""

Table of contents

"'

1. Studies and consortia for blood pressure analysis

Most of the studies contributing to the ICBP-GWAS were general population samples, and were recruited for classical or genetic epidemiological purposes, as described below. In addition to these population-based samples, we also included data from birth cohorts including a sampling of individuals born in a given year, single-sex samples, probands from a sample of twins, individuals in particular professions (civil service workers in the UK, female health professionals in the US), and controls from case-control studies of diabetes and myocardial infarction. In selecting studies for inclusion in ICBP-GWAS we sought to achieve a compromise that would maximize sample size, while minimizing potential heterogeneity of genetic effect sizes. In particular, cohorts of diabetic cases, hypertensive cases, MI cases, and cohorts with prevalent psychiatric disorders were not included for the GWAS discovery analysis ("Phase 0"). For replication analyses, most studies were population samples.

All participants provided written informed consent and studies were approved by their local Research Ethics Committees and/or Institutional Review Boards. Blood pressure, height, and weight were directly measured in all participants, except for the Women's Genome Health Study as described in Section 1.79 below. Sex and age were also recorded for all individuals. Summary demographic characteristics for studies in Europeans are listed in Supplementary Table 1, and for Non-Europeans in Supplementary Table 11. Additional descriptions of ascertainment methods for each participating study are given in Sections 1.1-1.82 below.

All studies with GWAS data performed genotyping using commercially available arrays with >300,000 SNPs. As described in Supplementary Table 2, in each study quality-control procedures excluded individual problematic samples and SNPs, using criteria such as excessive rates of genotyping error, a large proportion of missing genotypes, or marked deviations from Hardy-Weinberg equilibrium. All studies with GWAS data used hidden Markov model approaches[1-3] and HapMap reference panels[4] to impute genotypes at unmeasured SNPs and excluded SNPs, so that a common set of ~2.5M HapMap SNPs were available across the discovery samples[5,6].

Studies that performed genotyping of specific SNPs identified from our GWAS meta-analyses used a variety of assays. Five studies (YFS, EPIC-TURIN, FLEMENGHO, Nigerians, and JMGP) were genotyped using Taqman assays (Applied Biosystems, USA). Twelve studies (ARYA, ELSA, PREVEND, Prospect-EPIC, WHII, BRHS, BWHHS, GRAPHIC, INTERGENE, MRC NSHD, YMCA, and COBRA) were genotyped at KBiosciences using the KASPAr assay. Four studies (HYPEST, BRIGHT, EAS and NPHS-II) were genotyped using the KASPAr assay at Barts and The London Genome Centre. Fourteen cohorts (DiaGen, HUNT2, FINRISK97, FUSION2, METSIM, CLUE, MDC, MPP, FBPP-HyperGen, FBPP-GenNet, Jamaicans-GXE, Jamaicans-SPT, CCMB cohorts and IRAS) were genotyped using the iPLEX Sequenom MassARRAY platform. For quality control, data was used only for SNPs in Hardy-Weinberg equilibrium (p>0.001) and with call rate >90%; this excluded various SNPs for one or more cohorts, and one SNP (rs7401919) for all cohorts.

1.1 AGES Reykjavik

The Age Gene/Environment Susceptibility-Reykjavik Study originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages; another was designated as a control group and was not included in examinations until 1991. Other groups were invited to

participate in specific stages of the study. Between 2002 and 2006, the AGES-Reykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study[7]. The midlife data blood pressure measurement was taken from stage 3 of the Reykjavik Study (1974-1979), if available. Half of the cohort attended during this period. Otherwise an observation was selected closest in time to the stage 3 visit. The supine blood pressure was measured twice by a nurse using a mercury sphygmomanometer after 5 minutes rest following World Health Organization recommendations[8]. Individuals with previous MI were excluded from the analyses (N=12).

1.2 Amish

The Old Order Amish individuals included in this study were participants of several ongoing studies of cardiovascular health carried out at the University of Maryland[9,10]. Participants were mostly healthy volunteers and their family members from the Old Order Amish community of Lancaster County, PA. All protocols were approved by the Institutional Review Board at the University of Maryland and informed consent was obtained, including permission to use their DNA for genetic studies. Before any intervention, baseline BP was measured using an automated Datascope Accutorr Plus machine with the subject in the sitting position after he or she had been sitting quietly for 5 minutes, and the average of the last 2 measures was used for the analyses. Hypertension medication was discontinued in most subjects before the start of the study.

1.3 ARIC

The <u>A</u>therosclerosis <u>R</u>isk <u>I</u>n <u>C</u>ommunities Study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities[11]. Cohort members completed four clinic examinations each spread over about three years, conducted approximately three years apart between 1987 and 1998. The data used in this study are from the first visit in 1987-1989. A detailed study protocol is available on the ARIC study website (http://www.cscc.unc.edu/aric). Blood pressure was measured using a standardized Hawksley random-zero mercury column sphygmomanometer with participants in a sitting position after a resting period of 5 minutes. The size of the cuff was chosen according to the arm circumference. Three sequential recordings for systolic and diastolic blood pressure were obtained; the mean of the last two measurements was used in this analysis, discarding the first reading. Blood pressure lowering medication use was recorded from the medication history. Outliers (>4SD from the mean) with respect to the systolic or diastolic blood pressure distribution were excluded from the analysis. For this study the sample was restricted to individuals of European descent by self-report and principal component analysis using genome-wide genotypes.

1.4 ARYA

ARYA is a retrospective birth cohort consisting of 750 unselected young adults born between 1970 and 1973 in or near the city of Utrecht, the Netherlands[12]. Inclusion criteria were having birth anthropometric data and having at least one blood pressure measured in adolescence, both available from medical records of the Municipal Health Service. The only exclusion criterion was pregnancy at the time of examinations. Between 1999 and 2001, when participants had reached young adulthood, they underwent full cardiovascular disease risk profiling. Blood pressure was measured twice at two separate visits with a mean interval of 20.4 (SD 10.7) days. After 5 minutes rest in the seated position, blood pressure was measured using a semi-automated device (Dinamap) without replacing the cuff between the two measurements at the upper arm. After a further 5-15 minutes rest, the measurement was repeated. In the analyses,

the mean systolic and mean diastolic blood pressure was calculated as the average of four measurements.

1.5 B58C-T1DGC

The British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium is a sample from the national population-based sample followed periodically from birth to age 44-45 years (http://www.b58cgene.sgul.ac.uk/collection.php) and 2,580 individuals were included in this analysis. Blood pressure was recorded using the Omron 705CP machine three times, seated. The average of three readings was used for the analysis.

1.6 B58C-WTCCC

The British 1958 Birth Cohort – Wellcome Trust Case Control Consortium is a second sample from the national population-based sample followed periodically from birth to age 44-45 years (http://www.b58cgene.sgul.ac.uk/collection.php); 1,473 individuals were included in the analysis and are distinct from individuals included in the B58C-T1DGC cohort. Blood pressure was recorded using the Omron 705CP machine three times, seated. The average of three readings was used for the analysis.

1.7 BHS

The Busselton Health Study includes a series of seven cross sectional population health surveys of adult residents of the Shire of Busselton in the South-West of Western Australia, undertaken between 1966 and 1990. A cross-sectional community follow-up study in 1994-1995 included the collection of blood for DNA extraction for all survivors of previous surveys. A total of 4,554 individuals participated in this follow-up. BP was measured in the 1994-1995 follow-up study using a standard mercury sphygmomanometer (Baumanometer, New York) as described previously[13]. The participants were asked to refrain from caffeine for 12 hours and to not smoke prior to attending the survey. Three BP readings were recorded on the participant's survey chart to the nearest 2 mmHg and the average of the readings was used for the analyses.

1.8 BLSA

The Baltimore Longitudinal Study of Ageing is an ongoing prospective study of human ageing which started in 1958[14]. The study recruited volunteers predominantly from Washington DC and Baltimore, MD, USA. Healthy volunteers aged >17 years were recruited; only European-origin individuals were included in the analysis, there were no other exclusion criteria. Blood pressure was measured using a mercury sphygmomanometer in the seated position; the average of the 2nd and 3rd readings were recorded for both the right and left arm and used for the analyses.

1.9 BRHS

The British Regional Heart Study is a study comprising 7,753 men aged 40-59, who were recruited from general practices across Great Britain from 1978-1980. A wide range of phenotypic measures was taken for established risk markers such as lipids, blood pressure, and inflammatory and haemostatic markers. Most measures were taken both at recruitment and re-examination in 1998-2000, 20 years after recruitment. Serum samples were taken at the initial examination, and whole blood at the re-examination on 4,252 attendees. Blood pressure readings from the re-examination were used for this analysis. Blood pressure was measured twice in succession on the right arm, with the subject seated and the arm supported, using a Dinamap 1846 oscillometric blood pressure recorder. Over-reading of systolic blood pressure by the instrument was corrected in the analysis and readings were adjusted for observer

variation within each town. The mean of the two measurements was used for analysis (http://www.ucl.ac.uk/pcph/research-groups-themes/brhs-pub).

1.10 BWHHS

British Women's Heart and Health Study is a study comprising 4,286 women aged 60-79 years who were randomly selected from 23 British towns between 1999 and 2001. Of the 4,278 participants who gave consent for genetic testing, 15 were defined by the examining nurse as being non-white and were excluded from further analysis. Of the remaining 4,263 women, 3,800 (89%) had DNA available for genotyping. A Dinamap 1846SX vital signs monitor was used to measure blood pressure, mean arterial pressure and heart rate. Measurements were taken twice in succession, using the right arm, with the participant seated and the arm supported on a cushion. Arm circumference was measured and the appropriate cuff size was used. Since the Dinamap 1846X is known to systematically overestimate systolic blood pressure by 8 mmHg this was subtracted from values before analyses. Full details of the selection of participants and measurements used in the BWHHS can be found at the study website (http://www.lshtm.ac.uk/eph/ncde/research/bwhhs).

1.11 CARe

The Candidate Gene Association Resource (CARe) Study was initiated by NHLBI in 2006. DNA samples and phenotypic information were obtained in individuals of European and African ancestry from 9 NHLBI cohorts (the ARIC study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, CHS, the Cleveland Family Study (CFS), the Cooperative Study of Sickle Cell Disease, the Framingham Heart Study (FHS), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Sleep Heart Health Study[15]. The data used here are from 7,473 African Americans within the CARe study, taken from 5 of the cohorts (ARIC, CARDIA, CFS, JHS, MESA).

1.12 CCMB

The CCMB study comprises four cohorts (CRISIS, IMS, PMNS and WELGEN) and the details of each are described in individual Sections 1.18, 1.40, 1.64, and 1.77.

1.13 CHS

The Cardiovascular Health Study is a population-based cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted across four field centres. The original predominantly white cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional 687 African-Americans were enrolled in 1992-93 for a total sample of 5,888. Details of the study design are summarized elsewhere[16]. A total of 1,908 persons were excluded from the study sample due to prevalent coronary heart disease (N=1,195), congestive heart failure (N=86), peripheral vascular disease (N=93), valvular heart disease (N=20), stroke (N=166), or transient ischemic attack (N=56). Participants with missing BMI (N=10) or BP measurements (N=8) were also excluded. Research staff with central training in blood pressure measurement assessed repeated right-arm seated systolic and diastolic blood pressure levels at baseline with a Hawksley random-zero sphygmomanometer. Means of the repeated blood pressure measurements from the baseline examination from 3,277 CHS subjects of European ancestry were used for the analyses.

1.14 CLHNS

The Cebu Longitudinal Health and Nutritional Survey is an ongoing study of a cohort of Filipino infants born between 1983 and 1984 and their mothers that has been described previously[17]. The CLHNS was originally conceptualized as a study of infant feeding patterns

within the first two years postpartum, with the idea of studying how infant feeding decisions by the household interact with various social, economic, and environmental factors to affect health, nutritional, demographic, and economic outcomes. The study was expanded beyond the two-year postpartum period to date (latest survey on the mothers was in 2007; on index children in 2009) so that more issues related to selected health, demographic, and nutritional outcomes could be addressed. Blood pressure was measured while participants were seated and rested. This analysis used the mean of three measurements taken using mercury sphygmomanometers in 2005.

1.15 CLUE

The CLUE study comprises 2 large cohorts of volunteers who donated blood in 1974 and 1989 (CLUE I & CLUE II) in Washington County, MD, a study funded by the National Cancer Institute. CLUE I (based on the campaign slogan ''Give us a Clue to Cancer and Heart Disease'') consists of 26,147 individuals: CLUE II (Campaign against Cancer and Heart Disease) is an expansion of CLUE I and had 32,894 participants, collected in 1989. Approximately 30% of adult residents participated in both visits. Active participants have been followed up in 1996, 1998, 2000, 2003, and 2007 using questionnaires. The blood pressure measurement procedure differed between 1974 and 1989: In 1974 a nurse took three blood pressure measurements with a standard sphygmomanometer and recorded the lowest value; in 1989 a nurse used a random-zero sphygmomanometer to take three blood pressure measurements following a standard procedure and the third measurement was used[18]. Individuals overlapping with the ARIC study were removed. Clinical data and DNA samples from 7,065 subjects from CLUE were used in this study.

1.16 COBRA

The Control of Blood Pressure and Risk Attenuation study is a Wellcome Trust funded cluster randomised control trial with 2x2 factorial design of BP lowering in Karachi, Pakistan. Using multistage cluster sampling techniques, 12 communities were randomly selected from middle to low-income areas. The total number of subjects aged 5 years or above in the study clusters was 17,500. Out of these, one individual in the age groups of 5-14 and 15-39 years was randomly selected from each of the households and subjects aged 40 and over were also invited for interview and measurements. COBRA participants completed standardized clinical examinations and questionnaires at study baseline and a follow-up visit after median follow-up period of two years. Resting brachial blood pressure was measured with a calibrated automated device (Omron HEM-737 Intellisense TM Blood Pressure Monitor) in the sitting position after 5 minutes of rest. Three consecutive readings were obtained 5 minutes apart and the last two were used in the analysis. Research staff was trained in standardized measurement techniques with retraining at frequent intervals throughout the duration of the study. Blood samples were collected in all subjects aged 40 years and above (N=3,143) and COBRA contributed 2,131 subjects to the analysis.

1.17 CoLaus

The Cohorte Lausannoise is a population-based study aimed at assessing the prevalence and molecular determinants of cardiovascular risk factors in the population of Lausanne, Switzerland[19]. Participants in the study (4,969) were randomly selected from the population register of Lausanne in 2003 (N=56,694, aged 35-75 years). All individuals were of European origin, defined as having both parents and grandparents born in a defined list of European countries. Blood pressure was measured using the Omron HEM-907 machine, in the seated position. Three measures were taken on the left arm; the mean of the last two measures was used in the analyses.

1.18 CRISIS

The Coronary Risk of Insulin Sensitivity in Indian Subjects study recruited participants from rural, urban slum and urban middle-class locations from Pune, Western India to investigate the potential relationship of body fat with measures of insulin resistance and secretion, inflammatory markers, and glycaemia. Extensive phenotype measurements were taken for all participating individuals and blood pressure was measured in the supine position using an automated machine (UA 767PC; A and D Instruments, Abingdon, Oxford, UK).

1.19 DGI

The Diabetes Genetics Initiative (DGI) is a type 2 diabetes (T2D) case-control study of Swedish and Finnish individuals matched on age, gender and BMI[20]. GWAS data from 1,467 normoglycemic male and female controls were included in these analyses. Blood pressure traits were the average of two seated measurements using a mercury sphygmomanometer.

1.20 DiaGen

The Diabetes GENetic study is a large, prospective study of diabetes pathophysiology and genetics that began in 1997[21]. All participants had intensive metabolic, anthropometric and clinical phenotyping assessment and were followed up over many years. Blood pressure was measured automatically following the Riva Rocci method twice on both arms each after 30 minutes resting.

1.21 EAS

The Edinburgh Artery Study is an age-stratified random sample of men and women, aged 55–74 years, which was selected between August 1987 and September 1988 from the age–sex registers of ten general practices with a geographical and socio-economical catchment population spread throughout the city of Edinburgh, UK. Subjects were excluded if they were unfit to participate (e.g., due to severe mental illness or terminal disease); excluded individuals were replaced by other randomly sampled subjects. Physical examinations were performed by specially trained research nurses, using standardised operating procedures. Systolic and diastolic (phase V) blood pressures were recorded in the right arm, after 10 minutes rest in the supine position, using a Hawksley random zero sphygmomanometer. Participants with missing SBP or DBP measurements were excluded from analysis. Overall, 904 participants with valid measurements of SBP and DBP were included in this analysis.

1.22 ELSA

The English Longitudinal Study of Ageing is a national cohort of participants (48% men) aged over 50 years recruited from the Health Surveys for England in 1998, 1999, and 2001. Genetic data were collected at wave 2 of the study (2004/5); the phenotype measurements taken at wave 2 were used for this study. Participants were visited in the home, and blood pressure measurements were taken using an Omron HEM-907 blood pressure monitor by a nurse. Three measurements in the seated position following 10 minutes of rest were taken. Heavy physical activity, smoking, and alcohol use were avoided for 30 minutes prior to recording the blood pressure measurement.

