linking molecular pathways across omics levels to phenotypic outcomes. The BIOS consortium is funded by the Biobanking and Biomolecular Research Infrastructure (BBMRI) of The Netherlands, a research infrastructure financed by the Netherlands Organization for Scientific Research (NWO project 184.021.007).

*Nature Genetics: doi:10.1038/ng.3405*
CAGE-Amagasaki: Amagasaki Study. The Amagasaki Study (CAGE-Amagasaki) is an ongoing population-based cohort study of 5,743 individuals (3,435 males and 2,310 females), aged >18 years and recruited for a baseline examination between September 2002 to August 2003. Part of the Amagasaki Study samples (n=978) were included in the GWAS panel and not used for follow-up study. BP was measured using the COLIN BP-203RV-II machine (Omron Healthcare Co. Ltd., Japan) in the sitting position after at least 5 minutes of rest. The mean of two or three measurements taken on different occasions was used in the analysis.\(^3\)

CAGE-Amagasaki (2): Amagasaki Study. Details of the Amagasaki study sample and phenotype measurements are described above.\(^3\) Of the 5,743 samples available for genetic study, 978 were used for GWAS and included in the IGEN-BP GWAS meta-analysis. The remaining 4,765 samples were included in the de novo genotyping follow-up study.

CAGE-Fukuoka: Kyushu University Fukuoka Cohort Study. The Kyushu University Fukuoka Cohort Study (CAGE-Fukuoka) is a community-based prospective epidemiologic cohort of 12,959 subjects, who participated in the baseline survey during the period from February 2004 to August 2007.\(^4\) From this cohort, 12,569 subjects completed the questionnaire and also provided DNA for genotyping of SNPs to investigate lifestyle factors and genetic susceptibility of the so-called lifestyle-related diseases such as cardiovascular diseases, cancer, and diabetes mellitus. BP was recorded using the HEM-707 machine (OMRON, Japan) with subject in a sitting position at least for five minutes. The mean of two readings was used.

CAGE_GWAS1: Cardio-metabolic Genome Epidemiology Network. The Cardio-metabolic Genome Epidemiology (CAGE) Network is an ongoing collaborative effort to investigate genetic and environmental factors, and their interactions affecting cardiometabolic traits/disorders among Asian populations, including the Japanese, Vietnamese and Sri Lankan.\(^5\) CAGE participants were recruited in a population-based or hospital-based setting, depending on the design of member studies. From this network sample, a total of 1,547 Japanese samples were initially used for GWAS of blood pressure (BP) and hypertension (GWAS panel ver.1). These subjects were enrolled at four separate sites in Japan including the Tokyo, Nagoya, Osaka, and Shimane districts. For the case-control study, 842 hypertensive subjects and 678 normotensive controls of Japanese ancestry were selected from the GWASed samples in the CAGE Network (403 cases and 452 controls were included in 1,547 subjects above). Additional panels of 3,294 hypertensive subjects and 6,831 normotensive controls were selected from the CAGE Network (hospital-based and population-based samples from the Amagasaki Study, Kyushu University Fukuoka Cohort Study, and KING Study) to test the strength of the hypertension association. BP was measured using a standard mercury sphygmomanometer or a digital BP monitor (see description of the individual population-based studies below) in the seated position. The average of two or three readings was used for the analysis.

CAGE-KING: Kita-Nagoya Genomic Epidemiology study. The Kita-Nagoya Genomic Epidemiology (KING) study (ClinicalTrials.gov identifier: NCT00262691) is an ongoing community-based prospective observational study of the genetic basis of...
cardiovascular disease and its risk factor. The study recruited 3,975 Japanese subjects aged 50-80 years, who underwent community-based annual health checkups between May 2005 and December 2007. Part of the KING Study samples (n=1030) were included in the GWAS panel and not used for follow-up study. BP was measured using the USM-700G machine (Elquest Co., Japan) in the seated position. The average of two readings was used for the analysis.

**CAGE-KING (2): Kita-Nagoya Genomic Epidemiology study.** Details of the KING study sample and phenotype measurements are described above. Of the 3,975 samples available for genetic study, 1,030 were used for GWAS and included in the IGEN-BP GWAS meta-analysis. The remaining 2,945 samples were included in the de novo genotyping follow-up study.

**CAGE-Vietnam: Study on risk factors for diabetes and metabolic syndrome in Vietnam.** The study on risk factors for diabetes and metabolic syndrome in Vietnam (CAGE-Vietnam) is an ongoing collaborative study, aiming to longitudinally identify the prevalence and incidence of diabetes and metabolic syndrome-related diseases as well as to investigate risk factors contributing to escalation of the diseases; lifestyle-related factors and genetic factors are examined among the Vietnamese people. The participants were recruited in the capital (Hanoi) and lowland province (Thai Binh) of Vietnam in a population-based setting. A total of 3,877 individuals (1,507 in Hanoi and 2,370 in Thai Binh) were recruited at a new survey between January 2008 and September 2009. Among them, 1,567 participants (514 in Hanoi and 1,053 in Thai Binh) were used for the follow-up study from the similar survey in 2002. BP was measured twice, more than two minutes apart, in a resting and sitting position using an automatic digital sphygmomanometer (OMRON Healthcare Inc., Bannockburn, Illinois, USA) with an appropriately-sized cuff. A third measurement was performed if the BP difference between the first two measurements was >10 mmHg. The average of two or three measurements was used for the analysis.

**Cilento.** Cilento is a population-based study of isolated populations located in the area of the National Park of Cilento e Vallo di Diano. A total of 1,062 individuals were available with blood pressure measurements. The mean (SD) age of the Cilento cohort with BP was 52.6 (19.3). The study design was approved by the ethics committee of Azienda Sanitaria Locale Napoli. The study was conducted according to the criteria set by the declaration of Helsinki and each subject signed an informed consent before participating to the study. Blood pressure was measured on a sample of 1,062 adults. The measurements were performed twice for each arm with an interval of 15 min and in a seated position. A mercury sphygmomanometer was used by trained clinicians. We considered the average of the four measures in the analysis. 1,147 samples were typed with 370K and OmniExpress assays (Illumina, San Diego, USA). A per SNP call rate >95% was applied as QC filter prior to imputation. The number of SNPs that passed quality control and used for the imputation was 190,862. Imputation of genotypes was carried out using Mach and miniMach programs. The GWAS was performed using GenAbel and ProbAbel software. The relatedness between individuals was taken into account into the GWAS by using the mmscore function of GenABEL software.