1.23 EPIC-Norfolk

The European Prospective Investigation of Cancer is a population-based cohort study of 25,663 Europid men and women aged 39-79 years recruited in Norfolk, UK between 1993 and 1997[22]. 2,100 randomly selected control subjects were chosen from a BMI study in which genome-wide genotyping data had been obtained. Blood pressure was measured using the

Accutorr oscillometric BP machine; the mean of two readings was taken and used in the analysis.

1.24 EPIC Turin

The EPIC TURIN study is a longitudinal cohort of 10,603 volunteers, aged 35-64 years at baseline, from the Turin area, Italy. Blood pressure was measured using a mercury sphygmomanometer, in seated position, on the left arm. Full details of this cohort have been published previously[23].

1.25 ERF

The ERF (EUROSPAN) study is a family based study which includes over 3,000 participants descending from 22 couples living in the Rucphen region, the Netherlands, in the 19th century. All descendants were invited to visit the regional clinical research centre where they were examined and a fasting blood sample was drawn. All participants filled out a questionnaire on risk factors. The participants included in these analyses consisted of the first series of participants.

1.26 EUROSPAN

The EUROSPAN study consists of 5 cohorts (ERF, MICROS, NSPHS, ORCADES, and VIS), which are described in Sections 1.25, 1.55, 1.62, 1.63 and 1.78.

1.27 FBPP-GenNet

The Family Blood Pressure Project GenNet study recruited European-American (N=1,497) and African-American (N=1,101) participants at two field centers between 1995 and 2003, based on a hypertensive proband[24]. Non-Hispanic white subjects were recruited from Tecumseh, Michigan, and African-American subjects were recruited from Maywood, Illinois. Probands were defined as individuals aged 18-50 years with blood pressures in the upper 20th to 25th percentile of the age/gender-specific blood pressure distribution. Once the proband was identified, an attempt was made to enroll all siblings and parents of the proband, irrespective of their blood pressure or hypertension treatment status. Blood pressure measurements were carried out according to standard procedures in a sitting position after a resting period. Subjects were not allowed to smoke or drink coffee before the visit. The average of two manual BP measurements was used as the phenotype. DNA was available for 1,381 European-American and 848 African-American participants (www.biostat.wustl.edu/fbpp/FBPP.shtml).

1.28 FBPP-HyperGen

The Hypertension Genetic Epidemiology Network recruited two types of participants (hypertensive sibships and random samples of subjects) in European-American and African-American samples[24]. Recruitment of the study participants, including the hypertensive probands, was carried out at five field centers based largely on ongoing population-based studies. For European-Americans, HyperGEN recruited and characterized a total of 1,142 hypertensive subjects from 480 sibships, yielding a total of 992 self-reported sib-pairs, and a random sample of 472 biologically unrelated participants. For African-Americans, HyperGEN recruited and characterized a total of 1,261 hypertensive subjects from 596 sib-ships yielding a total of 826 self-reported sib-pairs, and a random sample of 446 biologically unrelated African Americans (www.biostat.wustl.edu/fbpp/FBPP.shtml). Blood pressure measurements were carried out according to standard procedures in a sitting position after a resting period.

1.29 Fenland

The Fenland Study is an ongoing population-based cohort study (started in 2005) designed to investigate the association between genetic and lifestyle environmental factors and the risk of obesity, insulin sensitivity, hyperglycemia, and related metabolic traits in men and women aged 30 to 55 years. Potential volunteers were recruited from general practice sampling frames in the Fenland, Ely, and Cambridge areas of the Cambridgeshire Primary Care Trust in the UK. Exclusion criteria for the study were: prevalent diabetes, pregnant and lactating women, inability to participate including terminal illness, psychotic illness, or inability to walk unaided. Currently, the study comprises more than 3,000 participants of whom the first 1,500 volunteers with complete anthropometric data were genotyped and included in the current analyses. All participants were measured at the MRC Epidemiology Unit Clinical Research Facilities in Ely, Wisbech and Cambridge. Blood pressure measurements were taken with an Accutorr automated sphygmomanometer using the average of three measurements made at one-minute intervals with the participant seated for 5 minutes prior to measurement. Of the 1,500 individuals that were genotyped 98 individuals were excluded as their genotyping data did not meet the quality control criteria applied such that 1,402 individuals were included in the genome-wide association analyses.

1.30 FHS

The Framingham Heart Study began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort consisted of 5,124 children and spouses of children of the original cohort. The mean age of the offspring cohort was 37 years; 52 percent were women. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled beginning in 2002. Details of study designs for the three cohorts are summarized elsewhere[25-27]. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was calculated. Systolic and diastolic blood pressures were measured twice by a physician on the left arm of the resting and seated participant using a mercury column sphygmomanometer. Pressures were recorded to the nearest even number. The means of two separate systolic and diastolic blood pressure readings at the first clinic examination of each cohort were used for GWAS analyses. For a subset of offspring cohort participants only one measurement was obtained. Individuals under 20 years of age, those who had a myocardial infarction, or congestive heart failure were excluded from the analyses because those conditions may affect blood pressure levels.

1.31 FINRISK97

FINRISK97 is a population-based, cross-sectional survey conducted in 1997 designed to study the prevalence of cardiovascular risk factors in Finland. Genotypes were available in 7,023 men and women free of exclusions. Blood pressures in Finrisk97 were averaged from 2 measures using a mercury column sphygmomanometer in seated participants resting for at least 5 minutes[28].

1.32 FLEMENGHO

FLEMENGHO is a family based population sample from a geographically defined area in North Belgium, and patients were recruited from August 1985 until December 2005[29]. The study population included 3,108 subjects. Blood for DNA extraction could not be obtained from 422 participants, and we also excluded 323 teenagers and 1,076 adults with lower than median age at enrolment (41.4 years). At the enrolment home visit, trained nurses measured

anthropometric characteristics and blood pressure. Blood pressure was measured five times consecutively, after the subject had rested for at least 5 minutes in the seated position. All participants had systolic and diastolic (Phase V) blood pressures recorded within 2 mmHg to the nearest even number using a standard mercury sphygmomanometer. The readings from 2 home visits were averaged and used for the analysis (2x5).

1.33 FUSION

The Finland-United States Investigation Of NIDDM Genetics study aims to discover variants predisposing to type 2 diabetes (T2D) and T2D-related quantitative traits (http://fusion.sph.umich.edu/)[30]. The FUSION GWAS sample includes 1,161 Finnish T2D cases and 1,174 normal glucose tolerant (NGT) controls and 122 offspring of case/control pairs (1 T2D, 119 NGT, 2 with impaired glucose tolerance). GWAS data from the controls and NGT offspring were used for these analyses. The blood pressure trait was the average of two seated measurements using a mercury sphygmomanometer after 5 minutes of rest[28]. FUSION analyses were adjusted for birth province.

1.34 FUSION2

The Finland-United States Investigation Of NIDDM Genetics controls are an independent sample from the FUSION study, these were used for stage 2 targeted genotyping; cohort details and blood pressure measurements are the same as described above for the FUSION study.

1.35 GRAPHIC

The Genetic Regulation of Arterial Pressure of Humans In the Community (GRAPHIC) study comprises 2,037 white European subjects from 520 nuclear families recruited from the general population via participating family practitioners in Leicestershire, UK[31]. Intensive cardiovascular phenotyping included 26-hour ambulatory blood pressure measures in all individuals and clinic blood pressure measures, for which three readings were made using an Omron HEM-705CP digital blood pressure monitor. Clinic blood pressure was defined as the mean of the second and third blood pressure readings. For the purposes of this analysis, clinic blood pressure recordings were used.

1.36 HABC

The Health ABC study is a prospective cohort study investigating associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling white (N=1,794) and black (N=1281), men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Blood pressure was collected at baseline using a conventional mercury sphygmomanometer, cuff, and stethoscope. Participants were allowed to rest quietly for 5 minutes prior to measurement. Caffeine, eating, heavy physical activity, smoking, and alcohol use were avoided for 30 minutes prior to recording the blood pressure.
http://www.nia.nih.gov/ResearchInformation/ScientificResources/HealthABCDescription.htm

1.37 HUFS

The Howard University Family Study is a population based family study of African-Americans in the Washington DC metropolitan area, USA[32]. The major objective of the HUFS was enroll and examine a randomly ascertained cohort of African-American families, along with a set of unrelated individuals, to study the genetic and environmental basis of common complex diseases including hypertension, obesity, and associated phenotypes.

Participants were sought through door-to-door canvassing, advertisements in local print media, and at health fairs and other community gatherings. Families were not ascertained based on any phenotype. During a clinical examination, blood pressure was measured in the sitting position using an oscillometric device (Omron). Three blood pressure readings were taken with a ten-minute interval between readings. The reported systolic and diastolic blood pressure readings were the average of the second and third readings.

1.38 HUNT 2

The Nord-Trøndelag Health Study samples were from the county of Nord-Trøndelag in central Norway and selected from the HUNT 2 study. Details of the study sample have previously been described[33]. Blood pressure was measured by specially trained nurses or technicians, using a Dinamap 845XT (Critikon), based on oscillometry. Cuff size was adjusted after measuring the arm circumference. The Dinamap was started after the participant had been seated for two minutes and blood pressure was measured automatically three times at one-minute intervals. The mean of the second and third measures were used in the analyses.

1.39 HYPEST

HYPertension in ESTonia is a population based case-cohort sample set consisting of unrelated subjects recruited between 2004 and 2007 across Estonia[34]. The aim of the study was to find hypertension risk factors in the Estonian population. All individuals have detailed epidemiological data and a documented history of multiple SBP and DBP readings (an average of 4.3 readings per individual) during a mean of 3.2 years. The HYPEST essential hypertension patients were selected based on the clinical diagnosis and profile of blood pressure specialists during the patients' ambulatory visits or hospitalization at the North Estonia Medical Center, Centre of Cardiology, or at the Cardiology Clinic, Tartu University Hospital, Estonia. The HYPEST healthy control cohort was recruited from among the long-term blood donors. After resting in the seated position, blood pressure readings were taken by a trained clinician using a standard mercury column sphygmomanometer with arm-circumference adjusted cuffs.

1.40 IMS

The Indian Migration Study is a sib-pair comparison study including urban factory workers who had migrated from rural areas together with their rural-dwelling sibling. Recruitment was from 4 Indian factories: Lucknow, Nagpur, Hyderabad and Bangalore. A 25% random sample of non-migrants was invited to participate in the study. Each participant was asked to invite one non-migrant full sibling of the same sex and closest to them in age still residing in their rural place of origin. Precedence was given to gender over age and where multiple sibs were available the one closest in age was invited. Non-migrants were also asked to invite a sib who resided in the same city but did not work in the factory. A wide range of phenotypic parameters were measured, including blood pressure. BP was measured using an automatic oscillometric device (Omron HEM 705 CP, Omron, Matsusaka Co, Japan) in the sitting position using the right upper arm and an appropriate sized cuff after a period of 5 minutes rest. Two readings were taken for BP and their average has been used in analyses.

1.41 InCHIANTI

The Invecchiare in Chianti study is a representative population-based study of older people living in the Chianti area of Tuscany, Italy[35]. All participants were >21 years of age and of white European origin. Blood pressure was measured using a mercury sphygmomanometer in the supine position; the average of the 2nd and 3rd readings was used for the analysis.

1.42 INTERGENE

This is a population based research program established to assess the INTERplay between GENEtic susceptibility and environmental factors for the risk of chronic diseases in western Sweden[36,37]. Randomly sampled women and men aged 25-74 years and living in the Västra Götaland region between 2001 and 2004 were invited to participate. For the purpose of the present study, 153 subjects of non-European origin were excluded and genotyping data were unavailable for a further 614 subjects, leaving 2,843 subjects with European ancestry. All subjects provided extensive questionnaire information on risk factors, medication, and diseases and were invited for a clinical examination by a team of trained research nurses. This included ECG, spirometry, blood sampling for genetic studies, as well as anthropometric and blood pressure measurements. Blood pressure was measured twice after a 10-minute rest, with a validated automatic device (Omron 711 Automatic IS; Omron Healthcare Inc, Vernon Hills, IL)[38] in the supine position, and the mean of two valid measurements was used for analyses.

1.43 IRAS

The IRAS Family Study was designed to study the genetics of insulin resistance and visceral adiposity. A cohort of 1,856 participants was initially recruited in 1999-2002 from three US communities[39]. Recruitment for the IRAS Family Study was based on recruiting family members of the original IRAS Study index cases on the basis of large family size, and not on extreme phenotype (e.g., insulin resistance, obesity). African-American families were recruited in Los Angeles, CA. Resting seated blood pressure was measured three times using a mercury manometer, after a 5-minute rest by centrally trained technicians using identical equipment. The mean of the last two measurements was used to calculate blood pressure used in the analyses.

1.44 Jamaicans-GXE

Jamaican-GXE participants were recruited from the city of Kingston, Jamaica as part of a larger project to examine gene by environment interactions on blood pressure among adults 25-74 years. Participants were either identified from the records of the Heart Foundation of Jamaica, a non-governmental organization based in Kingston, which provides low-cost screening services (height and weight, blood pressure, glucose, cholesterol) to the general public, or from among participants in family studies of blood pressure at the Tropical Metabolism Research Unit and from among staff members at the University of the West Indies, Mona. Screenees were eligible to participate if their body mass index (BMI) was in either the top or bottom third of BMI for the Jamaican population[40]; all participants were unrelated to each other. BP was measured 3 times in the brachial fossa in the sitting position with a mercury sphygmomanometer according to procedures described previously[41]; the mean of the last two measurements was used in the analysis.

1.45 Jamaicans-Spanish Town

Jamaican-Spanish Town participants were recruited from, in, and around Spanish Town, a stable, residential urban community neighboring the capital city of Kingston, Jamaica, as part of the International Collaborative Study of Hypertension in Blacks (ICSHIB) described in detail elsewhere[40]. A stratified random sampling scheme was used to recruit adult males and females aged 25 years and older. BP was measured 3 times in the brachial fossa in the sitting position with a mercury sphygmomanometer according to procedures described previously[41] and the mean of the last two measurements was used in the analysis.

1.46 Japanese (Takeuchi 2010)

These summary association results were extracted from Supplementary Table 2 of the work by Takeuchi et al.[42], which reported association statistics for SBP and DBP (adjusted by +15mmHg and +10mmHg for individuals on medication) with genotypes at some SNPs previously published. These data derive from a sample of 1,526 Japanese individuals genotyped using the Infinium HumanHap550 BeadArray.

1.47 JMGP

The Japanese Millenium Genome Project comprises 7 independent study cohorts for studies of cardiovascular diseases and related risk factors. The Ohasama, Shigaraki, Takashima, Suita, and Nomura Ehime studies are general population based genetic epidemiological studies of subjects recruited via a medical check-up process for community-residents. The Ohasama study is a population-based longitudinal epidemiological study focusing on the clinical implications of home BP measurement[43]. The Shigaraki[44] and Takashima[45] studies of Shiga University of Medical Science are general population-based longitudinal studies and subjects were recruited through a community-based annual medical check-up process. The Suita study is based on the residents of Suita city, an urban city located in the second largest area of Osaka, Japan[46]. Subjects were recruited through a biennial medical check-up process of the National Cardiovascular Center. The Ehime (Nomura study) is a longitudinal epidemiological study based on the Nomura Town residents[47]. The Yokohama (Yokohama City University) and Matsuyama (Ehime University) cohorts are derived from employees of large manufacturing industries located in Kanagawa and Matsuyama City, Ehime Prefecture (western part of Japan)[48], respectively. In our analyses, our descriptor Ehime includes individuals from both the Nomura Ehime population based samples and Matsuyama cohort (University cohort) combined. In all cohorts, clinical parameters were obtained from personal health records during the annual or biennial medical check-up process.

1.48 KARE

The Korea Association REsource project was initiated in 2007 to perform large-scale genome-wide association analyses of the Ansung and Ansan population-based cohorts in Korea[49]. The cohorts were collected as part of the Korean Genome Epidemiology Study and included 5,018 Ansung and 5,020 Ansan inhabitants between 40 and 69 years of age. Individuals were collected in the Gyeonggi Province, close to Seoul, Republic of Korea. All participants have been examined every two years since baseline, and more than 260 traits have been examined. Blood pressure measurements were taken three times in the supine position, following resting for 5 minutes. The average of 3 readings was used for these analyses.

1.49 KORA

The KOoperative Gesundheitsforschung in der Region Augsburg (third survey: S3/F3) is an epidemiological cohort recruited from the general population of Augsburg, Germany in 1994-1995[50,51]. A subset of this survey (1,644 subjects), were genotyped using the Affymetrix 500K array (http://epi.helmholtz-muenchen.de/kora-gen/). In this study subjects with BMI<35 kg/m^2 were included; diabetics were excluded. Final number of subjects entering the association analysis with blood pressure was 1,503. Blood pressure was measured using a random zero sphygmomanometer in the seated position at the first examination cycle. Three measurements were taken at least three minutes apart and the numbers entering the database were the mean of the last two measurements.