**CLHNS: Cebu Longitudinal Health and Nutrition Survey.** The Cebu Longitudinal Health and Nutrition Survey (CLHNS) is a community-based birth cohort study that
originally enrolled 3,327 pregnant women from the Metropolitan Cebu, Philippines area in 1983-4 (3,080 singleton live births), and has since followed them and their offspring. For this study of 1,787 CLHNS mothers, systolic and diastolic blood pressures were measured in triplicate after a 10-minute seated rest using mercury sphygmomanometers in the 2005 survey. The average of the three measurements was used in analysis.

**DIABNORD.** The DIABNORD Study is nested within the Västerbotten Health Survey, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. Participants with incident type 2 diabetes were identified from the Diabetes Register in Northern Sweden (DiabNorth). A total of 912 participants with incident type 2 diabetes from the DIABNORD Study had complete genotype and phenotype data necessary for the current analyses. Systolic and diastolic blood pressures were measured once, after a 5-min rest, using a mercury-gauge sphygmomanometer with the participant in a supine position. Participants were genotyped with Illumina HumanExome Beadchip 12 v1.1. This array contains ~250,000 variants, the majority being low-frequency (0.01≤minor allele frequency (MAF)≤0.05) or rare (MAF≤0.01) variants. The chip also contains tags for previously described GWAS hits.

**EGCUT: Estonian Genome Center of University of Tartu.** The Estonian cohort comes from the population-based biobank of the Estonian Genome Project of University of Tartu (EGCUT). The project is conducted according to the Estonian Gene Research Act and all participants have signed the broad informed consent (www.biobank.ee). In total, 52,000 individuals aged 18 years or older participated in this cohort (33% men, 67% women). The population distributions of the cohort reflect those of the Estonian population (83% Estonians, 14% Russians and 3% other). General practitioners (GP) and physicians in the hospitals randomly recruited the participants. A Computer-Assisted Personal interview was conducted during 1–2 h at doctors’ offices. Data on demographics, genealogy, educational and occupational history, lifestyle and anthropometric and physiological data were assessed. All diseases are recorded according to ICD-10 guidelines.

**FINCAVAS: The Finnish Cardiovascular Study.** The purpose of the Finnish Cardiovascular Study (FINCAVAS) is to construct a risk profile - using genetic, haemodynamic and electrocardiographic (ECG) markers - of individuals at high risk of cardiovascular diseases, events and deaths. All patients scheduled for an exercise stress test at Tampere University Hospital and willing to participate have been recruited between October 2001 and December 2007. The final number of participants is 4,567. In addition to repeated measurement of heart rate and blood pressure, digital high-resolution ECG at 500 Hz is recorded continuously during the entire exercise test, including the resting and recovery phases. About 20% of the patients are examined with coronary angiography. Genetic variations known or suspected to alter cardiovascular function or pathophysiology are analysed to elucidate the effects and interactions of these candidate genes, exercise and commonly used cardiovascular medications. Genotyping has been done with Illumina HumanCardio-Metabo BeadChip for 2,795 participants. After exclusions, both genotypes and blood pressure measurements were available for 2,241 participants. All participants gave their written informed consent.
GEMS: Gene Environment Multiphenotype Study. The GEMS study consists of participants from the Health Effects of Arsenic Longitudinal Study (HEALS) and Bangladesh Vitamin E and Selenium Trial (BEST) studies. Between 2000 and 2008, 20,033 participants were enrolled from rural Bangladesh in HEALS cohort to the health effects of chronic arsenic and other environmental exposures in Bangladesh population. Entry into HEALS was based on three eligibility requirements: (1) married males and females, (2) residents of the study area (Araihazar, Bangladesh) for at least 5 years (3) primarily drinking water from one of the 6000 tested wells for at least 3 years. Between 2006 and 2009, 7,000 Bangladeshi adults from similar geographic areas with benign skin lesions were enrolled using same study instruments in a 2x2 full factorial double-blind randomized controlled trial. The trial was designed to evaluate the efficacy of Vitamin E and Selenium in preventing non-melanoma skin cancer and other health outcomes. All participants had their blood pressure measured at recruitment using an automatic sphygmomanometer (HEM 712-C; Omron Healthcare GmbH, Hamburg, Germany). Measurements were taken in the participant’s homes with the cuff around the upper left arm of seated participants after 5 minutes of rest. Participants whose systolic blood pressure measured ≥ 140 mmHg, or diastolic blood pressure measured ≥ 90 mmHg were re-measured after 2-3 minutes of rest. The measurement with the lowest blood pressure was recorded. A subset of participants from each of the studies had their DNA processed, genotyped, and analyzed using the same protocols. After quality control, 1,990 BEST participants and 3,364 HEALS participants with genotyping information were used for this analysis. Each genotyped SNP was checked for association with the blood pressure phenotypes using the software GEMMA to implement a linear-mixed model to control for relatedness in this population. Age, age squared, gender, genotyping batch, and study population were used as covariates.

GeneBank: The GeneBank Study. The Cleveland Clinic GeneBank study is a single site sample repository generated from consecutive patients undergoing elective diagnostic coronary angiography or elective cardiac computed tomographic angiography with extensive clinical and laboratory characterization and longitudinal observation. Subject recruitment occurred between 2001 and 2006. Ethnicity was self-reported and information regarding demographics, medical history, and medication use was obtained by patient interviews and confirmed by chart reviews. All clinical outcome data were verified by source documentation. Coronary artery disease (CAD) was defined as adjudicated diagnoses of stable or unstable angina, myocardial infarction (MI) (adjudicated definition based on defined electrocardiographic changes or elevated cardiac enzymes), angiographic evidence of ≥ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of revascularization). Systolic and diastolic blood pressure was measured in each arm using a mercury sphygmomanometer in a standardized fashion with the subject in sitting position after 10 min of rest. The higher of the two values in each arm were used for the GWAS analyses of systolic and diastolic blood pressure. All patients provided written informed consent prior to being enrolled in GeneBank and the study was approved by the Institutional Review Board of the Cleveland Clinic. The GeneBank Study has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic disease.
**GenSalt: Genetic Epidemiology Network of Salt-Sensitivity.** The Genetic Epidemiology Network of Salt-Sensitivity (GenSalt) study is a unique NHLBI-sponsored family feeding-study designed to examine the interaction between genes and dietary sodium and potassium intake on BP. A detailed description of the GenSalt study design and participants has been reported previously. Briefly, 3,142 participants from 633 Han families from rural, north China were ascertained through a proband with untreated pre-hypertension or stage-1 hypertension identified from a population-based BP screening. A total of 1,906 GenSalt probands and their siblings, spouses, and offspring were eligible and a resulting 1,881 took part in the dietary intervention and GWAS genotyping. Three morning BP measurements were obtained according to a standard protocol during each of the 3-days of baseline observation. All BP readings were measure by trained and certified observers using a random-zero sphygmomanometer. BP was measure with the participant in the sitting position after 5 minutes of rest. In addition, participants were advised to avoid alcohol, cigarette smoking, coffee/tea, and exercise for at least 30 minutes prior to their BP measurements. The average of 9 systolic and diastolic BP measure from the 3-day baseline examinations is used in this analysis.

**GLACIER – exome: Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk – exome.** The Gene-Lifestyle interactions And Complex traits Involved in Elevated Disease Risk (GLACIER) Study is nested within the Västerbotten Health Survey, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. A total of 928 non-diabetic participants from the GLACIER Study had complete genotype and phenotype data necessary for the current analyses. Systolic and diastolic blood pressures were measured once, after a 5-min rest, using a mercury-gauge sphygmomanometer with the participant in a supine position. Participants were genotyped with Illumina HumanExome Beadchip 12 v1.1. This array contains ~250,000 variants, the majority being low-frequency (0.01≤minor allele frequency (MAF)≤0.05) or rare (MAF≤0.01) variants. The chip also contains tags for previously described GWAS hits.

**GLACIER MetaboChip: Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk.** The Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) Study is nested within the Västerbotten Health Survey, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. A total of 5917 participants from the GLACIER Study had complete genotype and phenotype data necessary for the current analyses. DNA was extracted from peripheral white blood cells and genomic DNA samples were diluted to 4 ng/μl as previously described. Samples were genotyped with the MetaboChip (Illumina iSelect) array set. MetaboChip contains ~200,000 SNPs of interest for metabolic, cardiovascular and anthropometric traits. Systolic and diastolic blood pressures were measured once, after a 5-min rest, using a mercury-gauge sphygmomanometer with the participant in a supine position.

**GOYA (Male): Genetics of extremely Overweight Young Adults.** The study individuals consist of randomly selected control group of one in every hundred men (n= 3,601) and all extremely overweight men (n=1,930) which were identified from the records of 362,200 Caucasian men examined at the mean age of 20 years at the Nature Genetics: doi:10.1038/ng.3405
draft boards in Copenhagen and its surroundings during the years 1943–77. Standing height and weight were measured at the draft board examinations. All extremely overweight men and a random sample of half the men, who were still living in the region, were invited to a follow-up survey (anthropometric measurements) in 1992–94 at the mean age of 46 years, at which time the blood samples were also taken (753 extremely overweight and 879 control men attended). Genome-wide genotyping on the Illumina 610 k quad chip was carried out at the Centre National de Génotypage (CNG), Evry, France. We carried out imputation to HapMap release 22 (CEU individuals) using Mach 1.0, Markov Chain Haplotyping. Blood pressure was measured in a sitting position on the left upper arm after 5 min rest. A London School of Hygiene sphygmomanometer was used. Run down of the mercury column was set to 2 mm/s.

Health2006. The Health2006 study is a population based study that originated in 2006 with participants of Danish origin aged between 18-69 years living in the southwestern part of the greater Copenhagen area, Denmark. A total of 3190 participants from the Health2006 study had complete genotype and phenotype data necessary for the current analysis. Genomic DNA was extracted from buffy coat using the Qiagen AutoPure LS system. Samples were genotyped with Human Cardio-Metabo BeadChip (Illumina) array set that contains ~200,000 SNPs. Two systolic and diastolic blood pressure measures were recorded with the patient in the sitting position using a standard mercury sphygmomanometer (Mercurio300) with an appropriate cuff size after at least a 5-min rest. The mean of the two blood pressure measures was calculated and used in the current analysis.18

HEXA: Health Examinee (HEXA) shared control study. The HEXA cohort is one of the KoGES population-based cohorts which were initiated in 2001 aiming to identify risk factors of life-style related complex diseases such as type 2 diabetes, hypertension, and dyslipidemia. Approximately 3,700 of 1,200,000 subjects aged 40-69 from the HEXA cohort were randomly selected as a shared control group for the Korean cancer and coronary artery disease (CAD) GWA studies. Genotyping was conducted with the Affymetrix Genome-Wide Human SNP array 6.0 in 200819.

HPS: Heart Protection Study. The Heart Protection Study (HPS) was a large randomised trial of statin therapy versus placebo in high-risk individuals. Between 1994 and 1997, 20536 men and women aged 40–80 years were recruited from 69 collaborating hospitals in the UK (with ethics committee approval). Participants were eligible for inclusion if they had non-fasting blood total cholesterol concentrations of at least 3.5 mmol/L (135 mg/dL) and either a previous diagnosis of coronary disease, ischaemic stroke, other occlusive disease of noncoronary arteries, diabetes mellitus, or (if men 65 years or older) treated hypertension. Blood pressure measurements and non-study medications included in the current analyses were recorded at or prior to the randomization visit. A random selection of 4000 self-reported Caucasians with lipids and other biomarker measurements were selected for genotyping using the Illumina 610K Quad panel, and data were available for 3,895 individuals after quality control exclusions. Further details of the Heart Protection Study are reported elsewhere20.

INGI-VB: INGI-Val Borbera. The INGI-Val Borbera population is a collection of 1,664 genotyped samples collected in the Val Borbera Valley, a geographically
isolated valley located within the Appennine Mountains in Northwest Italy. The valley is inhabited by about 3,000 descendants from the original population, living in 7 villages along the valley and in the mountains. Participants were healthy people 18-102 years of age that had at least one grandfather living in the valley.

**Inter99.** The Inter99 study is a population based randomized controlled trial (CT00289237, ClinicalTrials.gov) investigating the effects of lifestyle intervention on CVD where the baseline health examination was performed between 1999 and 2001 and followed up after five and ten years. The participants have been living in the Copenhagen area (with a Northern European origin from Denmark, Norway, Sweden, Iceland, and Faeroe Islands) and were identified in the central Danish Civil Registration System, and recruited by invitation. Baseline data from a random subsample of 6,124 participants aged approximately between 30-65 years with genotype and phenotype information participated in the current study. Samples were genotyped with HumanCardio-Metabo BeadChip (Illumina) array set that contains ~200,000 SNPs. Informed written consent was obtained from all participants. The study was approved by the Ethical Committee of Copenhagen. Two systolic and diastolic blood pressure measures were recorded with the patient in the lying position using a standard mercury sphygmomanometer with an appropriate cuff size after at least a 5-min rest. The mean of the two blood pressure measures was calculated and used in the current analysis.