1.50 KORCULA

The KORCULA study sampled Croatians from the Adriatic island of Korcula (N=969), between the ages of 18 and 88. The fieldwork was performed in 2007 in the eastern part of the island, targeting healthy volunteers from the town of Korčula and the villages of Lumbarda, Žrnovo and Račišće[52]. Over 150 quantitative traits were measured to each participant. Blood pressure was measured using standard procedures, briefly, the subject was seated in a quiet room, and they were advised to not have done any exercise, or have been exposed to the cold, eaten, or smoked for 30 minutes prior to the recording. Following 5 minutes of rest, blood pressure was recorded twice during the examination and the mean of the two readings was used for the analyses.

1.51 LOLIPOP

The LOndon LIfe Sciences POPulation study is an ongoing population-based cohort study of ~30,000 individuals (18,000 Indian Asians and 12,000 European white men and women), aged 35-75 years and recruited from the lists of 58 general practitioners in West London, United Kingdom[53,54]. Blood pressure was measured using an Omron 705CP sphygmomanometer (mean of 3 measurements) with the subject seated. For the European ancestry validation stage of the present study, lookups were performed in 3 subgroups of European White (EW) ancestry that had been genotyped using 3 GWAS platforms - LOLIPOP (EW_A), LOLIPOP (EW_P) and LOLIPOP (EW_610) - comprising in total 1,603 individuals. For the South Asian ancestry replication stage of the present study, we used genotype data from 7 SNPs that had been genotyped in 12,900 individuals during previous work[55], and complemented these with lookups for 22 SNPs in 2 subgroups of Indian Asian ancestry that had been genotyped using 2 GWAS platforms - LOLIPOP (IA317) and LOLIPOP (IA610) - comprising in total 8,688 individuals. Analyses in GWAS datasets were adjusted for ancestry principal components, and analyses in the larger IA sample were adjusted for self-reported religion.

1.52 MDC

The Malmö Diet and Cancer study is a community-based prospective epidemiologic cohort of 28,449 persons recruited for a baseline examination between 1991 and 1996. From this cohort, 6,103 persons were randomly selected to participate in the Cardiovascular Cohort (MDC-CC), which seeks to investigate risk factors for cardiovascular disease[56]. Blood pressure was measured using a mercury sphygmomanometer once after 10 minutes of rest in the supine position.

1.53 MESA

The Multi-Ethnic Study of Atherosclerosis investigation is a population-based study of 6,814 men and women age 45 to 85 years, without clinical cardiovascular disease, recruited from six United States communities (Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles County, CA; northern Manhattan, NY; and St. Paul, MN). The main objective of MESA is to determine the characteristics of subclinical cardiovascular disease and its progression. Sampling and recruitment procedures have been previously described in detail[57]. Adults with symptoms or history of medical or surgical treatment for cardiovascular disease were excluded. During the recruitment process, potential participants were asked about their race/ethnicity. Self-reported ethnicity was used to classify participants into groups[58]. After a 5-minute rest BP was measured three times at 1 minute intervals using a Dinamap PRO 100 automated oscillometric device (Critikon, Tampa, FL) with the subject in seated, and the average of the second and third BP measurements was used in the analysis. Additional individuals were derived from the MESA Family Study, an ancillary study to MESA whose goal is to identify

genes contributing to the risk for cardiovascular disease, by looking at the early manifestations of atherosclerosis within families, mainly siblings. MESA Family studied siblings of index subjects from the MESA study and sib-pairs in new families ascertained through index subjects meeting MESA enrollment criteria. In a small proportion of subjects, parents of MESA index subjects participating in MESA Family were studied but only to have blood drawn for genotyping. The MESA Family cohort was recruited from the six MESA Field Centers during May 2004 - May 2007. The number of non-classic MESA family members recruited was 1,633 (950 African-Americans and 683 Hispanic-Americans) from 594 families, yielding 3,026 sib-pairs. Participants underwent the same examination as MESA participants.

1.54 METSIM

The **MET**abolic **S**yndrome **I**n **M**en study includes men aged 45-72 years, randomly selected from the population of the town of Kuopio, Eastern Finland, Finland (population 95,000). Detailed sample characteristics of this cohort have been previously reported[59]. The present analysis is based on the first 7,055 subjects examined for METSIM. Blood pressure was measured in the seated position after 5 minutes rest using a mercury sphygmomanometer. The average of 3 measurements was used in the analysis.

1.55 MICROS

The **Micro**-Isolates in **S**outh **T**yrol (MICROS, EUROSPAN) study (http://www.biomedcentral.com/1471-2350/8/29) is part of the genomic health care program 'GenNova' and was carried out in three villages of the Val Venosta on the populations of Stelvio, Vallelunga and Martello. This study was an extensive survey carried out in South Tyrol (Italy) in the period 2001-2003. Study participants were volunteers from three isolated villages located in the Italian Alps, in a German-speaking region bordering with Austria and Switzerland. Due to geographical, historical and political reasons, the entire region experienced a prolonged period of isolation from surrounding populations. The 1,096 participants included in this study are those which had both phenotypic and GWAS data available. Blood pressure was taken after 3 minutes rest and 3 consecutive measurements were recorded on the right upper arm using an Omron HEM-705CP. The median for each person was used in the analysis.

1.56 MIGen

The **M**yocardial **I**nfarction **G**enetics Consortium cohort is composed of a subset of the controls of a case-control study aimed at identifying genetic variants associated with early-onset myocardial infarction. Most of the controls are selected from population based cross-sectional or cohort studies and come from five different studies: Heart Attack Risk in Puget Sound (Seattle, USA), REGICOR (Girona, Spain), MGH Premature Coronary Artery Disease Study (Boston, USA), FINRISK (Finland); Malmö Diet and Cancer Study (Malmö, Sweden). There is a minimal overlap of samples between the resources (N=30). For the majority of studies, blood pressure was measured twice using calibrated sphygmomanometers, in the seated position after at least 5 minutes of rest; the mean of the two measurements was used in the analysis. The first two principal components from an identical by state (IBS) analysis were used to adjust for potential population stratification.

1.57 MPP

The **M**almö **P**reventive **P**roject is a screening program for cardiovascular risk factors and comprises 33,346 Swedish subjects (22,444 men and 10,902 women) from the city of Malmö in southern Sweden[60]. There are 14,600 with DNA after removing subjects who were also

participants in MDC-CC (see above). Blood pressure was measured using a mercury sphygmomanometer (mean of 2 measurements) after 10 minutes of rest supine.

1.58 MRC NSHD

The Medical Research Council (MRC) National Survey of Health and Development (NSHD) is an ongoing prospective birth cohort study consisting of a stratified random sample of all births in England, Scotland and Wales in one week in March 1946 (http://www.nshd.mrc.ac.uk/)[61]. The original cohort comprised 2,547 women and 2,815 men who have been followed up over 20 times since their birth. In 1999, when the cohort members were aged 53 years, 2,989 individuals were interviewed in their homes by research nurses. During these visits, blood pressure was measured twice, with the survey member seated and after 5 minutes of rest, using an Omron HEM-705 automated digital oscillometric sphygmomanometer (Omron); the 2nd blood pressure reading was used for this analysis.

1.59 NFBC1966

The North Finland Birth Cohort of 1966 was designed to study factors affecting preterm birth, low birth weight, and subsequent morbidity and mortality (http://kelo.oulu.fi/NFBC/). The longitudinal data collection includes clinical examination and blood sampling at age 31 years, from which data in the current study are drawn. The attendees in the follow-up (71% response rate) were adequately representative of the original cohort[62] as is the final study sample in the present analyses. 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.

1.60 Nigerian

The participants were recruited from Ibadan and Igbo-Ora, Yoruba-speaking communities in southwest Nigeria as part of a long-term study on the environmental and genetic factors underlying hypertension[63]. A screening examination was completed by trained research staff, using a standardized protocol[40] and trained local interviewers obtained a medical history and a family history in the participant's native language. BP observers were trained and certified by a previously described procedure[40]. An automated device (Omron HEM-412C) was used for all BP measurements and three measurements were taken three minutes apart. The average of the final two readings was used in the analysis.

1.61 NPHS-II

The Northwick Park Heart Study II is a prospective study of 3,012 healthy middle-aged men aged 50-64 years at recruitment, sampled from nine UK general practices between 1989 and 1994. Full details of recruitment, measurements, follow-up and definitions of incident disease have been reported elsewhere[64]. Exclusion criteria were: history of unstable angina or acute myocardial infarction, a major Q wave on the ECG, regular anti-platelet or anticoagulant therapy, cerebrovascular disease, and life-threatening malignancy. Blood pressure was recorded with a random-zero sphygmomanometer (average of 2 measurements) at baseline and on five following annual visits. Baseline measures were used for these analyses.

1.62 NSPHS

The Northern Swedish Population Health Study (EUROSPAN) represents a family-based prospective population study located in the parish of Karesuando, in the subarctic region of the County of Norrbotten, Sweden. This parish has about 1,500 inhabitants, of whom 740 participated in the study. Historic population accounts show that there has been little

immigration or other dramatic population changes in this area during the last 200 years. The study includes a comprehensive health investigation and collection of data on family structure, lifestyle, diet, medical history, and samples for clinical chemistry, RNA and DNA analyses. Blood pressure was taken once by the auscultatory method using a sphygmomanometer and a stethoscope.

1.63 ORCADES

The Orkney Complex Disease Study (EUROSPAN) is an ongoing family-based cross-sectional study in the isolated Scottish archipelago of Orkney. Genetic diversity in this population is decreased compared to Mainland Scotland, consistent with the high levels of endogamy historically. Data for participants from a subgroup of ten islands were used for this analysis. Fasting blood samples were collected and over 200 health-related phenotypes and environmental exposures were measured in each individual. BP was recorded twice five minutes apart using a calibrated Omron digital sphygmomanometer, after at least 10 minutes of supine rest; the mean of the readings was used for the analyses.

1.64 PMNS

The Pune Maternal Nutritional Study is the first Birth Cohort study in India to investigate the relationship between maternal nutrition and offspring risk of type 2 diabetes and cardiovascular disease. The study recruited non-pregnant married women in six villages near Pune, Western India. The subjects included in the ICBP study are parents of children studied in the PMNS. Blood pressure was measured in the supine position using an automated machine (UA 767PC; A and D Instruments, Abingdon, Oxford, U.K.).

1.65 PREVEND

The Prevention of REnal and Vascular ENd stage Disease 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[65,66]. 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.

1.66 PROCARDIS

The Precocious Coronary Artery Disease study (PROCARDIS) (www.procardis.org) is a European consortium investigating the genetics of precocious coronary artery disease (CAD) in German, Italian, Swedish, and British CAD patients and controls[67]. Country of origin was a covariate in all analyses. The controls were included in this study; these had no personal history of CAD, hypertension, or diabetes. Blood pressure was measured twice using various sphygmomanometers, in the seated position after at least 5 minutes of rest; the mean of the two measurements was used.

1.67 PROMIS

Pakistan Risk Of Myocardial Infarction Study is a case-control study of acute first-ever MI in urban Pakistan. A locally piloted and validated epidemiological questionnaire was administered to participants by medically qualified research officers that sought more than 200

items of information in relation to ethnicity, demographic characteristics, lifestyle factors (e.g., tobacco and alcohol consumption, dietary intake, and physical activity), personal and family history of cardiovascular disease, and medication use. The control subjects were included in this study. Blood pressure was measured by research medical officers who were registered physicians using standard blood pressure apparatus. Measurements were taken twice ten minutes apart whilst the participants were seated. The first reading was taken at least 15 minutes after an eligible participant had arrived at the recruitment centre; the value used in the analyses is the mean of the two measurements.

1.68 Prospect-EPIC

Prospect-EPIC is one of the two Dutch contributions to the European Prospective Investigation into Cancer and Nutrition (EPIC)[68]. Participants were recruited between 1993 and 1997 among women living in Utrecht and its vicinity and who attended the regional population-based breast cancer screening program. A total of 17,357 women aged 49-70 years were included. For laboratory analysis a 10% random sample of 1,736 samples was used. Blood pressure was measured using an automated and calibrated Oscillomat (Bosch & Son, Jungingen, Germany); the average of two readings after 10 minutes rest in the seated position was used for the analysis.

1.69 RS-I and RS-II

The Rotterdam Study (RS-I) and Rotterdam Extension Study (RS-II) are prospective population-based cohort studies; the RS-I comprises 7,983 subjects aged 55 years or older. Participants completed an interview at home and at the research centre, where participants were subsequently examined. Baseline data were collected between 1990 and 1993. In 1999, inhabitants who turned 55 years of age or moved into the study district since the start of the study were invited to participate in an extension of the RS (RS-II), 3,011 participated (67% response rate). The rationale and design of the RS have been described in detail elsewhere[69]. At the research centre, we obtained two seated blood pressure measurements of the right brachial artery with a random zero sphygmomanometer. The mean of two consecutive measurements was used in association analyses. We excluded participants who were older than 85 years of age and those who had a history myocardial infarction or congestive heart failure, because of the impact of these conditions on blood pressure levels.

1.70 SardiNIA

The SardiNIA study is a longitudinal study examining age-related quantitative traits in individuals from the Ogliastra region of Sardinia, Italy[70]. The SardiNIA GWAS examined 4,305 related individuals (age >14 years), of whom 3,998 individuals were included in this study. Blood pressure was measured using a mercury sphygmomanometer; the average of the second and third reading was used for the analyses.

1.71 SHIP

The Study In Health In Pomerania is a population-based survey in West Pomerania, the northeast area of Germany[71]. A sample from the adult population aged 20 to 79 years was drawn based on population registries of cities and towns in the region. SHIP finally comprised 4,308 participants (corresponding to a final response rate of 68.8%). 3,310 individuals with GWAS data were included in this study. Blood pressure was measured three times, seated, after 5 minutes of rest, using a digital blood pressure monitor (HEM-705CP, Omron Corporation, Tokyo, Japan), after a rest period of 3 minutes for each measurement. The mean of the second and third measurements was used in the analyses.

1.72 SiMES

Singapore Malay Eye Study is a population-based cross-sectional epidemiological study of 3,280 individuals from one of the three major ethnic groups residing in Singapore[72,73]. All subjects were Malay and aged 40-80 years. Two readings of blood pressure were taken from participants after 5 minutes of rest, seated, 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 was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated.

1.73 SP2

Singapore Prospective Study Program is a population-based study of diabetes and cardiovascular disease in Singapore that has been described previously[74]. The SP2 has recruited 10,633 Chinese, Malay, and Indian subjects from four cross-sectional studies that were conducted in Singapore between 1984 and 1998. Subjects were aged 18-69 at baseline and represented a random sample of the Singapore population. Two readings of blood pressure were taken from participants after 5 min of rest, seated, 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 was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated.

1.74 SPLIT

The SPLIT study is an ongoing cross-sectional study that samples Croatians from the town of Split, between the ages 18 and 85. The sampling started in 2008, and continues throughout 2010[75]. A wide range of phenotypic measurements is available. Blood pressure was measured using standard procedures. Briefly, the subject was seated in a quiet room and they were advised to not have done any exercise, or have been exposed to the cold, eaten, or smoked for half an hour prior to the recording. Following 5 minutes of rest, blood pressure was recorded twice during the examination and the mean of the two readings was used for the analyses.

1.75 SU.VI.MAX

The **SUpplementation en VItamines et Mineraux AntioXydants study** is a longitudinal study performed on a national sample of healthy volunteers from France between 1996 and 2001. 1,823 individuals, aged 35-65 years at baseline were included in this study[76]. Blood pressure was measured using a mercury sphygmomanometer in the seated position; the average of three readings taken from the first examination (1996) was used in the analysis.

1.76 TwinsUK

The TwinsUK Study comprises a sample of healthy female whites recruited through the TwinsUK registry in London (http://www.twinsuk.ac.uk/). All participants were recruited from the general population without presence or interest in any particular disease or trait through national media campaigns. One of each twin pair was selected, with ages ranging from 18 to 76 years. Blood pressure was measured using an Omron HEM-907 machine, seated. Three readings were taken, the first was discarded and the average of the other two was used in the analyses. In the GWAS discovery (stage 0) data from 873 Twins UK participants was used (**TwinsUK**). During the study data from a further 2,163 independent individuals

from study became available for analyses (**TwinsUK2**), which were used for lookups in stage 3 validation.

1.77 WELGEN

The <u>Wel</u>come <u>Gen</u>etic study comprises consecutively recruited young type 2 diabetes patients from the Diabetology Research Centre, King Edward Memorial Hospital and Research Centre (KEMHRC), Pune, Western India. Patients were diagnosed and classified according to WHO 1999 criteria, subjects with ketoacidosis at diagnosis, clinically judged to be insulin dependent, with exocrine pancreatic disease (fibrocalculous pancreatic diabetes) and those who fulfill clinical criteria of Maturity Onset Diabetes of the Young (MODY) were excluded from the study. Phenotypic details were collected using an investigator-administered questionnaire and blood pressure, anti-hypertensive medications, lipid profile and various obesity measures were also collected. BP measurements were made in the seated position after 5 minutes rest using a Diamond Regular BP apparatus (India).