**InterAct.** The InterAct study is a large-scale case-cohort study nested within European Prospective Investigation into Cancer and Nutrition (EPIC) cohorts. InterAct includes 12,403 incident cases of T2D and a comparison cohort of 15,376. Details of the study and its participant characteristics have previously been described. A maximum of 18,676 (7,655 incident cases, 11,021 non-cases) had genotypes available from either the Illumina 660W-Quad Bead Chip (N=9,290) imputed into 1000 Genomes or Illumina CardioMetabo chip (N=9,361) and were eligible for inclusion in the current study. For the blood pressure measurements in EPIC centres, uniform procedures were recommended but no standard method or common type of instrument was introduced. Analyses were performed within incident cases and the subcohort separately, and were further stratified by genotyping platform, given different SNP coverage on each platform.

**JMGP: The Japanese Millennium Genome Project.** The Japanese Millennium Genome Project comprises seven independent study cohorts for studies of cardiovascular diseases and related risk factors. Among them, four cohorts of general Japanese population, namely Nomura (Ehime University), Ohasama (Tohoku University), Shigaraki, and Takashima (Shiga University), and two cohorts derived from employees of large manufacturing industries, i.e. Matsuyama (Ehime University) and Yokohama (Yokohama city University), were included in this analysis. Study subjects were recruited through a community-based or company-based annual medical check-up process, and all clinical data used in this study were obtained from the personal medical check-up records of the subjects. All study procedures were approved by the ethics committees of each university, and signed informed consent was obtained from all participants. DNA was extracted from peripheral blood using a QIAamp DNA blood kit (Qiagen GmbH, Hilden, North Rhine-Westphalia, Germany). Genotype was analysed by a TaqMan probe assay using commercially available primers and probes purchased from the Assay-on-
Demand system (Life Technologies Corporation, Carlsbad, California, USA). Fluorescence of PCR products was measured using an ABI PRISM 7900HT sequence detector (Life Technologies, Minato-ku, Tokyo, Japan)25.

**KARE: The Korea Association Resource study.** The Korea Association Resource (KARE) study was initiated in 2007 to undertake a large-scale GWA analysis for Type 2 Diabetes and numerous complex quantitative traits amongst the 10,038 participants (aged between 40 and 69) of the Ansung (n=5,018) and Ansan (n=5,020) population-based cohorts26. The two KARE study cohorts were established as part of the Korean Genome Epidemiology Study (KoGES) in 2001. Both cohorts were sampled from KyungGi-Do province, close to Seoul, the capital of the Republic of Korea and adopted the same investigational strategy. More than 260 traits have been extensively examined among KARE participants through epidemiological surveys, physical examinations, and laboratory tests. Three BP measurements were obtained from each study participant using a random zero sphygmomanometer. The average of these three measures is used in the current analysis.

**KORA: Cooperative Health Research in the Region of Augsburg.** KORA (Cooperative Health Research in the Region of Augsburg) is a research platform of independent population-based health surveys and subsequent follow-up examinations of individuals of German nationality resident in the region of Augsburg in Southern Germany. Written informed consent was obtained from all participants and the studies have been approved by the ethics committee of the Bavarian Medical Association. Study design, sampling method and data collection have been described in detail elsewhere27. The survey S4 was conducted in 1999-2001, respectively, and comprised 4261 subjects aged 25 to 74 years. Of these, 3080 participated in the follow-up examination F4 in 2006-2008. Anthropometric variables and clinical parameters were determined at all examinations. DNA methylation measurements from a representative subsample of 1727 KORA F4 participants were considered. DNA methylation was measured with the Infinium HumanMethylation450K BeadChip® (Illumina, Inc., CA, USA). Sample preparation and measurement have been described in detail elsewhere28.

**LBC1921: Lothian Birth Cohort 1921.** LBC1921 consists of 550 (234 male) relatively healthy individuals, assessed on cognitive and medical traits at a mean age of 79.1 years (SD = 0.6). They were born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) of Scotland. A full description of participant recruitment and testing can be found elsewhere29. Ethics permission for the study was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent. Blood pressure was measured once in a sitting position. A B026 Omron 705IT monitor was used to measure blood pressure. Whole blood extracted DNA samples were genotyped at the Wellcome Trust Clinical Research Facility using the Illumina 610-Quadv1 array (San Diego). Individuals were excluded based on unresolved gender discrepancy, relatedness, call rate (≤ 0.95), and evidence of non-Caucasian descent. Standard quality control measures were applied including the following thresholds: call rate ≥ 0.98, minor allele frequency ≥ 0.01, and Hardy-Weinberg Equilibrium test with P ≥ 0.001. The first four components
from a multidimensional scaling (MDS) analysis of the SNP data were extracted and used as covariates in the analyses to control for population stratification. 2,543,887 SNPs common SNPs were imputed using the HapMap phase II CEU data as the reference sample. NCBI build 36 (UCSC hg18) was used and genotype data were imputed using MACH software. Linear regression analysis assuming an additive genetic model was performed using MACH2QTL.

**LBC1936: Lothian Birth Cohort 1936.** LBC1936 consists of 1091 (548 male) relatively healthy individuals who underwent cognitive and medical testing at a mean age of 69.6 years (SD = 0.8). They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland. A full description of participant recruitment and testing can be found elsewhere. Blood pressure was measured three times in a sitting position and the mean measurement used. A B026 Omron 705IT monitor was used to measure blood pressure. Whole blood extracted DNA samples were genotyped at the Wellcome Trust Clinical Research Facility using the Illumina 610-Quadv1 array (San Diego). Individuals were excluded based on unresolved gender discrepancy, relatedness, call rate (≤ 0.95), and evidence of non-Caucasian descent. Standard quality control measures were applied including the following thresholds: call rate ≥ 0.98, minor allele frequency ≥ 0.01, and Hardy-Weinberg Equilibrium test with P ≥ 0.001. The first four components from a multidimensional scaling (MDS) analysis of the SNP data were extracted and used as covariates in the analyses to control for population stratification. 2,543,887 SNPs common SNPs were imputed using the HapMap phase II CEU data as the reference sample. NCBI build 36 (UCSC hg18) was used and genotype data were imputed using MACH software. Linear regression analysis assuming an additive genetic model was performed using MACH2QTL.

**LifeLines: LifeLines Cohort Study.** LifeLines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behaviour, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity and complex genetics. In addition, the LifeLines project comprises a number of cross-sectional sub-studies which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. Written informed consent was obtained from every participant.