1.78 VIS

The **VIS (Croas; EUROSPAN) study** includes Croatians, aged 18 to 93 years, who were recruited during 2003 and 2004 in a population-based study in the villages of Vis and Komiza on the Dalmatian island of Vis, Croatia[75,77]. Biochemical and physiological measurements were performed, detailed genealogies reconstructed, questionnaire of lifestyle and environmental exposures collected, and blood samples and lymphocytes extracted and stored for further analyses. Blood pressure was measured using standard procedures, briefly, the subject was seated in a quiet room, and they were advised to not have done any exercise, have been exposed to cold, eaten, or smoked for 30 minutes prior to the recording. Following 5 minutes of rest, blood pressure was recorded twice during the examination and the mean of the two readings was used for the analyses.

1.79 WGHS

The <u>W</u>omen's <u>G</u>enome <u>H</u>ealth <u>S</u>tudy (**WGHS**) is a prospective cohort of female North American health care professionals representing participants in the Women's Health Study (WHS) trial who provided a blood sample at baseline and consent for blood-based analyses. Participants in the WHS were 45 years or older at enrolment and free of cardiovascular disease, cancer or other major chronic illness. For the primary WHS endpoints of cardiovascular disease, full medical records were obtained for reported endpoints and reviewed by an endpoints committee of physicians unaware of assignment. The current data are derived from 23,294 WGHS participants for whom whole genome genotype information was available at the time of analysis and for whom self-reported European ancestry could be confirmed by multidimensional scaling analysis of 1,443 ancestry informative markers in PLINK v. 1.06. Baseline BP in the WGHS was ascertained by a self-reported questionnaire, an approach which has been validated in the WGHS demographic, namely female health care professionals[78-80]. Questionnaires recorded systolic blood pressure in 9 categories (<110, 110-119, 120-129, 130-139, 140-149, 150-159, 160-169, 170-179, ≥180 mmHg), and diastolic blood pressure in 7 categories (<65, 65-74, 75-84, 85-89, 90-94, 95-104, ≥105 mmHg). The midpoint of each category was used for analysis. Hypertension was defined as a history of physician-diagnosed HTN and ongoing HTN treatment, or SBP ≥ 140 or DBP ≥ 90 mmHg. To account for treatment effects, 10 and 5 mmHg were added to the measured systolic and diastolic blood pressures respectively, if a participant was taking antihypertensive medication[81].

1.80 WHII

The Whitehall II Study recruited 10,308 participants (70% men) between 1985 and 1989 and involved 20 London based civil service departments. In this longitudinal study blood pressure was recorded at phase 1 (1985-1988), phase 3 (1991-1993), phase 5 (1997-1999) and phase 7 (2003-2004). DNA was stored from phase 7 from over 6,000 participants. The study individuals are all highly phenotyped for cardiovascular and other ageing related health outcomes. Blood pressure measurements were taken 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.

1.81 YFS

Cardiovascular risk in Young Finns Study was set up to determine the contribution of childhood lifestyle, biological, and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents from all over Finland participated in the baseline study. Thereafter these subjects were followed up with several examinations including comprehensive risk factor assessments. The 27-year follow-up was performed in 2007 and the blood pressure measurements at this time point were used for this study. Blood pressure was measured by nursing staff three times using a random-zero sphygmomanometer and the average of the three measurements was taken. Individuals were excluded if BMI, systolic or diastolic blood pressure measurements or genotype data were missing.

1.82 YMCA

Young Men Cardiovascular Association Study is a cohort of 1,157 unrelated young, apparently healthy men recruited from randomly selected secondary schools in Southern Poland[82]. Phenotyping for conventional cardiovascular risk factors included taking medical history, basic anthropometry, and fasting blood biochemistry. Blood pressure measurements were conducted in a sitting position using a mercury sphygmomanometer and three readings were averaged to estimate the final systolic and diastolic blood pressure values.

2. Blood pressure and hypertension association analyses

Most of the studies included in ICBP GWAS are cross-sectional, with observations at a single time-point. Substantial numbers of participants had blood pressure measured while taking antihypertensive or blood pressure-lowering medication. The prescription of these medications generally depends on assessments of blood pressure levels while off medication, with individuals with higher blood pressure generally more likely to be treated. If measured blood pressure values are used for association analyses this generates a bias, which is not corrected (and indeed may be exacerbated) by using medication status as a covariate[83]. Following the recommendation of Tobin, Sheehan et al[83], we imputed the off-treatment blood pressure value for treated individuals by adding 15mmHg to measured SBP and 10mmHg to measured DBP for all treated individuals. We used these imputed values plus measured values for un-treated individuals for all analyses described here. Alternative adjustment methods (addition of 10/5mmHg) gave very similar results in an interim analysis.

In general, association analyses for SBP and DBP were conducted using linear regression, and analyses for dichotomous hypertension using logistic regression. All analyses assumed an additive genetic model, that is, for each SNP we coded the genotype as 0/1/2, indicating the

subject's number of copies of the designated coded allele. For imputed SNPs, we either used (a) the expected number of copies of the coded allele as the genotype or (b) a quadratic approximation to the missing data likelihood[2]. These approaches are both proper score tests and are asymptotically equivalent[84,85].

For dichotomous hypertension, we sought uniform definition across all studies, which in many cases had recorded medications taken by each subject. We defined cases as individuals with $SBP \geq 140mmHg$, or $DBP \geq 90mmHg$, or who were taking anti-hypertensive or blood pressure-lowering medication for any reason (the indication for each medication was typically not recorded). We used all other individuals as controls.

All association analyses were either stratified by sex or used sex as a covariate. In addition, age, age-squared, and BMI were always included as covariates (except where these covariates were uniformly identical, i.e. in cohorts of individuals born in a given year). Where available and appropriate, additional covariates were used to correct for potential within-cohort stratification. In studies with GWAS data these were typically ancestry principal components[86], but additional recruitment site or self-declared ancestry variables were used by some studies. Because of the strong phenotypic correlation between blood pressure and BMI, the inclusion of BMI as a covariate reduces residual variance and may therefore increase power to detect associations between SNPs and blood pressure. However, inclusion of BMI as a covariate could potentially reduce power in some cases, such as for variants of pleiotropic effect with direct impact on both blood pressure and BMI (unless their effect goes in the opposite direction for the two). SNPs that affect blood pressure only via an effect on BMI are not of specific interest in our study since such variants are best identified by GWAS for BMI.[87-89]

For all association analyses, covariates were either (a) included directly in the analysis for each SNP, or (b) used to compute a residual (from regression) that was then tested for association with each SNP in turn. These approaches give extremely similar results in practice.

3. Meta-analysis

We conducted separate genome wide meta-analyses for SBP and DBP. Before meta-analysis, the association results for each cohort were filtered to exclude SNPs not in HapMap, SNPs with alleles different from HapMap, and SNPs with observed/expected ratio of coded genotype scores less than 0.3. The latter ratio is a measure of imputation quality; for an allele with frequency p, the expected variance at Hardy Weinberg proportions for an allele dosage (coded as 0, 1, 2 for a directly genotyped SNP) is 2p*(1-p); for poorly imputed SNPs the allele dosages typically "shrink" towards 2p and the variance of actual scores is often smaller than expected. A genomic control correction was applied[90] by multiplying all standard errors in a given study's results by a factor $\sqrt{\lambda}$, where λ is the median χ^2 (of association statistic)/0.4549. Where studies stratified their analyses by sex, genomic control was applied within each sex stratum. For each SNP, the per-coded-allele effects were combined across studies (and across sex strata, when used) using inverse variance weighted meta-analysis. Finally, a second genomic control correction was applied to the meta-analysis standard errors in the same way as for the individual studies.

Results from stage 0 (discovery) and stages 1&2&3 (validation) were also combined using inverse-variance weights. As stages 1 and 2 contained few SNPs and were greatly enriched for SNPs truly associated with blood pressure, the median χ^2 statistic is not a reliable measure of stratification, so genomic control was not applied here.

All data handling and meta-analyses were conducted by two separate analysts (T.J. and G.B.E.) working independently at separate locations, and were checked to ensure concordance.

4. Sequential experimental design

For discovery of new blood pressure loci, our aim was to maximize statistical power while not 'wasting' resources on genotyping unlikely to materially alter results. We therefore adopted a sequential experimental design with three stages of validation genotyping in samples of European ancestry as illustrated in Supplementary Figure 1. The P-value thresholds used for skipping targeted genotyping at the end of Stages 0 and 1 do not reflect any *a priori* determined most powerful sequential experiment design, but rather reflect the historical process whereby a large GWAS dataset (WGHS, see 1.79) used in Stage 3 only became available after commencement of Stage 2 genotyping, which was restricted to 27 SNPs.

Stage 0 (completed in May 2009) consisted of meta-analysis of GWAS data from N=69,395 individuals plus follow-up genotyping at selected SNPs[55]. At this time, 9 SNPs had sufficiently strong evidence, defined as $P < 2.5 \times 10^{-8}$ and previously reported to be associated with blood pressure after discovery and validation in subsets of our Stage 0 dataset,[55,91] such that no further validation using follow-up genotyping was deemed necessary.

From the remaining SNPs, using an informal combination of the Stage 0 *P*-value, functional and bioinformatic data, and literature searches, we took 27 SNPs into Stage 1 for further genotyping in N=42,644 individuals. These included 4 SNPs at loci previously reported (*ZNF652*, *ULK4*, *TBX5-TBX3*, *CACNB2*)[55,91] but which had stage 0 $P > 2.5 \times 10^{-8}$ and which we therefore thought warranted further effort to confirm or reject association. Note that the stage 0 *P*-values for these 4 SNPs differ slightly from previously published data[55,91] because previous work by the CHARGE consortium used a +10mmHg/+5mmHg increment to impute off-treatment BP, whereas we used a +15mmHg/+10mmHg increment here. In Stage 1 (completed August 2009) *P*-values were calculated for a joint analysis of all Stage 0 and 1 data. After Stage 1, nine of the 27 SNPs had $P < 10^{-10}$, and no further genotyping was deemed necessary. Nine SNPs (with $10^{-10} < P < 5 \times 10^{-7}$) were considered to merit further genotyping, and the results from eight were deemed "inconclusive" ($P > 5 \times 10^{-7}$).

The nine SNPs selected for further genotyping were examined in Stage 2 in an additional N=63,374 individuals of European ancestry.

After Stage 2 genotyping had commenced (August 2009), GWAS data in a further N=9,041 + 23,294 (WGHS) individuals became available and were added to our study for all SNPs genotyped in Stage 1, as well as all SNPs which had attained $P < 2.5 \times 10^{-8}$ in Stage 0. In Stage 3 (May 2010 for receipt of WGHS data for all SNPs), *P*-values were calculated for a joint analysis of all data (Stage 0, Stage 1 if available, Stage 2 if available, and Stage 3).

Noting that our rules for "early stopping" of the follow-up experiment were $P < 2.5 \times 10^{-8}$ at the end of stage 0, and $P < 10^{-10}$ at the end of stage 1, we determined that when all available data were meta-analysed for each SNP at the end of stage 3, a significance threshold $P < 2.5 \times 10^{-8}$ would conservatively ensure a false positive rate less than 5×10^{-8} per SNP per phenotype, after accounting for interim decisions in the sequential design.

To estimate the realised false positive rate for our sequential design, we calculated the per-SNP per-phenotype false positive rate contributions from each of the four mutually-exclusive possible declarations of significance. Under standard assumptions of test statistic normality,

independence of data collected at different stages, and assuming sample sizes (proportional to) 70k, 40k, 40k and 25k for stages 0, 1, 2 and 3, these are as follows:

$$\text{Pr } [P_0 < 2.5\text{x}10^{-8} \text{ and } P_{03} < 2.5\text{x}10^{-8}] = 2.83\text{x}10^{-9}$$
$$\text{Pr } [P_0 > 2.5\text{x}10^{-8} \text{ and } P_{01} < 10^{-10} \text{ and } P_{013} < 2.5\text{x}10^{-8}] = 5.53\text{x}10^{-11}$$
$$\text{Pr } [P_0 < 2.5\text{x}10^{-8} \text{ and } 5\text{x}10^{-7} < P_{01} < 10^{-10} \text{ and } P_{013} < 2.5\text{x}10^{-8}] = 9.12\text{x}10^{-9}$$
$$\text{Pr } [P_0 < 2.5\text{x}10^{-8} \text{ and } P_{01} > 5\text{x}10^{-7} \text{ and } P_{013} < 2.5\text{x}10^{-8}] = 1.52\text{x}10^{-8}$$

Here, P_0, P_{01}, P_{03}, and P_{013} respectively denote the P-value combining data from Stage 0 only, Stages 0 and 1, Stages 0 and 3, and Stages 0, 1 and 3. Therefore, the total realised false positive rate (FPR) is $2.72\text{x}10^{-8}$ per SNP per phenotype. As we tested SBP and DBP, which are correlated phenotypes, this approach corresponds to an expectation, which is bounded above (conservatively) by $5.44 \text{ x}10^{-8}$. To illustrate the importance of taking into account the interim decisions, we calculate that simply declaring significance at $P_{0123} < 5\text{x}10^{-8}$ for all SNPs at the end of stage 3, keeping all other assumptions the same, the realised FPR would have been $5.33\text{x}10^{-8}$ per phenotype.

To determine whether the previously reported association[55] at the *PLCD3* locus was supported by all currently available data, we reasoned as follows:

(i) In the meta-analysis of all available data, the SNP rs12946454 at this locus did not reach our pre-specified significance threshold accounting for our sequential design (final $P_{03} = 4.35\text{x}10^{-8}$ for SBP; see Supplementary Table 4). It is not appropriate to compare this against a classical $P < 5\text{x}10^{-8}$ threshold because this is one of the SNPs for which follow-up genotyping was not carried out because of the small P-value at the end of stage 0.

(ii) Restricting the analysis to data newly acquired for this study (and previously not reported), there is no support for association of rs12946454 ($P = 0.02$ for SBP and $P = 0.15$ for DBP in ~48,000 individuals, which are not significant after adjusting for the 36 loci followed up).

5. Expression SNP-blood pressure analysis

For each of the 29 sentinel genome-wide significant blood pressure SNPs (Table 1 and Supplementary Table 5), all proxy SNPs with $r^2 > 0.8$ were identified in HapMap CEU (releases 21, 22, and HapMap 3 vers. 2) using SNAP[92]. All index SNPs and their proxies were then looked up in a database of expression SNP (eSNP) results from the following tissues: fresh lymphocytes[93], fresh leukocytes[94], leukocyte samples in individuals with celiac disease[95], lymphoblastoid cell lines (LCL) derived from asthmatic children[96], HapMap LCL from 3 populations[97], a separate study on HapMap CEU LCL[98], peripheral blood monocytes[99], adipose[117] and blood samples[100,101], 2 studies on brain cortex[99,102], 3 large studies of brain regions including prefrontal cortex, visual cortex, and cerebellum (Emilsson, personal communication), liver[103], osteoblasts[104], and additional fibroblast, T cell and LCL samples[105]. The collected eSNP results met criteria for statistical significance for association with gene transcript levels as described in the original papers. The index BP SNP (Supplementary Table 6A) or the best proxy SNP (Supplementary Table 6B) that is also an eSNP, and each of its eSNP associations, are reported, along with the strongest known eSNP (regardless of BP association) for each transcript in tissues showing BP association at that eSNP locus.

A selection of transcripts from some previously identified BP GWAS loci were measured by reverse transcriptase-polymerase chain reaction (RT-PCR) in RNA from leukocytes and platelets from 1,800 Framingham Heart Study offspring cohort (examination cycle 8) participants as previously described[106]. A targeted set of SNPs was genotyped by a custom

Illumina iSelect panel that included BP index SNPs and previously reported eSNPs at the RT-PCR loci)[55,91,107].

In examining cis-expression associations, either based on genome-wide expression surveys or RT-PCR and regional SNP analysis, we define coincident associations as those where the strongest BP SNP at a locus was also the strongest known eSNP for at least 1 transcript at the locus. We highlight 5 such loci in Table 6A. There is a potential for confounding by linkage disequilibrium, such that SNP associations detected through GWAS (whether to transcript level or BP) are more likely to occur for SNPs in LD with many SNPs. One cannot exclude the possibility that a given BP index SNP is associated also with levels of a given transcript, but that there is no causal relationship between the transcript level and BP.

6. Non-synonymous SNP look ups

We searched for all non-synonymous SNPs that were in high LD ($r^2 > 0.8$ in 1000 Genomes Low Coverage Pilot CEU phased haplotypes, available from ftp://ftp.1000genomes.ebi.ac.uk) with one of the 29 index genome-wide significant blood pressure SNPs (Table 1 and Supplementary Table 5). Results are presented in Supplementary Table 7.