Systolic and diastolic blood pressure, and pulse rate are measured every minute for a period of 10 minutes using an automated DINAMAP Monitor i.e. 10 measures for each of the indices. The size of the cuff was chosen according to the arm circumference. The reported level for each of the blood pressure indices is an average estimate of the last 3 measures. Genome-wide genotyping was performed in all participants using the Illumina HumanCytoSNP12 v2 beadchip assay (Illumina, Inc; San Diego, CA, USA). Genotypes were called with the Illumina GenomeStudio software package (Illumina, Inc). SNPs were excluded if the genotype call rate was <95%, the minor allele frequency was <1% or the Hardy-Weinberg P-value was <10-4. Subjects were excluded if the sample call rate was
<95%, if the subject was not Caucasian (assessed with EigenStrat, or if the subject was highly related to another individual and had a lower sample call rate. Untyped SNPs were imputed with IMPUTE using the Hap-Map CEU Phase II reference set (release 22, build 36).

**LOLIPOP: London Life Sciences Population study.** LOLIPOP is a population based cohort study of ~30,000 South Asian and European white men and women, aged 35-75 years, recruited from the lists of 58 General Practitioners in West London, UK\(^{32}\). A nurse-led interviewer-administered questionnaire was used to collect data on medical history, family history, current prescribed medications and cardiovascular risk factors. Physical assessment included measurements of height, weight, waist and hip circumference. Blood pressure was measured in the sitting position using an Omron 705CP, and the mean of three measurements recorded.

**LURIC: Ludwigshafen Risk and Cardiovascular Health Study.** LURIC is an ongoing prospective study of more than 3,300 individuals of German ancestry in whom cardiovascular and metabolic phenotypes (CAD, MI, dyslipidemia, hypertension, metabolic syndrome and diabetes mellitus) have been defined or ruled out using standardized methodologies in all study participants. Inclusion criteria for LURIC were: German ancestry (limitation of genetic heterogeneity), clinical stability (except for acute coronary syndromes) and availability of a coronary angiogram. Exclusion criteria were: any acute illness other than acute coronary syndromes, any chronic disease where non-cardiac disease predominated and a history of malignancy within the last five years. A 10-year clinical follow-up for total and cause specific mortality has been completed\(^{33}\).

**NFBC86: Northern Finland Birth Cohort 1986.** NFBC1986 is a longitudinal one-year birth cohort study from an unselected population\(^{34}\). The cohort included all the mothers (N = 9,362) with children whose expected date of birth fell between July 1st 1985 - June 30th 1986 in the two northernmost provinces on Finland (Oulu and Lapland). Altogether 9,432 children were live-born into the cohort. The longitudinal data collection includes clinical examination and blood sampling at age 15/16 years, from which data in the current study are drawn. Blood pressure was measured using a mercury sphygmomanometer, seated, from the right arm after 15 minutes rest. The average of two readings taken 5 minutes apart was used for the analyses. Both questionnaire and national medication reimbursement data were used for anti-hypertensive medication information. A random sample of 1,949 individuals were genotyped using illumina HumanOmniExpressExome chip and had both genotypes and blood pressure phenotypes available.

**NHAPC: The Nutrition and Health of Aging Population in China.** The Nutrition and Health of Aging Population in China (NHAPC) is a population-based cohort study among 3,289 individuals (3,210 of them are Chinese Hans), aged 50 to 70 years, recruited from Beijing and Shanghai. The study design, methods and measurements of this cohort study have been described in detail elsewhere\(^{35}\). Briefly, the participants were recruited using a multistage sampling method from 2 urban districts and 1 rural district of each city. Data on demographic variables, health status, health behavior, and physical activity was collected using a standardized questionnaire, and standard anthropometric measurements were collected using a standardized protocol when the participants attended a physical examination. Blood
pressure was measured (Omron HEM-705CP, OMRON Healthcare, Vernon Hills, IL) on the right arm of the participant in a seating position after 5 min of rest. Written informed consent was obtained from all participants, and study protocols were approved by the institutional review board of the Institute for Nutritional Sciences.

**POPGEN:** **Popgen control cohort.** Popgen is a population-based sample that has been recruited by the popgen Biobank (www.popgen.de) in Kiel, Germany, between 2005 and February 2006 as reference sample for genetic-epidemiological studies. Participants were either recruited through official population registries in Kiel (Germany) or as blood donors. At the first follow-up examination (2010-2012), participants were invited in our study center and characterized with respect to basic cardiovascular risk factor. BP was measured twice on the right arm using a cuff of appropriate size. The mean of both measurements was considered in the analyses. Weight was determined using a calibrated scale (SECA) by subtracting 2 kg from the value obtained in full clothing (without shoes).

**PREVEND:** **Prevention of REnal and Vascular ENd stage Disease.** The Prevention of REnal and Vascular ENd stage Disease (PREVEND) study is an ongoing prospective study investigating the natural course of increased levels of urinary albumin excretion and its relation to renal and cardiovascular disease. Inhabitants 28 to 75 years of age (n=85,421) in the city of Groningen, The Netherlands, were asked to complete a short questionnaire, 47% responded, and individuals were then selected with a urinary albumin concentration of at least 10 mg/L (n = 7,768) and a randomly selected control group with a urinary albumin concentration less than 10 mg/L (n = 3,395). Details of the protocol have been described elsewhere (www.prevend.org). Blood pressure was measured in the supine position every minute for 10 and 8 minutes, respectively, with an automatic DINAMAP XL Model 9300 series monitor (Critikon, Tampa, Florida). Systolic and diastolic blood pressures were calculated as the mean of the last two measurements at the two visits.

**PROMIS:** **Pakistan Risk Of Myocardial Infarction Study.** The Pakistan Risk of Myocardial Infarction study (PROMIS) is a case-control study of acute myocardial infarction (MI) in participants of South Asian origin. All MI cases in PROMIS were enrolled within 24 hours of onset of chest symptoms and had typical ECG changes and a positive troponin-I test. Control subjects were individuals without a history of cardiovascular disease. They were frequency-matched to cases by sex and age (in 5-year age intervals). Controls were concurrently identified in the same hospitals as index cases and were either (1) visitors of patients attending the outpatient department, (2) patients attending the outpatient department for routine non-cardiac complaints, or (3) non–blood-related visitors of index MI cases. People with recent illnesses or infections were not eligible. Information was recorded on personal and parental ethnicity, spoken language, dietary intake, lifestyle factors, and other characteristics. PROMIS has received approval by the research ethics committee at the Center for Non-Communicable Diseases, Pakistan. Informed consent has been obtained from each participant recruited into the study, including for use of samples in genetic, biochemical and other analyses.

**PROSPER:** **PROspective Study of Pravastatin in the Elderly at Risk.** All data come from the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER).
A detailed description of the study has been published elsewhere\textsuperscript{38,39}. PROSPER was a prospective multicenter randomized placebo-controlled trial to assess whether treatment with pravastatin diminishes the risk of major vascular events in elderly. Between December 1997 and May 1999, we screened and enrolled subjects in Scotland (Glasgow), Ireland (Cork), and the Netherlands (Leiden). Men and women aged 70-82 years were recruited if they had pre-existing vascular disease or increased risk of such disease because of smoking, hypertension, or diabetes. A total number of 5,804 subjects were randomly assigned to pravastatin or placebo. A large number of prospective tests were performed including Biobank tests and cognitive function measurements. A whole genome wide screening has been performed in the sequential PHASE project with the use of the Illumina 660K beadchip\textsuperscript{40}. Of 5,763 subjects DNA was available for genotyping. Genotyping was performed with the Illumina 660K beadchip, after QC (call rate $<$95%) 5,244 subjects and 557,192 SNPs were left for analysis. These SNPs were imputed to 2.5 million SNPs based on the HAPMAP built 36 with MACH imputation software.

**RHS:** Ragama Health Study. The Ragama Health Study (RHS) is a population-based study of South Asian men and women aged 35-64 yrs living in the Ragama Medical Officer of Health (MOH) area, near Colombo, Sri Lanka.\textsuperscript{41} Consenting adults attended a clinic after a 12-h fast with available health records, and were interviewed by trained personnel to obtain information on medical, sociodemographic, and lifestyle variables. A 10-mL sample of venous blood was obtained from each subject. The concurrent study was performed in two tea plantation estates in the Lindula MOH area, near Nuwara Eliya (180 km from Colombo), to investigate the gene-environment interaction in a community with differing lifestyles (e.g., physical activity and diet). BP was measured using the Omron 750CP (Omron Co., Japan) in the seated position. The average of two readings was used for the analysis. The RHS is a collaborative effort between the Faculty of Medicine, University of Kelaniya and the National Center for Global Health and Medicine, Japan.

**RS:** The Rotterdam Study. RS is a large prospective population-based cohort study including men and women of 45 years and over. The design and rationale of this study are described in detail elsewhere. In summary, the Rotterdam Study aims to investigate the determinants, incidence and progression of chronic disabling diseases in the elderly. The first cohort, Rotterdam Study I (RS-I) was initiated in 1989 including 7983 persons aged 55 years and older and was extended in 1999 with 3011 participants in Rotterdam Study II (RS-II). In 2005, Rotterdam Study III (RS-III) added another 3932 individuals aged 45 years and older. All participants were examined in detail at baseline and follow-up visits every ±5 years. In summary, a home interview was conducted and the subjects underwent an extensive set of examinations at the research center, including cardiovascular and metabolic health. DNA was isolated from whole blood using standard procedures. Epigenome-wide methylation scans were carried out on the DNA collected at the baseline visit of the RS-III cohort (the RSIII-dataset), or at the fifth follow-up measurement in all three Rotterdam cohorts (RS-BIOS dataset).

**SCES:** Singapore Chinese Eye Study. The Singapore Chinese Eye Study (SCES) is the Singapore Chinese equivalent to SiMES, where the sampling was similarly performed in the same 15 residential districts and included 3,353 Singapore Chinese individuals.\textsuperscript{42} Two readings of blood pressure were taken from participants after 5
min resting using an automated blood pressure monitor (Dinamap model Pro Series DP110X-RW, 100V2, GE Medical Systems Information Technologies Inc, Milwaukee, WI) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant’s arm. A third reading was performed if the difference between the two readings of either the systolic blood pressure (SBP) was greater than 10mmHg or the diastolic blood pressure (DBP) was greater than 5mmHg. The mean values of the closest two readings were calculated.

SCHS: Singapore Chinese Health Study. The design of the Singapore Chinese Health Study has been previously described. Briefly the cohort is drawn from permanent residents or citizens of Singapore aged 45-75 at study entry, who reside in government-built housing estates (~86% of Singapore residents live in such facilities). The study subjects are restricted to the two major dialect groups of Chinese in Singapore: The Hokkiens, who originated from southern Fujian Province, and the Cantonese, who came from Guangdong Province (Both provinces are in south eastern China. Between April 1993 and December 1998, 63,257 individuals completed an in-person interview that included questions on usual diet, demographics, height and weight, use of tobacco, usual physical activity, menstrual and reproductive history (women only), medical history, and family history of cancer. A follow-up telephone interview took place between 1999 and 2004 for 52,325 cohort members (83% of recruited cohort). By April 2005, all surviving cohort subjects had been contacted for biospecimen donation. Samples were obtained from 32,535 subjects, representing a consent rate of about 60%. The institutional review boards at the National University of Singapore, the University of Minnesota, and the University of Pittsburgh approved this study.

At the time of the blood draw, blood pressure measurements were taken with the Omron HEM-705CP digital blood pressure (BP) monitor validated by the British Hypertension Society (1996). The subjects were seated with both feet parallel and flat on the floor. Subjects had not smoked or taken any vigorous exercise in the two hours before the measurement. A standard cuff was applied to the left arm. After five minutes of rest, three measures of systolic and diastolic blood pressure were taken with three minute intervals between each measurement.

For a nested case-control type 2 diabetes GWAS, we excluding subjects with prevalent diabetes, cardiovascular disease, or cancer at the baseline interview, and excluding those without stored tissue samples, 24,932 subjects remained. Among them, we identified 2615 incident diabetes cases and 2615 controls matched on age, gender, dialect group (Cantonese or Hokkien), and date of blood draw. DNA extraction was conducted at the Molecular Epidemiology and Biomarker Research Laboratory at the University of Minnesota (approximately 2/3rds samples) or the Genome Institute of Singapore (approximately 1/3rds samples) using the Qiagen method. DNA concentrations were measured by the PicoGreen method and prepared for genotype analysis. Stage 1 genotyping was performed at the Genome Institute of Singapore according to the manufacturer’s recommendations using an Affymetrix ASI (Asian) Axiom array. Genotype calling was performed by Affymetrix Corporation. A standard series of QC steps were followed in order to identify SNPs and cases and control samples for genetic association analyses. Starting with 510,584 SNPs provided for 4,918 callable study samples, we dropped additional samples with less than 98 percent of SNPs called (n=22) and SNPs (n=3,075) with less than 98 percent of the remaining samples called leaving 507,509 SNPs total. We estimated relatedness between pairs of samples as the expected number of
alleles shared identically by descent, $r_{ij}$, using PLINK. We dropped two pairs of unintended duplicate samples that were discovered to have $r_{ij}$ close to one; we also dropped samples that appeared to be closely related $r_{ij}>.2$ to more than one other sample in the study and one of each remaining pair of samples with $r_{ij}>.2$ (n=180 total including the duplicates). We compared reported sex of each sample to sex as inferred on the basis of X chromosome heterozygosity, dropping 29 uncertain or conflicting samples. We computed principal components of the genotype matrix and dropped 9 individuals who were more than 5 standard deviations from the mean on any of the first 4 principal components. One additional sample was dropped because of missing covariate information. A total of 4,677 samples (2338 cases and 2339 controls) remained after QC.