7. Metabolomic and lipidomic analyses

To identify candidate mechanisms that might mediate associations with BP, the 29 BP-associated SNPs were tested for association with levels of circulating serum metabolites. For this purpose, results from three NMR-based metabolomic studies were meta-analyzed with total sample size N=7,032, and (separately) results from five mass spectrometry-based lipidomic studies were meta-analyzed with total sample size N=4,023.

The three metabolomics studies comprised 4,703 individuals from the Northern Finland Birth Cohort 1966 (NFBC1966, see Section 1.59) genotyped using the Illumina CNV370-Duo DNA Analysis BeadChip, 1,904 individuals from the Cardiovascular Risk in Young Finns Study (YFS, see Section 1.81) genotyped using a customized Illumina 670 BeadArray[108], and 425 individuals from the Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study, consisting of unrelated Finnish individuals, aged 25–74 years from the Helsinki area[109], recruited in 2007 as an extension of the FINRISK 2007 study, genotyped using the Illumina 610-Quad SNP array (Illumina Inc., San Diego, CA, USA), using standard protocols. Each of these studies measured levels of circulating metabolites in fasting serum samples, all analysed using the same high throughput proton NMR platform. This methodology provided information on 138 serum measures including lipoprotein subclass distribution and lipoprotein particle concentrations, low-molecular-weight metabolites such as amino acids, 3-hydroxybutyrate, and creatinine, and detailed molecular information on serum lipids including free and esterified cholesterol, sphingomyelin, (poly)(un)saturated and ω-3 fatty acids. Further details of the NMR spectroscopy and metabolite quantification have been described previously[101,110]. All three studies imputed genotypes at unmeasured SNPs using standard hidden Markov model methods and HapMap CEU reference panels. Individuals missing BP or on lipid-lowering medication and pregnant women were excluded from the analyses. The analyses were adjusted for sex, BMI, age (in YFS and DILGOM) and for the first ten principal components from the genome-wide data to account for population structure.

The residuals were transformed by inverse normal transformation to normal distributions for each cohort separately and tested for associations with the 29 BP-associated SNPs. Effect size estimates from the three cohorts were combined using inverse variance weighted meta-analysis using the R statistical package.

The five lipidomics studies were conducted in five genetically diverse European populations: the Erasmus Rucphen Family (ERF) study (see Section 1.25), the MICROS study (see Section 1.55), the Northern Swedish Population Health Survey (NSPHS; see Section 1.62), the Orkney Complex Disease Study (ORCADES; see Section 1.63) and the CROAS (VIS) study (see Section 1.78). For all five studies, participants were not selected on the presence of lipid-related or other pathology. A total of 4,023 individuals had both genotype data and lipidomic measurements available. The lipidomic measurements were all conducted at the Institute of Clinical Chemistry and Laboratory Medicine of Regensburg University, Germany, using electrospray ionization tandem mass spectrometry (ESIMS/MS), as validated and described previously[111,112]. In brief, samples were analysed by direct flow injection using a precursor ion scan of m/z 184 specific for phosphocholine containing lipids including phosphatidylcholine (PC), sphingomyelin (SM)[112] and lysophosphatidylcholine (LPC)[111]. A neutral loss scan of m/z 141 was used for phosphatidylethanolamine (PE) [113] and PE-based plasmalogens (PE-pl) were analysed according to the principles described by Zemski-Berry[114]. Fragment ions of m/z 364, 380 and 382 were used for PE p16:0, p18:1 and p18:0 species, respectively. Quantification was achieved by calibration lines generated by addition of naturally occurring lipid species to plasma and internal standards belonging to the same lipid class (PC14:0/14:0, PC 22:0/22:0, PE 14:0/14:0, PE 20:0/20:0, LPC 13:0, LPC 19:0). Calibration lines were generated for the following naturally occurring species: PC 34:1, 36:2, 38:4, 40:0 and PC O 16:0/20:4; LPC 16:0, 18:1, 18:0; PE 34:1, 36:2, 38:4, 40:6 and PE p16:0/20:4. Correction of isotopic overlap of lipid species as well as data analysis was performed by self-programmed Excel macros for all lipid classes according to the principles described previously[112]. The performed analysis does not always allow an exact assignment. In this case, an "O" is added to the subspecies name, e.g., PC O 36:5 and PC O 32:1. This denotes that the two species are most likely be assigned to PC species containing an ether bond (alkyl) and may constitute plasmalogens. However, we cannot exclude the possibility that PC O 36:5 may be assigned to PC 35:5, an unlikely odd carbon number species. Similarly, PC O 32:1 may be assigned to PC31:1. Persons with measurements more extreme than +/-4 standard deviations from the mean were excluded from further analysis. Measurements were first adjusted for age and sex, and residuals that were not normally distributed were inverse normal transformed prior to association analysis.

Genotyping platforms and imputation methods were identical to those used for the BP association analyses (see Supplementary Table 2). As all of the studies included related individuals, testing for association between (residual) lipid level and allele dosage was performed using a mixed model by 'mmscore' option in GenABEL software[115]. This option combines the Family Based Score Test for Association (FASTA) method of Abecasis et al.[116] and kinship matrix estimated from genotyped SNPs[117]. Effect size estimates from the five studies were meta-analyzed using inverse variance weighting using METAL software[118]. Multiple testing-corrected significance levels were calculated using a Bonferroni correction for 4,031 tests (139 metabolites times 29 SNPs).

Testing the 29 BP index SNPs for association with these panels of metabolites failed to identify any metabolites that might mediate causal effects, after Bonferroni correction for multiple testing for the 138 and 139 metabolites examined in each dataset respectively (see Supplementary Table 9).

8. CNV analyses

We obtained lists of CNV tagging SNPs (CNV-tSNPS) from four sources: (i) 261 CNV-tSNPs, which tag CNVs at $r^2 > 0.8$, generated at the Broad Institute by typing HapMap samples on the Affymetrix 6.0 array[119]. (ii) 2,174 CNV-tSNPs, which tag CNVs at $r^2 > 0.8$, made available by the Genomic Structural Variation consortium and based largely on typing 450 HapMap samples on a custom-made Agilent 105K array capable of genotyping ~3,320 CNVs in HapMap CEU[120]. (iii) 3,113 CNV-tSNPs, generated using HapMap phase III samples genotyped using Affymetrix 6.0 and Illumina 1M arrays, generated by the HapMap 3 project (http://hapmap.ncbi.nlm.nih.gov/). (iv) 2,905 CNV-tSNPs, generated using ~19,000 samples of European ancestry genotyped by the Wellcome Trust Case Control Consortium (3,000 controls and 2,000 cases for each of 8 diseases)[121], using a custom-made Agilent 105K array [120]. By combining these lists (without excluding multiple CNV-tSNPs that tag the same CNV), we identified 6,378 CNV-tSNPs for which we had meta-analysis association statistics for SBP and DBP. For these SNPs, we applied a simple Bonferroni corrected threshold of $P < 0.05/6378$, after genomic control was applied genome-wide to all meta-analysis association P-values.

Only CNV tSNP rs6599167 was significantly associated with DBP ($P=9.3 \times 10^{-7}$, $r^2=0.77$ with our index SNP rs3774372 at the *ULK4* locus). The tagged CNV is rare compared to the tagging SNP (rs6599167 MAF=0.18, CNV MAF=0.018) and is poorly tagged by any HapMap SNP (maximum $r^2=0.090$). Further exploration would require directly genotyping the CNV and testing for association with DBP in large sample sizes.

9. MAGENTA analyses

To test whether the overall genome-wide association results were enriched for members of specific biological pathways, we used the gene set enrichment framework MAGENTA v2.1[122] to calculate a corrected gene association P-value based on the most significant SNP association P-value of all SNPs in the gene region (defined as 110 kb upstream to 40 kb downstream from transcript start/stop). MAGENTA corrects values for gene size, number of SNPs/gene, and recombination. Results were tested against databases within MAGENTA v2.1 including: KEGG (from June 2008), PANTHER Biological Processes (from Jan. 2010), PANTHER Molecular Function (from Jan. 2010), and Ingenuity Pathway (from June 2008). For each pathway, enrichment of highly ranked gene scores above the 95th percentile of all gene scores in the meta-analyses was evaluated compared to 10,000 randomly sampled gene sets of identical size from the genome.

Analysis was performed for the meta-analysis results for DBP and SBP. Additionally, to examine whether common factors might be present across the two analyses, an analysis was performed using the minimum p-value for each SNP of the DBP and SBP results. Supplementary Table 10 shows the results for pathways with a nominal P-value less than 0.01. This highlights some biological categories which could fit with processes linked with blood pressure. However, given the number of categories tested in each database, the nominal GSEA P-values might best be compared against the following Bonferroni corrected P-values to control the false positive rate for each individual database: Ingenuity pathways $P < 0.0005$, KEGG pathways $P < 0.0003$, PANTHER BioProc $P < 0.0002$, PANTHER MolFunc $P < 0.0002$, PANTHER pathways $P < 0.0004$). A further 15-fold multiple testing correction could additionally be applied, because three sets of GWAS results were used (DBP, SBP, best of DBP and SBP) and five databases were used.

10. Risk score analyses using multi-SNP predictors

We defined genetic risk scores in the following way: Using a set of m SNPs, for the i-th SNP in the j-th individual denote x_{ij} as the 0/1/2 coded genotype (for directly genotyped SNPs) or expected allele dosage (which takes real values between 0.0 and 2.0 for imputed SNPs). Using results from Stages 0-3, define the set of regression coefficients to be $w_1, w_2, ..., w_m$. Then the risk score for subject j is defined to be

(1) $s_j = s_0 + w_1 x_{1j} + w_2 x_{2j} + ... + w_m x_{mj,}$

where s_0 is the intercept. In all our analyses, we specify the coefficients $w_1, w_2, ..., w_m$ to be the effect sizes, in mmHg per coded allele, estimated in single SNP analyses of both SBP and DBP in stages 0. For a risk score based on mean BP (defined as (SBP+DBP)/2), we used coefficients $(w_i^{(SBP)} + w_i^{(DBP)})/2$, where $w_i^{(SBP)}$ and $w_i^{(DBP)}$ are the observed effect of the i-th SNP on SBP and DBP, respectively (this can be motivated formally by appealing to the linearity of both addition and expectation).

We also note that, when considering multiple SNPs that are in linkage equilibrium with each other, and small effect sizes per SNP, effect sizes estimated jointly for all SNPs using a multiple regression model are effectively identical to those estimated in a series of single SNP regression models. Thus regression on the risk score can be reconstructed from regressions on each of the m SNPs in turn, without further access to individual-level data.

The calculations involved are of the same type as for meta-analysis; the coefficient of the risk score is a weighted mean of the per-SNP regression coefficients, where each is weighted by its corresponding w_i. The estimated variance of the risk score is given by similarly weighting the estimated variances (squared standard errors) of each per-SNP regression coefficient. The assumption of zero LD between SNPs ensures that these contributions are independent. Importantly, as with inverse-variance weighted meta-analysis, in large samples this procedure gives valid p-values under the null, i.e. when there is no relationship between the "lookup" phenotype and any variants at the SNPs contributing to the risk score.

Using SNP-specific results in this way, we estimated and tested the coefficient of the risk score in independent "lookup" results using linear regression for continuous phenotypes, logistic regression for binary phenotypes, and proportional hazards regression for time-to-event phenotypes. These estimates and tests inherit the covariate adjustment performed in the original SNP-specific analysis.

11. Consortia and studies providing association results for non-BP phenotypes and HTN

We obtained summary phenotype-genotype association results for up to 30 SNPs of interest, by requesting "look-ups" in the results of meta-analyses and analyses that had already been conducted by consortia and research groups studying non-BP phenotypes. We also obtained phenotype-genotype results for hypertension from the WGHS (see Section 1.79) and the BRIGHT study.

In this section, we summarise relevant information about cohort recruitment and ascertainment, methods of phenotype measurement, phenotype definitions for case/control phenotypes, and phenotypic covariates used in the association analyses. We do not present technical details, such as the choice of genotyping platform or genotype imputation methodology, or which extra covariates were used to control for population stratification, when these are already

documented in the published literature. Briefly, almost all genotyping was performed using standard GWAS arrays, genotype imputation was performed using hidden Markov model approaches with a HapMap CEU reference panel, and ancestry principal components[86] or location-of-origin were used as covariates and/or variance components were used to adjust for familial relatedness in the individual analyses.

11.1 BRItish Genetics of HyperTension study (BRIGHT)

We obtained association data for 28 SNPs directly genotyped using the KASPAR assay and one perfect proxy SNP genotyped using the HumanCVD BeadChip (Illumina) in the BRIGHT study which includes hypertensive cases selected from families with more than one affected sibling and normotensive controls[123]. There was no overlap with samples used for the ICBP-GWAS discovery analysis. Hypertensive cases had a diagnosis of hypertension prior to 50 years, and BP recordings $\geq$ 150/100mmHg on a single reading or $\geq$ 145/95mmHg on 3 consecutive readings. Normotensive controls had blood pressure recordings of SBP $\leq$ 140mmHg and DBP $\leq$ 90mmHg and were not taking any anti-hypertensive medications and were recruited from similar geographical regions as the cases. Blood pressure was measured in both cohorts using the Omron-705CP blood pressure monitor. After QC exclusions the sample analyzed comprised 2,406 hypertensive cases and 1,990 normotensive controls.

11.2 CHARGE - Heart Failure Working Group

We obtained association data for SNPs of interest from the meta-analysis of 4 cohorts with a total of 20,926 participants free of clinical heart failure at baseline, in whom 2,526 incident heart failure events occurred during follow-up[124]. All cohorts included in the heart failure analysis are also included in the ICBP-GWAS discovery analysis.

11.3 EchoGen (LM mass and LV weight)

Association statistics for left ventricular (LV) mass and LV wall thickness (LV mass) were obtained from the discovery meta-analysis described previously[125]. The discovery analysis for this study combined data from 5 cohorts with total sample size N=12,612. Four of the cohorts, CHS, RS, KORA F3, FHS, with total N=9,312, overlap those used for the ICBP-GWAS discovery analysis.

Subjects underwent routine transthoracic echocardiography, and methodology for measurement of LV dimensions, and calculation of mass and wall thickness, have been reported previously[125]. Within-cohort association analyses regressed LV mass and LV wall thickness onto additively coded (expected) genotype dose, with age, sex, height and weight as covariates, using linear regression (with random effects to account for relatedness where necessary). Results were combined across cohorts using an inverse variance weighted meta-analysis.

11.4 NEURO-CHARGE (stroke)

Association statistics for risk of incident stroke were obtained from the discovery meta-analysis of the CHARGE consortium, as described previously[126]. The discovery analysis for these phenotypes combined data from 4 cohorts with total sample size N=19,602, of which there is 100% overlap the ICBP-GWAS discovery samples.

For the stroke analysis, individuals who were stroke-free at recruitment were followed up for an average of 11 years, and there were 1,544 incident strokes (of which 1,164 were ischemic strokes). The association analysis was a survival (time-to-event) analysis using a proportional hazards model, adjusted for age and sex as covariates.

11.5 UK-US Stroke Collaborative Group

A total of 2,957 subjects were included in the study of an association with ischemic stroke (1,473 cases and 1,484 controls). Cases came from the BRAINS (N=435), SWISS (N=296), and ISGS (N=742) studies; the controls came from ISGS and Coriell (N=782) and the Baltimore Longitudinal Study of Aging (BLSA, N=702). Probands from SWISS were extracted from family groups.

The British Repository of DNA in Stroke (BRAINS) is an on-going multi-centre, hospital-based prospective study from the UK[127]. All recruits were extensively clinically phenotyped and have imaging-confirmed ischaemic stroke using either CT or MRI brain scans. BRAINS has a number of spokes including BRAINS-SA (South Asian) which recruits British Asian stroke patients and BRAINS-Euro recruiting stroke patients of European descent. For the purposes of this study samples were selected from BRAINS-Euro.

Siblings with Ischemic Stroke Study (SWISS) is a prospective multicenter affected sibling pair study of first-ever or recurrent ischemic stroke[128]. Subjects were recruited from 54 enrolling hospitals across the US and Canada. All recruits were extensively clinically phenotyped and have imaging-confirmed ischaemic stroke using either CT or MRI brain scans. Samples were collected between October 2000 and December 2009.

Ischemic Stroke Genetics Study (ISGS) is a 5-center prospective case-control study of first-ever ischemic stroke cases and concurrently enrolled controls individually matched for age, sex and recruitment site[129]. Samples were collected between May 2003 and September 2008.
No UK-US SCG stroke cases have overlap with samples used for the ICBP-GWAS discovery analysis.

Controls utilized in this study are ISGS controls (no overlap with ICBP-GWAS discovery analysis), neurologically normal controls from Coriell[130] (no overlap with ICBP-GWAS discovery analysis), and participants of the Baltimore Longitudinal Study of Aging (BLSA, which overlapped N=702 samples used for the ICBP-GWAS discovery analysis). The controls for the present analysis were restricted to individuals with no known first degree relatives with stroke or other neurological disease.