**SiMES: Singapore Malay Eye Study.** The Singapore Malay Eye Study (SiMES) is a population-based and cross-sectional study which aimed to investigate the epidemiology of eye diseases in Singapore Malays aged 40-80. Initially, resident adults were selected through an age-stratified random sampling from the 15 residential districts in South-western Singapore to obtain approximately equal numbers in each decade between the ages 40-80. 3280 Malay adults were eligible and participated in the study. Two readings of blood pressure were taken from participants after 5 min resting using an automated blood pressure monitor (Dinamap model Pro Series DP110X-RW, 100V2, GE Medical Systems Information Technologies Inc, Milwaukee, WI) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant’s arm. A third reading was performed if the difference between the two readings of either the systolic blood pressure (SBP) was greater than 10mmHg or the diastolic blood pressure (DBP) was greater than 5mmHg. The mean values of the closest two readings were calculated.

**SINDI: Singapore Indian Eye Study.** The Singapore Indian Eye Study (SINDI) is the South Asian Indian equivalent to SiMES, where the sampling was similarly performed in the same 15 residential districts and included 3,400 Asian Indians living in Singapore. Two readings of blood pressure were taken from participants after 5 min resting using an automated blood pressure monitor (Dinamap model Pro Series DP110X-RW, 100V2, GE Medical Systems Information Technologies Inc, Milwaukee, WI) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant’s arm. A third reading was performed if the difference between the two readings of either the systolic blood pressure (SBP) was greater than 10mmHg or the diastolic blood pressure (DBP) was greater than 5mmHg. The mean values of the closest two readings were calculated.

**SMART: Secondary Manifestations of ARTerial disease.** The Secondary Manifestations of ARTerial disease (SMART) study is a prospective outpatient cohort study among patients aged 18-74 years newly referred to the University Medical Center Utrecht, The Netherlands, because of atherosclerotic vascular disease or for treatment of atherosclerotic risk factors. The objective of SMART is to determine the prevalence of concomitant asymptomatic arterial disease and risk factors in patients presenting with a manifestation of arterial disease or risk factor, and to study the incidence of future cardiovascular events and their predictors in these high-risk patients. Details of the protocol have been described elsewhere. DNA for wet-lab genotyping to replicate discovery results of the current study was available in a total
of 8,361 SMART participants. Wet-lab genotyping for single nucleotide polymorphism (SNP) analysis was carried out by KBiosciences, Hertfordshire, UK. (www.kbioscience.co.uk), whose personnel were blinded to patient status, using their proprietary KASPar PCR technique and Taqman Genotype calling was carried out using an automated system, the results of which were checked manually by study personnel using SNPviewer software.

**SMHS: Shanghai Men’s Health Study.** The SMHS is an on-going population-based prospective cohort study conducted in urban Shanghai, China\(^{46}\). All male permanent residents from eight typical communities who were 40-74 years of age and had no prior history of cancer were eligible. Trained interviewers visited the homes of 83,058 eligible men identified through the Shanghai Resident Registry who lived in the study communities during the baseline study period and recruited 61,480 (response rate of 75\%) men between 2002 and 2006. The baseline survey was completed by in-person interview using a structured questionnaire designed to collect information on demographic characteristics, lifestyle habits, including dietary intake, cigarette smoking and alcohol consumption, medical history, and use of medications, including antihypertensives and hormones. The prevalence of hypertension was assessed by the question, ‘Have you ever been diagnosed with hypertension by a physician?’ Two blood pressure measurements were taken for 98.2\% of participants (n=60,401) after the participants sat quietly for more than five minutes using an aneroid sphygmomanometer according to a standard protocol. An average of two blood measurements was used for this analysis.

**SMSS: Suzhou Metabolic Syndrome Study.** Conducted in 2008, Suzhou Metabolic Syndrome Study (SMSS) was an observational cohort study on relationship between metabolic syndrome and components and cardiovascular disease events in 6 rural townships in of Suzhou in China’s Jiangsu province. A total of 18,461 individuals who were ≥20 years of age, of Han ethnicity, had no evidence of end organ damage, including coronary heart disease, stroke, chronic renal disease and tumors, and signed informed consent were recruited to participate in the study. Baseline data on demographic information, lifestyle risk factors, family history of cardiovascular disease, and personal medical history were obtained using a standard questionnaire administered by trained staff. Body weight and height were measured by using a regularly calibrated stadiometer and balance-beam scale with participants wearing light clothing and no shoes. Body mass index (BMI) was calculated as weight in kilograms divided by the square of the height in meters. Three sitting consecutive BP measurements (30 seconds between each) were taken by trained staff using a standard mercury sphygmomanometer according to a standard protocol, after the subjects had been resting for 30 minutes. The first and fifth Korotkoff sounds were recorded as systolic blood pressure (SBP) and diastolic blood pressure (DBP), respectively. The mean of the three readings was used in analysis. Hypertension was defined as SBP ≥140 mmHg and/or DBP ≥90 mmHg and/or use of antihypertensive medication in recent two weeks. Overnight fasting blood samples were obtained, plasma (serum) and white blood cell samples were isolated, and DNA was extracted from white blood cells for all the participants, plasma(serum),white blood cell and DNA samples were frozen at -80°C until laboratory testing. A modified hexokinase enzymatic method was applied to test plasma glucose levels. Total cholesterol (TC), high-density lipoprotein cholesterol (HDL cholesterol) and triglycerides (TG) were analyzed enzymatically using a
Beckman Synchron CX5 Delta Clinical System (Beckman Coulter, Inc., Fullerton, California, USA) with commercial reagents. Low density lipoprotein cholesterol (LDL cholesterol) levels were calculated using the Friedewald equation for participants who had less than 400 mg/dL TG. 4,994 ones who were randomly selected from 18,461 participants were genotyped by using TaqMan methods.