Cases were genotyped using Illumina 650K Quad arrays at the Department of Molecular Neuroscience and Reta Lilla Weston Laboratories, Institute of Neurology, University College London. Controls were genotyped using Illumina 550Kv1 or 550Kv3 arrays at the Laboratory of Neurogenetics, National Institute on Aging, NIH (Bethesda, MD). Individuals $> \pm 6$ SD from the mean European ancestry estimates for a combined CEU/TSI sample, after multi-dimensional scaling analyses for component vectors 1 and 2 were excluded. Cryptic relatedness was filtered using pair-wise proportional sharing estimates of < 0.15 for inclusion, effectively removing duplicates, first and second-degree relatives. Samples with call rate <0.95 and samples discordant between genotypic sex estimated from X chromosome heterogeneity and self-reported sex were excluded. Final sample size was 1,473 cases and 1,484 controls.

SNPs were filtered for call rate≥ 0.95, MAF > 0.01, HWE p-value $> 10^{-7}$ before imputation to remove likely genotyping errors. Imputation was conducted using MaCH 1.0.16 using phased HapMap CEU haplotypes from the 1000 genomes August 2009 release from Sanger Association analyses were conducted assuming a logistic regression model, adjusted for component vectors 1 and 2 from a multi-dimensional scaling analysis of the case/control samples with allele dosage at each SNP as a predictor of stroke (using MaCH v 1.0.16 for imputed SNPs and PLINK v 1.07 for genotyped SNPs). Genome-wide, association statistics had a genomic control lambda of 1.0268.

11.6 CARDIoGRAM (CAD)

The <u>C</u>oronary <u>AR</u>tery <u>DI</u>sease <u>G</u>enome-wide <u>R</u>eplication <u>A</u>nd <u>M</u>eta-analysis (CARDIoGRAM) consortium combines data from 14 GWAS, all published and several unpublished, in individuals with European ancestry including >22,233 cases with CAD and/or MI and >64,762 controls, and unifies samples from Atherosclerotic Disease VAscular functioN and genetiC Epidemiology study, CADomics, Cohorts for Heart and Aging Research in Genomic Epidemiology, deCODE, the German Myocardial Infarction Family Studies I, II, and III, Ludwigshafen Risk and Cardiovascular Heath Study/AtheroRemo, MedStar, Myocardial Infarction Genetics Consortium, Ottawa Heart Genomics Study, PennCath, and the Wellcome Trust Case Control Consortium (WTCCC). These studies have a case-control design or are prospective cohort studies both having detailed phenotyping for CAD and/or MI as previously described[131,132]. Control subjects have been derived from population-based studies in most investigations. 2,287 cases and 22,024 controls from the CHARGE consortium, 1,121 controls from MIGen, and 1,473 controls from B58C used by the WTCCC, overlap samples used in the ICBP-GWAS discovery analysis.

For all of the participating studies, genome-wide scans were performed in the years 2006-2009 using either Affymetrix or Illumina platforms followed by imputation of genotypes in most studies. Statistical methods have been standardized across the studies, and an analysis platform has been created to allow summarized analyses on CAD, MI, and related phenotypes. Data were combined across all studies (prevalent CAD case/control comparisons). Overall prevalence of MI within the CAD cases was 66%.

11.7 C4D Consortium (CAD)

The C4D consortium comprises CHD cases and controls of European origin from PROCARDIS and the Heart Protection Study (HPS) and of South Asian origin from the LOLIPOP and PROMIS studies[133]. Data analyzed with respect to risk of CHD all relate to the European origin participants from PROCARDIS and HPS. The PROCARDIS study included 5,720 cases of CHD (74.9% male, 80.4% myocardial infarction) and 1,684 controls recruited from Germany, Italy, Sweden and the UK. The controls were supplemented with N=2,697 unselected individuals provided from the National Blood Service and used as common controls by the WTCCC. Many of the PROCARDIS cases are related to each other, and familial clustering was accounted for by using robust sandwich estimation of the variance, using Stata v 10.

The HPS included 2,704 CHD cases included in the HPS trial[24] and 2,887 controls from the 1958 Birth Cohort. The HPS was a large clinical trial of patients with either (i) CAD (i.e., MI, unstable or stable angina, coronary artery bypass grafting, or angioplasty); or (ii) occlusive disease of non-coronary arteries (i.e., stroke, leg artery stenosis; or (iii) diabetes mellitus; or (iv) treated hypertension (if also male and aged 65 or older). A total of 20,536 individuals (15,454 men and 5,082 women) were recruited. Previous MI was reported by 8,510, some other history of CAD by 4,876, and no history of CAD by 7,150. Among the 13,386 with known CAD, 1,460 had cerebrovascular disease, 4,047 had peripheral arterial disease, and 1,981 had diabetes mellitus. GWAS genotyping was performed for a random sample of 4,000 participants with sufficient DNA and complete phenotypic information, and of these there were 2,704 CAD cases after quality control exclusions.

No CAD cases used by C4D overlapped samples used for the ICBP-GWAS discovery analysis, but N=795 controls from PROCARDIS did overlap samples used for the ICBP-GWAS discovery analysis.

11.8 CKDGen

Association statistics for estimated glomerular filtration rate (eGFR) were obtained from the discovery meta-analysis of the CKDGen consortium, as described previously[134]. eGFR was calculated from calibrated creatinine using the 4-variable MDRD Study equation. The discovery analysis for these phenotypes combined data from 20 cohorts with total sample size N=67,093. 14 of the cohorts: AGES, Amish, ARIC, BLSA, CHS, 1,300 samples from ERF, FHS, KORA F3, MICROS, ORCADES, RS, RSII, SHIP and Vis, with total N=39,361, overlap the cohorts used for the ICBP-GWAS discovery analysis. Association statistics for dichotomous chronic kidney disease (CKD), urinary albumin/creatinine ratio (UACR), and dichotomous microalbuminuria, were obtained in collaboration with the CKDGen consortium by querying their datasets. CKD was defined as eGFR <60 ml/min/1.73m^2, in the same set of samples in which eGFR was studied. The discovery analysis for urinary albumin/creatinine phenotypes combined data from 12 cohorts with total sample size N=31,580. 12 of the cohorts, Amish, ARIC, BLSA, CHS, CoLaus, EPIC, Fenland, FHS, KORA F3, MICROS, NSPHS, and SHIP, with total N=30,342, overlap cohorts used for the ICBP-GWAS discovery analysis. Microalbuminuria was defined as UACR >25 mg/g [women] or >17 mg/g [men]. For CKD there were 5,807 cases and 61,286 controls available. For microalbuminuria 3,698 cases and 27,882 controls were used. eGFR and urinary albumin/creatinine ratio were (natural) log transformed prior to analysis. Within-cohort association analyses regressed phenotype on genotype using an additive genetic model with age and sex as covariates, using linear regression for continuous phenotypes and logistic regression for dichotomous phenotypes. Family-based methods were used where relevant. Results were combined across cohorts using an inverse variance weighted meta-analysis.

11.9 KidneyGen consortium

Association statistics for serum creatinine were obtained from the discovery meta-analysis of the KidneyGen consortium, as described previously[135]. The discovery analysis for this study combined data from 9 cohorts, with total sample size N=23,812. Six cohorts, CoLaus, SardiNIA, 873 samples from TwinsUK, Fenland, InCHIANTI, NFBC1966, with total sample size N=17,699, overlap the cohorts used for the ICBP-GWAS discovery analysis.

Serum creatinine concentrations were log10 transformed prior to analysis. Within-cohort association analyses were linear regressions of transformed serum creatinine on genotype, using an additive genetic model with age, sex and ancestry principal components as covariates. Results were combined across cohorts using a standard inverse variance weighted meta-analysis. Effect sizes were converted to a natural log transformed scale for presentation in Table 2 and Suppl. Table 14, for comparability with other phenotypes.

11.10 Heart and Vascular Health Study

The setting for this study was Group Health (GH), a large integrated health care system in western Washington State. Data were utilized from an ongoing case-control study of incident myocardial infarction (MI) and stroke cases with a shared common control group. Methods for the study have been described previously[136-138] and are briefly summarized below. All study participants were GH members and aged 30-79 years. MI and stroke cases were identified from hospital discharge diagnosis codes and were validated by medical record review. Controls were a random sample of GH members frequency matched to MI cases on age (within decade), sex, treated hypertension, and calendar year of identification. The index date for controls was a computer-generated random date within the calendar year for which they had been selected. For MI cases, the index date was the date of admission for the first acute MI. Participants were excluded if they were recent enrollees at GHC, had a history of prior MI or stroke, or if the

incident event was a complication of a procedure or surgery. Eligibility and risk factor information were collected by trained medical record abstractors from a review of the GH medical record using only data available prior to the index date and through a telephone interview. Medication use was ascertained using computerized GH pharmacy records. There is no overlap between these samples and those used for the ICBP-GWAS discovery analysis.

Genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system. Genotypes were called using the Illumina BeadStudio software. Samples were excluded from analysis for sex mismatch or call rate < 95%. The following exclusions were applied to identify a final set of 301,321 autosomal SNPs: call rate < 97%, HWE $P < 10^{-5}$, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap, inconsistencies across genotyping batches. Imputation was performed using BIMBAM with reference to HapMap CEU using release 22, build 36 using one round of imputations and the default expectation-maximization warm-ups and runs. Logistic regression was used to assess the association of each SNP with stroke and MI, adjusting for the matching factors of age, sex, hypertension status and index year. We used an additive model for genotype, robust standard errors, and estimated risk for each additional copy of the variant allele. All analyses were implemented in R. SNPs with variance on the allele dosage ≤0.01 were excluded from analysis. There were 1,172 MI cases and 1,314 controls; the stroke analysis included 501 ischemic stroke cases and 1,314 controls. The genomic control lambda was 1.05 for MI and 1.07 for stroke.

11.11 Women's Genome Health Study (WGHS)
Samples, phenotyping, and genotyping are described above in Section 1.79. There is no overlap with the samples used in the ICBP-GWAS discovery analysis[139].

12. Analyses of SNPs not reaching genome-wide significance

We tested for evidence of association with SBP and DBP for sets of SNPs that did not reach the significance thresholds required for validation genotyping in the present study. To achieve this, we constructed cohort-wise partitions of our principal GWAS dataset (with N=69,899 individuals), to make pairs of non-overlapping discovery and validation datasets. We constructed these partitions such that the discovery dataset constituted ~50%, ~60%, ~70% and ~80% of the sample size while the validation dataset constituted the remainder of the sample size. For each target partition size, we used a Monte Carlo algorithm to identify 50 distinct cohort-wise partitions that were as close as possible to the target partition proportions. For each partition we performed separate meta-analyses within the discovery and validation datasets, using identical methodology to our principal meta-analysis (Section 3).

SNPs were ranked by association P-values in each discovery meta-analysis and aggressive LD-based pruning was used to remove all SNPs in pair-wise LD $r^2 > 0.05$ with any SNP with a smaller P-value. This procedure defined sets of effectively independent SNPs with discovery P-values in predefined intervals: $P \leq 10^{-7}$, $10^{-7} < P \leq 10^{-6}$, ... , $10^{-3} < P \leq 10^{-2}$, such that SNPs included in any given discovery P-value interval are not included if they are in linkage disequilibrium with SNPs with more significant discovery P-values. For each set of SNPs we defined a linear predictor using coefficients that were effect size estimates from the discovery dataset, and then regressed the phenotype onto this predictor in the non-overlapping testing dataset, using the method described in Section 10.

As an empirical negative control for our approach, we defined phenotypic predictors with SNPs and coefficients determined from analyses of alternative phenotypes not directly related to blood pressure. In place of meta-analysing discovery sets from ICBP-GWAS, we used the meta-analysis results of Kathiresan et al.[140] for each of three lipid phenotypes. We used only the validation sets from ICBP-GWAS that contained independent samples to the ones used by Kathiresan et al.[140] (which for 20% of the total ICBP-GWAS sample size was 8/50 distinct partitions). We found that the lipid phenotype-derived predictors explained relatively little of the variation (<1%) in blood pressure, and no predictors were significantly associated after accounting for multiple testing (largest 2 ln L = 6.09, p = 0.24 adjusting for 18 tests).

In contrast, BP-derived predictors based on SNPs within all *P*-value bins were significantly associated with BP in the independent testing subsamples ($P=6\times10^{-5}$ to $P=5\times10^{-10}$ for 12 independent 1 d.f. tests), as shown in Supplementary Figure 5. This provides robust evidence for the existence of as yet undetected common variants among these SNPs.

13. Estimating the number of as-yet undiscovered signals

Our approach to estimating the total number of independent signals associated with continuous blood pressure phenotypes, and the total fraction of variance explained, is based on the approach described recently by Park et al.[141]. The underlying principle is as follows: For a set of signals detected in a discovery GWAS, data independent of the data used for discovery is free of bias caused by the winner's curse effect, and therefore can be used to consistently estimate the true effect sizes. This allows consistent estimates of the power of the discovery GWAS to detect each of the signals that was detected.

For details on these analyses, see Appendix A.

14. Measurement of natriuretic peptides in Framingham and Finrisk97

Measurements of brain natriuretic peptide (BNP) and N-terminal pro-atrial natriuretic peptide (pro-ANP) were obtained on Framingham Heart Study offspring cohort participants attending their sixth examination (1995-1998). Natriuretic peptide levels were measured with a high-sensitivity, non-competitive immunoradiometric assay, based on a 2-site sandwich antibody system (Shionogi Co, Osaka, Japan). The lower limit of detection was 4 pg/mL for BNP and 94 pmol/L for N-terminal pro-ANP. The sample size with BNP levels and genotype data was 2,891 individuals.

In the Finrisk97 sample, plasma BNP concentration was determined using the Abbott Architect BNP assay, with an inter-assay coefficient of variation of 4.3%. Plasma N-terminal pro-ANP was determined using an immunoluminometric sandwich assay targeted against the mid-region of the peptide (BRAHMS, AG, Berlin, Germany), with an inter-assay coefficient of variation of 2.3%. The N-terminal pro-BNP was determined using the Roche Elecys assay, with an inter-assay coefficient of variation of 1.4%. A total of 6,756 individuals were available with genotype and all natriuretic peptide measurements.

The G allele of the index SNP in the current report at the *MTHFR-NPPB* locus was associated with lower SBP and DBP (Table 1), and with higher ANP levels (FHS *P*>0.05, FR97

P=7.9x10^{-46}), BNP levels (FHS P=3.4x10^{-9}, FR97 P=1.7x10^{-19}) and N-terminal pro-BNP (FHS P=1.8x10^{-10}, FR97 P=6.6x10^{-30}). Opposite directional associations with BP and ANP/BNP levels are expected, and are consistent with ANP/BNP levels mediating the association with BP as a consequence of the vasodilatory and natriuretic properties of the peptide.

15. Websites accessed

Name	Website	Date accessed
1000 Genomes Project Data Site	ftp://ftp.1000genomes.ebi.ac.uk	July 25th, 2010
NHGRI GWAS catalogue	http://www.genome.gov/gwastudies/	Last on November10th, 2010
PubMed	http://www.ncbi.nlm.nih.gov/pubmed	Last on April 28th, 2011
SNAP annotation and proxy tool	http://www.broadinstitute.org/mpg/snap/ldsearchpw.php	Last on November10th, 2010
UCSC Genome Browser	http://www.genome.ucsc.edu	Last on November10th, 2010

16. Supplementary Tables (see also file "ICBPresub4_SoMtables_final.xls")

Supplementary Tables 1 through 14 are presented in a separate file: "ICBPresub4_SoMtables_final.xls". (Please see separate Supplementary Tables file).

Supplementary Table 1: Demographic data of European-ancestry cohorts

Demographic data on all European ancestry GWAS (stage 0) and follow-up (direct genotyping – stage 1 & 2 and lookup – stage 3) cohorts. For each cohort the alias is indicated in column 1 (see Supplementary Material for details on each study). Numbers of samples genotyped (N), basic descriptive statistics include mean (SD) for: age, systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI); % HTN = % of individuals defined as hypertensive (SBP ≥ 140mmHg, or DBP ≥ 90mmHg, or taking anti-hypertensive or blood pressure lowering medication for any reason); % anti-HTN treat. = % of individuals taking anti-hypertensive medications. Standard deviations of the residual computed by regression of treatment corrected BP on age, age^2, sex, and BMI and used in the association are given in column 10 and 11. Columns 13 and 14 show information used by the cohorts for adjustment of population stratification and sample relatedness (kinship). Abbreviations; NA= non-

applicable, PC and PCA= Principal Components and principal components analyses; IBS = identity by state.

Supplementary Table 2: Genotyping and imputation methods

Basic information on genotyping and imputation methods for each genome-wide study. Column 2 indicates if the study participated in the genome-wide scan (GWAS) or only replication SNPs were searched (lookup). The commercial genotyping platform and the genotype calling algorithm are stated in columns 3 and 4. The different studies used similar, but slightly different criteria to filter the genotyping data before the imputation. Information on the imputation is provided (software used, HapMap genotype freeze on which the imputation was based, and number of SNPs used). The last 6 columns indicate which software was used for association analysis and indicates the genomic control parameter for each phenotype used.

HWE=Hardy-Weinberg equilibrium; MAF = minor allele frequency, NA = not available, imput. = imputation; ME = Mendelian error.

Supplementary Table 3: GWAS results pruned by LD

All SNPs having passed quality control with an association P-value of $\leq 9.85 \times 10^{-6}$ for at least one of SBP or DBP were pruned to obtain a SNP list with independent SNPs (LD $r^2 < 0.2$). For genotyping and other purposes, a proxy SNP ("target SNP") rather than the SNP with the best association *P*-value ("sentinel SNP") was chosen for follow-up in some cases. If the sentinel SNP is in a gene, the alias name is indicated in column 6, together with all genes within a 100kb window (column 7). Position is given in hg18 coordinates. The distance from the closest SNP with lower *P*-value ("higher ranking SNP") is indicated in column 8. If a proxy SNP for the sentinel SNP was chosen for follow-up, the target SNP identity is indicated in column 9 with the reason for choice indicated. The 3 last columns of the table indicate the number of phases of replication for the target SNP if the replication was successful and the locus name. *P*-values indicated are corrected by genomic control within and between studies (see methods above). NA = not applicable.

Supplementary Table 4: Staged validation results

The table gives association statistics and sample sizes by stage and SNP. Because all SNPs genotyped in stage 2 progressed into stage 3 (Supplementary Figure 1), no interim analysis was conducted at completion of stage 2.

∗ Index SNPs reaching our positive criteria for skipping further targeted genotyping (green boxes in Supplementary Figure 1), which were $P_0 \leq 2.5 \times 10^{-8}$ and previously published for stage 0, and $P_{01} \leq 10^{-10}$ for stage 1; + Index SNPs reaching our negative criterion for skipping further targeted genotyping (red box in Supplementary Figure 1), which was $P_{01} > 5 \times 10^{-7}$ for stage 1; AAF means All Assays Failed for targeted single SNP (non-array) genotyping; ∗∗ Index SNPs reaching our criterion for declaring genome wide significance, which was $P_{0123} \leq 2.5 \times 10^{-8}$.

Supplementary Table 5: Summary of newly discovered or previously reported loci

Summary association statistics for all 28 newly discovered and previously reported loci[55,91]. Although the overall *P*-value for rs12946454 (*PLCD3*) achieves P $\leq$ 5x10^{-8}, the data summarized here were acquired via a sequential experiment that technically invalidates classical thresholds for genome wide significance. Comparison against a more appropriate threshold (P $\leq$ 2.5x10^{-8}; see Section 4) and consideration of the fact that the overall *P*-value is less significant with the inclusion of newly acquired data in N=30,174 individuals, suggest that this association should not be considered to be well validated. Genomic positions are given in build 36 coordinates. Abbreviations: Cod. all.=coded allele.

Supplementary Table 6: eSNP evidence

All BP index SNPs and their proxies were searched against eSNP sources as described in the Supplementary Materials (Section 5). Only the eSNP with the lowest BP *P*-value is shown here. Supplementary Table 6A: 9 BP index SNPs that are eSNPs based on genome-wide eQTL studies. Supplementary Table 6B: 3 eSNPs from genome-wide surveys that are in high LD with a proxy of BP index SNPs. Supplementary Table 6C: BP index SNPs that show evidence as being eSNPs using RT-PCR measurement of selected transcripts at several associated loci. †Proxy SNPs with r^2>0.8 to the sentinel SNPs were identified in HapMap CEU (releases 21, 22, and HapMap 3 vers. 2) using SNAP[92]. ‡ The lowest blood pressure *P*-value for the eSNP for either SBP or DBP after genomic control adjustment is listed. * A BP-eSNP relationship for this locus was previously reported[91]. n/a means not available. ¥ Measured by RT-PCR in a separate set of experiments[106].

Supplementary Table 7: Non-synonymous SNP lookups

For all 29 genome-wide significant loci, we determined whether either the index SNP, or a proxy with r^2>0.8, estimated in 1000 Genomes haplotype data for the CEU panel, resulted in non-synonymous changes in amino acids within the translated protein sequence for corresponding genes.

Supplementary Table 8: Sex- and BMI-interaction results

For the 29 significant SNPs from the main effect analysis, a sex-by-SNP and BMI-by-SNP interaction analysis was performed. Datasets with unrelated individuals were analyzed in strata by sex, with the sex*SNP interaction beta and SE estimated from within the strata, coding male = 1 and female = 2 (beta$_{interaction}$=beta$_{male}$ - beta$_{female}$; SE$_{interaction}$=$\sqrt{(SE_{male}^2+(SE_{female}^2))}$. The analyses were analogous for BMI, coding BMI< 25 = 0 and BMI $\geq$ 25 = 1. For studies that included related individuals the following regression was performed and the interaction term extracted (e.g. SBP ~ sex*W + age*W + (age^2)*W + ... + G + W + G:W; e.g. DBP ~ BMI*S + age*S + (age^2)*S + ... + G + S + G:S).

Supplementary Table 9: Results of the metabolomic/lipidomic lookups

Association of BP-associated SNPs with metabolic levels in a meta-analysis of the NFBC1966/YFS/DILGOM studies (upper section) and a meta-analysis of five EUROSPAN studies (lower section). Results are shown only for SNP-metabolite combinations with multiple

testing adjusted association $P < 1$, along with short and full names of the metabolite, effect size and standard error.

Supplementary Table 10: MAGENTA analyses

Genes near the top 29 SNPs of three phenotypes (SBP, DBP, and the minimum of SBP and DBP) were tested for the over-representation with genes from known pathways using three gene-set databases. The most significant biological pathways and genes are shown for each phenotype, the nominal P-value and the P-value after correction for false discovery are indicated in columns 6 and 7.

Supplementary Table 11: Demographic data of non-European ancestry cohorts

Demographic data on all non-European ancestry cohorts. For each cohort the alias is indicated in column 1 (see Supplementary Material for details on each study). Numbers of samples genotyped (N), basic descriptive statistics include mean (SD) for: age, systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI); % HTN = % of individuals defined as hypertensive (SBP $\geq$ 140mmHg, or DBP $\geq$ 90mmHg, or taking anti-hypertensive or blood pressure lowering medication for any reason); % anti-HTN treat. = % of individuals taking anti-hypertensive medications. Standard deviations of the residual computed by regression of treatment corrected BP on age, age^2, sex, and BMI and used in the association are given in column 10 and 11. Columns 13 and 14 show information used by the cohorts for adjustment of population stratification and sample relatedness (kinship). Abbreviations; NA= non-applicable, PC = Principal Components; IBS = identity by state.

Supplementary Table 12: Association analyses in samples of non-European ancestry, by SNP

Association of 29 index SNPs in individuals of non-European ancestry. These data are partly presented in Table 1 of the main text. Unadjusted association P-values are reported, with accompanying adjusted P values (Q-values) to correct for multiple testing over all SNPs and phenotypes within each ancestry. See Supplementary Table 13 for risk score analyses by individual cohort. The 29 SNPs were also tested for association with SBP and DBP in 2,030 individuals of Hispanic origin, but no significant associations were observed (results not presented). Abbreviations: EA = East Asian ancestry, SA = South Asian ancestry, AA = African and African-American ancestry.

Supplementary Table 13: Association analyses in samples of non-European ancestry, risk score analysis by cohort

For each cohort, we used all available SNPs passing QC (given in the # SNPs column) to test association between BP (either DBP or SBP) and a phenotype specific risk score, parameterized using estimates from the stage 0 GWAS meta-analysis in Europeans (Supplementary Figure 6). In columns SBP.a and DBP.a, the value 'a' is the estimated effect (in mmHg) in the target cohort predicted by the risk score, with associated standard errors in columns SBP.SE(a) and DBP.SE(a). We combined estimates of 'a' over cohorts using fixed effects inverse variance weighted meta-analyses.

Supplementary Table 14: Details on other outcome lookups.

Combined risk score for 29 genome wide significant SNPs tested for association with blood pressure and hypertension in an independent sample, with dichotomous cardiovascular outcomes of stroke, coronary artery disease (CAD), chronic kidney disease, and with continuous measures of hypertensive target organ damage. "Beta" here is regression coefficient for outcome onto risk score, in both units of risk score standard deviations, and also in units of mmHg predicted mean BP. These data are partly presented in Table 2 of the main text. See Supplementary Material Section 11 for phenotype definitions. All analyses of risk score in relation to other outcomes were unadjusted for BP levels, except for the HVH sample that was (by design) matched for frequency of treated hypertension between MI cases and the common controls. Top vs. bottom quintile and decile contrasts are exponentiated for traits on natural log (ln) scales, but not for traits on linear scales (SBP, DBP, LV mass and LV wall thickness). Pseudo-R^2 is real proportion of variance explained (R^2) for linear model analyses (continuous traits), and ($2\Delta lnL/N$) for binary traits (equivalent to Cox and Snell pseudo-R^2 for logistic regression analyses), where $2\Delta LnL$ is twice increase in log likelihood when risk score included in model ($=(a/aSE)^2$), and N is total sample size. * UK-US Stroke Collaborative Group
References for the contributing studies (if available) are: BRIGHT[123]   C4D[133], CARDIoGRAM[132], CHARGE-HF [124], CKDGen[134], EchoGen[125], KidneyGen[135], NEURO-CHARGE[126], WGHS[139].

Supplementary Table 15: Description of all validated loci by the International Consortium of Blood Pressure Genome-wide Association study and GWAS database lookups.

Column 1 denotes locus (nearest RefSeq gene(s)). Column 2 indicates the sentinel SNP defining the BP association, position on HapMap build 36, the chromosome band, indicates if the SNP is a non-synonymous SNP or a if a non-synonymous SNP is in linkage disequilibrium (r^2) with the sentinel SNP: the location of the SNP relative to the nearest gene(s) is also indicated. Column 3 describes loci in the region and biological plausibility. Column 4 indicates if prior association with BP exists to this region; Column 5 indicates if experimental models exist with a BP or vascular phenotype, or whether there has been prior linkage of the same region in experimental hypertensive models. Column 6 describes association with other diseases or traits either at this SNP and all the proxies in high linkage disequilibrium ($r^2>0.8$) or at all RefSeq genes within 100kb. Column 7 shows the results from look ups of sentinel or proxy SNPs in two databases: NHGRI and an Open Access Database of GWAS[142,143]. Only findings for the other disease/trait reaching a significance of $p<1\times10^{-6}$ or lower are shown. The number of proxy SNPs in the look up were: rs17367504: 5; rs2932538: 41; rs13082711: 27; rs3774372: 117; rs419076: 22; rs1458038: 1; rs13107325: 1; rs13139571: 8; rs1173771: 7; rs11953630: 21; rs1799945: 4; rs805303: 6; rs4373814: 3; rs1813353: 4; rs4590817: 1; rs932764: 4; rs11191548: 21; rs7129220: 28; rs381815: 2; rs633185: 8; rs10850411: 10; rs3184504: 3; rs17249754: 9; rs1378942: 8; rs2521501: 2; rs12946454: 1; rs17608766: 0; rs12940887: 0; rs1327235: 2; rs6015450: 13.

| Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
MTHFR-NPPB	rs17367504 ch1: 11,785,365 1p36.3 in intron in *MTHFR*	MTHFR is involved in methionine and homocysteine metabolism. MTHFR variants influence homocysteinemia[144]. Mild hyperhomocysteinemia is associated with endothelial dysfunction[145]. Other plausible candidate genes in this region include CLCN6, NPPA and NPPB. *CLCN6* encodes a neuronally expressed chloride channel that has not been previously implicated in BP control. *NPPA/NPPB* locus has been associated with levels of gene product and lower BP[107]	Yes[55,107,146-148]	Mthfr-deficient mice show endothelial dysfunction[149] Mice KO for ANP show salt-sensitive hypertension[150]. Nppb-null mice[151] display cardiac fibrosis.	*MTHFR* appears to influence susceptibility to occlusive vascular disease such as stroke[152,153], to neural tube defects[154] and to cancer[155]. Usually, interregional differences have been reported. *NPPA* has been associated with atrial fibrillation[156]	**Systolic blood pressure**[55], Plasma homocysteine[157,158] Atrial fibrillation[156]

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
MOV10	**rs2932538** chr1:113,018,066 1p13.2 intronic in an EST that spans *MOV10*	MOV10 protein is a putative helicase, which may be involved in mRNA silencing, and is associated with the 60S ribosome subunit It interacts with hepatitis delta antigen and facilitates viral replication. MOV10 may be involved in telomerase progression. The signal is 2kb downstream of *CAPZA1*, which is a member of the F-actin capping protein alpha subunit family and encodes the alpha subunit of the barbed-end actin binding protein. This modulates protein kinase signaling to cardiac myofilaments[159]. The top SNP is 27kb downstream of *RHO kinase C*, which encodes a member of the Ras superfamily of small GTP-binding proteins.	No	Rho kinase inhibition with fasudil in L-NAME treated spontaneously hypertensive rats has been found to alter RHO kinase C expression and improve parameters of renal function[160]	None	-
SLC4A7	**rs13082711** chr11:100,098,748 11q22.1 ~10kb 5' of *SLC4A7*	The only adjacent gene encodes the solute carrier family 4 member 7 (*SLC4A7*). An electro-neutral sodium bicarbonate co-transporter, also known as, NBC3 and NBCn1[161]. *SLC4A7* is located on basolateral epithelial cell surface in the thick ascending limb and the distal collecting ducts and in intercalated cells (either apically or basolaterally) in the connecting and collecting tubules[162]. In the thick ascending limb SLC4A7/NBCn1 may be important for ammonium reabsorption.	No	NBCn1 mediates the Na+-dependent bicarbonate transport important for pH regulation in smooth muscle cells of mouse mesenteric, coronary, and cerebral small arteries [163]. KO models with disrupted SLC4A7 suffer from blindness and deafness but there is no renal or BP phenotype documented [164]	The *SLC4A7* region has recently been linked to erythrocyte lead levels in dizygotic twin pairs[165]. There is no specific evidence that SLC4A7 functions as a lead transporter. Prior association of blood lead and heavy metal levels with increased blood pressure. Note we found association of SLC39A8 (which transports heavy metals, e.g. cadmium) and BP	Breast cancer[166]
ULK4	**rs3774372** chr3:41,852,418 3p22.1 missense SNP in *ULK4*	Two SNPs (rs1716975 and rs2272007) linked to rs3774372 are associated with altered *ULK4* gene expression in lymphoblastoid cell lines. ULK4 encodes a serine/threonine protein kinase, the role of which is currently unknown.	Yes[91]	No	No	**Diastolic blood pressure**[91]
MECOM	**rs419076** chr3:170,583,580 3q26.2 in intron 1 in MDS1	rs419076 is located within the *MECOM* locus that may contain two genes, the *MDS1* gene (myelodysplasia syndrome protein 1) and the *EVI1* gene (ecotropic viral integration site 1 isoform a). MDS1 exists in normal tissues both as a unique transcript and as a normal fusion transcript with EVI1[167]. Little is known about the function of MDS1 as a unique transcript. EVI1 is a zinc-finger protein and GATA-binding transactivator[168] and transcriptional repressor[169].	Yes[55,91]	No	EVI1 may be required for heart development[170] EVI1 and/or MDS1 are found as fusion transcripts with the transcription factor AML1 in leukemic cells[169]	Systolic blood pressure[91], Diastolic blood pressure[55] Hypertension[171] nasopharangeal cancer[172], serum magnesium[173]
FGF5	**rs1458038** chr3:41,852,418 ~20kb 5' of *FGF5*	FGF5 stimulates cell growth and proliferation in different cell types including cardiomyocytes. The FGF5 has been related to angiogenesis in the heart. Related to diastolic blood pressure but mechanism is unclear.	Yes[42,55,147,174,175]	No	Involved in angiogenesis in the heart	Diastolic blood pressure[55] Blood pressure and hypertension[42,175]