SP2: Singapore Prospective Study Program. Subjects were recruited from 4 population-based, cross-sectional surveys conducted in Singapore, including the Thyroid and Heart Study 1982–1984, the National Health Survey 1992, the National University of Singapore Heart Study 1993–1995, and the National Health Survey 1998. Briefly, all studies included a random sample of individuals from the Singapore population, with disproportionate sampling stratified by ethnicity to increase the number in the minority ethnic groups (Malays and Asian Indians). Between 2003 and 2007, we revisited these subjects. Of the 10,747 subjects who were eligible, we successfully recruited 7,742 subjects who completed a questionnaire. 5,157 subjects in this cohort also completed the subsequent clinical examination which included venesection for the collection of, among other things, DNA for genetic analysis. Two readings of blood pressure were taken from participants after 5 min resting using an automated blood pressure monitor (Dinamap Pro100V2; Criticon, Norderstedt, Germany) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant’s arm. A third reading was performed if the difference between two readings of either the systolic blood pressure (SBP) was greater than 10mmHg or the diastolic blood pressure (DBP) was greater than 5mmHg. The mean values of the closest two readings were calculated.

SRS: Shanghai-Ruijin Study. All participants are of Han Chinese ancestry recruited from Shanghai metropolitan area. After removing individuals with genotype call rates less than 98%, and keeping only one of the duplicates or first degree relatives with more complete call rates; the total sample size entered our analysis is 902. The group of hypertensives (n=447, 224 males and 223 females) were selected with following criteria: (1) Have positive family history, having at least two hypertensive family members; (2) SBP>140 mmHg and/or DBP>90mmHg if untreated during the past two weeks before taking blood pressure measurement; or take anti-hypertensive drugs; (All blood pressure measurements are the average of three independent measures in one day.) (3) Age of onset >30 years. (on average >35 years), to minimize the possibility of early-onset secondary hypertension; (4) Fasting glucose <6.5 mmol/L (to exclude diabetic cases) and no family members are diagnosed to have diabetes; (5) BMI <25 kg/m2, to exclude the confounding of obesity. (6) Secondary hypertension, diabetes, heart failure, renal or hepatic dysfunction, cancer excluded. Normotensive control subjects (n=455, 230 males and 225 females) are selected to have: (1) SBP<=120mmHg and DBP<=80mmHg; (2) Age between 40 to 80 years.; (3) BMI < 25 kg/m2; (4) Fasting glucose <6.5 mmol/L; (5) Diabetes, heart disease, liver and kidney disease, cancer excluded.

SWHS: Shanghai Women's Health Study. The SWHS is a population-based prospective cohort study of 74,941 women (response rate of 92.7%), aged 40-70 years at recruitment (March, 1997-May, 2000). Eligible women living in seven urban districts of Shanghai were recruited through in-home visits by medical professionals. At baseline, information about demographics, reproductive history,
medical history, dietary habits, weight history, physical activity, and occupational history was collected, and weight and height measurements were taken. Specific questions such as: ‘Have you ever been diagnosed with HT by a doctor’ and ‘Have you ever taken antihypertensive medications’ were asked at recruitment. During the first follow-up (2000-2002; response rate of 91%), two blood pressure measurements were taken for study participants after they sat quietly for more than five minutes using an aneroid sphygmomanometer according to a standard protocol An average of two BP measurements was used for this analysis.

**TWSC: Taiwan Super Control Study.** Control participants of the Taiwan Super Control Study (Taiwan) who took part in the current analysis included 1,000 randomly selected individuals whose genetic data were extracted from the Han-Chinese Cell and Genome Bank in Taiwan. In brief, more than 3300 healthy controls were recruited via a stratified, 3-staged probability clustering sampling scheme through the registry of all the 329 non-aboriginal townships or city districts in Taiwan and their genomic DNA was extracted from peripheral blood using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN, USA). Standard protocols for BP measurements established by the Nutrition and Health Survey in Taiwan were followed. BP was measured three times with two consecutive pulse measurements in between using the Omega 1400 NBP (Invivo Research Laboratories Inc., Orlando, FL, USA).

**WHII: Whitehall II.** Whitehall II recruitment of 10,308 participants (70% men) between 1985 and 1988 involved 20 London based Civil service departments. Genetic samples were collected in 2004 from over 6,000 participants. Blood pressure data are used from data collected in 2004 by a nurse, using an Omron HEM-907 blood pressure monitor. Three measurements were taken with the participants in a seated position following a 10 minute rest period. Genotyping includes Cardiochip and Metabochip information augmented by imputation using the 1000 Genomes dataset.

**YFS: The Young Finns Study.** The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 27-year follow-up study was conducted in 2007 (ages 30-45 years) with 2,204 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. Genotyping has been done with Illumina Human 670k Custom BeadChip for 2,443 participants. After exclusions, both genotypes and blood pressure measurements were available for 1,987 participants. All participants gave their written informed consent.
Consortium members

BIOS-consortium

Bastiaan T. Heijmans, Peter A.C. ‘t Hoen, Joyce van Meurs, Dorret I. Boomsma, René Pool, Jenny van Dongen, Jouke J. Hottenga, Marleen MJ van Greevenbroek, Coen D.A. Stehouwer, Carla J.H. van der Kallen, Casper G. Schalkwijk, Cisca Wijmenga, Sasha Zhernakova, Ettje F. Tigchelaar, P. Eline Slagboom, Marian Beekman, Joris Deelen, Diana van Heemst, Jan H. Veldink, Leonard H. van den Berg, Cornelia M. van Duijn, Bert A. Hofman, André G. Uitterlinden, P. Mila Jhamai, Michael Verbiest, H. Eka D. Suchiman, Marijn Verkerk, Ruud van der Breggen, Jeroen van Rooij, Nico Lakenberg, Hailiang Mei, Jan Bot, Dasha V. Zhernakova, Peter van ‘t Hof, Patrick Deelen, Irene Nooren, Matthijs Moed, Martijn Vermaat, René Luijk, Marc Jan Bonder, Maarten van Iterson, Freerk van Dijk, Michiel van Galen, Wibowo Arindarto, Szymon M. Kielbasa, Morris A. Swertz, Erik. W van Zwet, Aaron Isaacs, Rick Jansen, Lude Franke.

CARDioGRAMplusC4D


**LifeLines Cohort Study**
Behrooz Z Alizadeh, Paul IW de Bakker, H Marike Boezen, Pim van der Harst, Gerjan Navis, Marianne Rots, Salome Scholtens, Harold Snieder, Ronald P Stolk, Morris Swertz, Bruce HR Wolffenbuttel, Cisca Wijmenga.

**The InterAct Consortium**
Supplementary references