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
SLC39A8	**rs13107325** chr4:103,407,732 4q24 missense SNP in *SLC39A8*	A missense variant within *SLC39A8* (solute carrier family 39 (zinc transporter) member 8, that encodes a zinc transporter. The mouse orthologue transports zinc, cadmium (Cd) and manganese[176]	No	None	Transgenic Cd-sensitive mice also exhibit acute renal failure and proximal tubular damage when exposed to cadmium[176,177]. Association between cadmium levels and BP/hypertension exists in animal models[178].	**HDL cholesterol[179, 180]**
GUCY1A3-GUCY1B3	**rs13139571** chr4:156,864,963 4q32.1 in intron 10 in *GUCY1A3*	The guanylate cyclase 1, soluble, alpha 3, *GUCY1A3* lies 23Kb upstream of *GUCY1B3* encoding α and β subunits of the soluble guanylate cyclase (sGC). A signal transduction enzyme activated by nitric oxide (NO). sGC converts GTP to the second messenger cGMP influencing cardiovascular homeostasis including smooth muscle tone and growth, vascular permeability, platelet reactivity and leukocyte recruitment[181].	Yes	Spontaneously[182] hypertensive rats have impaired endothelium dependent vasodilatation, decreased *GUCY1A3* and *GUCY1B3* mRNA levels and reduced sGC activity [181,183]. KO mice deficient for the β subunit have raised SBP, platelet dysfunction [184]. The smooth-muscle β subunit knockout mice are hypertensive with no other phenotype [185].	KO of β subunit develops fatal gastrointestinal obstruction[184]	-
NPR3-C5orf23	**rs1173771** chr5: 32,850,785 5p13.3 ~20kb of both *C5orf23* and *NPR3*	NPR3 codes for the natriuretic peptide clearance receptor (NPR-C) expressed in the heart, kidney, and vascular smooth muscle[186,187] Genetic variation reducing production of NPR-C or altering function may reduce clearance of natriuretic peptides, and lower BP. NPR-C/Gi coupling opens a Gi-gated inwardly rectifying potassium channel (GIRK) and may underlie CNP (C-natriuretic peptide) acting as an endothelium-derived hyperpolarising factor affecting smooth muscle and relaxing resistance vessels. CNP actions may exert anti-atherogenic and anti-platelet properties through NPR-C[188]. The NPPA/NPPB locus on chr1 has been associated with levels of gene product and lower BP[107].	No but NPPA/B genotype and gene products have been related to BP (see column 3)	NPR3 knockout mice demonstrate substantial delays in ANP/BNP plasma clearance[187,188]	The *NPR3* locus has been associated with height & skeletal parameters in Europeans[189,190]. Osteocrin augments NPRC induced cGMP production in chondrocytes and osteoblasts in bone and may explain CNP/NPRC involvement in skeletal phenotypes[191]. A study indicates suggestive association of *NPR3* with post coronary bypass left ventricular dysfunction[192].	Height[189,190]
EBF1	**rs11953630** chr5:157,777,980 5q 33.3 250Kb from *EBF1*	The. EBF1 (Early B-cell factor 1) is associated with olfactory signal transduction[193] The marker lies c.800kb upstream of SOX30[194] and ADAM19 an enzyme (a disintegrin and metalloproteinase) family, it is involved in cell-cell and cell-matrix interactions[195].	No	No	ADAM19 is associated with renal disease[195].	Type I Diabetes[171], Serum bilirubin levels[196], Serum markers of iron status[197,198], **Hemoglobin levels**[199,200], Hematocrit[200], Mean corpuscular volume[200], Hematological parameters[190]

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
HFE	**rs1799945** chr6:26,199,158 6p22.1 encodes a histidine to aspartic change at amino acid 63 (H63D) in *HFE*	SNP rs1799945 encodes a histidine to aspartic change at amino acid 63 (H63D) of the hemochromatosis gene (*HFE*)[201]. It is unclear how loss of function mutations in *HFE* could influence blood pressure but it could be a downstream consequence of altered total body iron stores or a function of altered HFE-hepcidin-signaling pathways. The minor alleles of SNP rs198846 and perfect proxy rs1799945 (MAF 0.15), in the HLA region on chromosome 6, were associated with higher DBP ($P = 3x10^{-11}$). The minor allele rs1800562 (MAF 0.04) is also weakly associated with increased DBP in the discovery GWAS ($P = 3x10^{-4}$) and modestly stronger but not meeting the threshold for further genotyping at stage 1 validation ($P = 2x10^{-5}$).	No but C282Y associated with BP treatment[202].	KO model for *HFE* exists but no BP phenotypes reported[203].	This variant is a low penetrance allele for hemochromatosis[201] and is involved in microvascular complications of diabetes[204]. Hemochromatosis is characterized by iron overload with multi-organ dysfunction caused by mutations in several genes that regulate hepcidin-induced degradation of ferroportin. The most common cause is homozygosity for the minor allele of rs1800562 (C282Y); this polymorphism, with H63D, is also a likely regulator of Haemoglobin and other red cell phenotypes in the general population[200].	Hemoglobin, haematocrit, MCH and MCV levels[200]
BAT2-BAT5	**rs805303** chr6:31,724,345 6p21.33 intronic in *BAT3*	This signal is in a region of extended haplotypes in the major HLA complex 1[205]. rs805303, is in an intron of *BAT3* (HLA-B-associated transcript 3). Implicated in the control of apoptosis and modulation of TGF-β signaling through TGF-β type I and type II in renal mesangial cells[206]. The region includes other genes (see locus column).	No	None	HLA region associated with lung cancer and rheumatoid arthritis[207,208]. The nsSNP rs1046089 in BAT2 has recently been associated with malaria susceptibility[209]. Another SNP in this gene is associated with risk of stroke[210].	Type I Diabetes[211], Weight[212], Lung adenocarcinoma[213], Lung cancer[207,214], Multiple sclerosis[215], Rheumatoid Arthritis[171,216]
CACNB2(5') *CACNB2(3')*	**rs4373814** chr10:18,459,978 10p12.33 ~10kb 5' of *CACNB2* **rs1813353** chr10:18747454 intronic in *CACNB2*	The two SNPs (rs4373814 and rs1813353; pairwise r2=0.015) associated with SBP & DBP point to two distinct causal variants. The CHARGE consortium reported BP association for rs11014166 (r2 0.93 with rs1813353; r2=0.22 with rs4373814) in *CACNB2*[91]. *CACNB2* is expressed in the heart and encodesthebeta-2 subunit of a voltage-gated calcium channel. As a member of a family of voltage-gated calcium channel genes, CACNB2 may regulate BP through interaction of the beta-2 subunit with alpha-1 calcium channels (Ca$_V$1.2)[217].	Yes [91,147]	KO model not phenotyped for BP[218].	Coding variation in *CACNB2* appears to be responsible for Brugada syndrome[219]. KO mice deaf due to impaired auditory nerve conductance[218].	Systolic blood pressure[91], Diastolic blood pressure[91], Hypertension[91], Protein quantitative trait loci[220], Bipolar disorder[171]

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
C10orf107	**rs4590817** chr10:63,137,559 intronic in C10orf107	Chromosome 10 open reading frame with no clear neighbouring candidate gene or functional implication in blood pressure.	Yes [55,91,147]	No	None	Systolic blood pressure[55]
PLCE1	**rs932764** chr10:95,885,930 10q23.33 intronic in *PLCE1*	The phospholipase–C–epsilon -1 (PLCE-1) isoform catalyses the hydrolysis of polyphosphoinositides affecting second messenger cascades influencing cellular growth, differentiation and gene expression. PLCE1 is important for normal podocyte development within the glomerulus of the kidney[221]. Within this interval of high LD there is an adipocyte differentiation gene (*NOC3L*) and the pseudogene *PIPSL*. Neither represent good candidates.	Yes[222,223].	No	Implicated in familial nephrotic syndromes with end-stage kidney disease[221,222]. In focal segmental glomerulonephritis characterized by glomerulosclerosis[222,223]. A type of glomerulosclerosis, the Kimmelstiel-Wilson lesion, has been described in severe hypertension with nephropathy and proteinuria.	Eosophageal squamous cell carcinoma[224], gastric adenocarcinoma[225]-
CYP17A1-NT5C2	**rs11191548** ch10: 104,836,168 5'-UTR of *NT5C2*	*CYP17A1* encodes the cytochrome P450 enzyme CYP17A1 that mediates steroid 17α-hydroxylase and 17.20-lyase activity. The first enzymatic action is a key step in the biosynthesis of mineralocorticoids and glucocorticoids that affects sodium handling in the kidney and the second is involved in sex-steroid biosynthesis. Missense mutations in *CYP17A1* cause one form of adrenal hyperplasia characterized by hypertension, hypokalemia and reduced plasma renin activity[226,227]. *AS3MT* is involved in arsenic metabolism[228]. Chronic arsenic exposure causes high BP in humans[228] and animals potentially via increased oxidative stress[229]. For the other genes present in this locus (C10orf32, CNNM2 and NT5C2), no physiological link with blood pressure could be found.	Yes[55,91,147]	As3mt KO mice have decreased ability to metabolise arsenic[230]		**Systolic blood pressure**[55], intracranial aneurysm[231]
ADM	**rs7129220** chr11:10,307,114 ~20kb 3' of *ADM*, intronic in an EST	*ADM* codes for a pro-peptide that is cleaved to generate a widely expressed, 52 amino acid protein, adrenomedullin with vasodilator and BP regulatory properties. Plasma adrenomedullin is associated with mean arterial pressure in men and with pulse pressure in women[232] and pro-adrenomedullin was reported to be associated with pulse pressure and hypertensive target organ damage in African Americans[233].	Yes	*ADM*[+/-] mice exhibit elevated blood pressure and diminished nitric oxide production[234]. *ADM* mice expressing 50% -140% mRNA do not exhibit basal BP changes[235]	Polymorphisms in the human *ADM* gene have been associated with genetic predisposition to diabetic nephropathy and proteinuria with essential hypertension.	Plasma HDL-cholesterol[236], human stature[237], bipolar disorder[238]

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
PLEKHA7	**rs381815** chr11: 16,858,844 intronic in *PLEKHA7*	Pleckstrin homology domain-containing family member A7 is located at zonula adherens in epithelial cells and is thought to be responsible for microtubule adherence at the apical end of these cells. The role of this gene product in blood pressure is unclear	Yes[91,147]	No	No	**Systolic blood pressure**[91], **Diastolic blood pressure**[91] Blood pressure and hypertension[175]
FLJ32810-TMEM133	**rs633185** chr11:100,098,748 11q22.1 intronic in *FLJ32810*	FLJ32810 encodes a Rho-type GTPase activating protein with SH3 and pleckstrin binding domains; it has no known or suspected roles in blood pressure control or hypertension. SNP is >200kb from *TMEM133* & the progesterone receptor (PGR) gene encodes a member of the steroid receptor superfamily and mediates the physiological effects of progesterone, which plays a central role in reproductive events associated with the establishment and maintenance of pregnancy.	No	No	None	None
ATP2B1	**rs17249754** chr12: 88,584,717 ~5kb 5' of *ATP2B1*	*ATP2B1* encodes a plasma membrane calcium/calmodulin handling ATPase (PMCA1) which is implicated in calcium efflux from cells. It is expressed in the endothelium.	Yes 42,91,239	Rat aortic smooth muscle expression is enhanced in the spontaneously hypertensive rat		**Diastolic blood pressure**[42,91], **Hypertension**[91], **Systolic blood pressure**[42,91], **Biomedical quantitative trait**[221], serum magnesium[173]
SH2B3	**rs3184504** chr12: 110,368,991 missense SNP in *SH2B3*	This SNP is intronic within the ataxin gene, which has been linked to the autosomal dominant cerebellar ataxias. It is within a cluster of highly correlated SNPs spanning 200kb and is perfectly correlated with a missense SNP in *SH2B3* (rs3184504, R262W, r² in CEU to rs653178 = 1.0), The same minor allele of rs3184504 has been associated with type 1 diabetes and celiac disease. The SH2B3 protein (also known as lymphocyte-specific adapter protein, LNK) is one of a subfamily of SH2 domain-containing proteins and is implicated in growth factor, cytokine, and immunoreceptor signaling. Within the associated interval lies ALDH2 (acetaldehyde dehydrogenase type 2) a key enzyme in alcohol metabolism which has previously be associated with BP levels	Yes[55,91,147]	No	Knockout mice of SH2B3 are viable but sensitised to cytokines	Systolic blood pressure[55,91], Diastolic blood pressure[55,91], Type 1 diabetes[171,240], Hematocrit and hemoglobin[200], **Plasma eosinophil count**[241], **Celiac disease**[242] myocardial infarction[241], retinal vascular caliber[243]
TBX5-TBX3	**rs10850411** chr12:113,872,179 12q24.21 ~200kb from *CR591392* and 300kb from *TBX3*	T-box genes share a common DNA Binding domain and encode transcription factors involved in the regulation of developmental processes. TBX5 acts as an activator and competes with TBX3 which acts as a repressor for various myocardial genes. A mutation in TBX5 causes Holt Oram syndrome which may be associated with altered expression of natriuretic peptides.	Yes[91,147]	TBX5 Knockouts have marked reduction in natriuretic peptides. Not phenotyped for BP[244].	Holt Oram Syndrome which affects cardiac development and paroxysmal AF[245] TBX3 and TBX5 are both associated with PR. Common variants modulate heart rate, PR interval and QRS duration[246,247].	Hypertension, Type 2 Diabetes, Coronary Artery Disease[171], Gallstone disease[248].

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
CYP1A1-ULK3	**rs1378942** ch15: 72,864,420 intronic in *CSK*	Human CYP1A2 is one of the major CYPs in human liver and metabolizes a number of drugs (e.g. caffeine, theophylline, propranolol and verapamil), carcinogens and endogenous compounds (e.g. melatonin and estrogens). There are large inter-individuals variability in CYP1A2 expression according to *CYP1A2* genotypes. *CYP1A2* influence the association of coffee intake with hypertension[249]. *In vitro* experiments showed that *CSK* is involved in the reorganisation of the actin cytoskeleton[250]. Actin cytoskeleton might play a role in blood pressure control by influencing Na-K ATPase activity[251].	Yes[42,49,55,91,147,249]	Cyp1a2-null mice and cyp1a1-null mice both exist, but marked interspecies differences in substrate specificity limit the extrapolation of animal findings to humans. Cyp1a1 KO mice, show that CYP1A1 activity is needed for halogenated aromatic hydrocarbons-associated endothelial dysfunction and hypertension[252]. The ARID3b gene is embryonic lethal when knocked out in mouse, with branchial arch and vascular developmental abnormalities[253].	CYP1A1 and CYP1A2 share common xenobiotic response elements[254] and influence the metabolism of numerous drugs. *CYP1A2* variants influence the association of coffee intake with myocardial infarction[255].	Diastolic blood pressure[55,91]
FURIN-FES	**rs2521501** chr15:89,238,392 15q26.1 in intron 18 of *FES*	SNP is 11kb downstream of *FURIN*, encoding a type-1 membrane bound protease and member of the subtilisin-like proprotein convertase family (known as PCSK3). It activates proteins including: receptors (insulin and the hepatocyte growth factor), plasma proteins (complement c3, von Willebrand factor) and hormones (nerve growth factor, endothelin, parathormone, transcription growth factor type beta). It solubilises hemojuvelin and pro-hepcidin (see HFE) the pro-renin receptor, and cleavage of the α&γ subunits of the epithelial sodium channel (eNAC) [256, 257,258]. Mutations of eNAC cause Mendelian forms of hyper/hypotension. The v-fes feline sarcoma viral oncogene (*FES*) regulates innate immune response.	Yes[259]	No	No	Attention deficit hyperactivity disorder[260].

Locus Name	Sentinel SNP, Chr: position, chromosomal band, relative SNP position	Description of genes in region and biological plausibility	Related to human BP ?	Animal model with BP or vascular phenotype or regional linkage to BP ?	Relationship of gene, region, SNP or gene product with another disease	Lookup in NHGRI GWAS table and non-NHGRI GWAS database: by sentinel SNP and all proxies r²>0.8, by all RefSeq genes within 100kb
GOSR2	**rs17608766** chr17:42,368,270 17q21.32 intronic in *GOSR2*; within a tight LD block across *GOSR2*	Golgi SNAP receptor complex member 2 (GOSR2) gene encodes three different isoforms of a trafficking membrane protein which transports proteins among the medial- and trans-Golgi compartments[261] This locus has been linked with hypertension and/or blood pressure in familial linkage studies[262,263]. Recently the non-synonymous Lys67Arg SNP within GOSR2 (rs197922) associated with hypertension in whites[264] (see also other diseases).	Yes[264]	Homologous to a region of chromosome 10 which contains a QTL for blood pressure in a rat model of hypertension[265,266].	The same SNP has been associated with aortic root dilatation in a meta-analysis of genes for cardiac structure[125]. An association of a nsSNPLys67Arg within *GOSR2* (rs197922) with coronary heart disease and hypertension in whites has been reported in r² 0.26 with rs17608766[264].	-
ZNF652	**rs12940887** chr17:44,757,806 17q21.32 intronic in *ZNF652*	GBPG has reported association with DBP of SNP rs16948048 near ZNF652 and PHB, which is in high LD (r2 = 0.9) with the SNP reported here.	Yes[55,147,175]	No	ZNF652 has been implicated in tumorigenesis[267]	Diastolic blood pressure[55]
JAG1	**rs1327235** chr20:10,917,030 20p12.2 ~300kb 5' of *JAG1*, intronic within an EST	The *JAG1* (jagged 1) locus is in the center of the short arm of chromosome 20[268]and encodes a ligand for the Notch receptor. Jagged/Notch interactions are critical for determination of cell fates in early development.	Yes, see other diseases	Endothelial-specific deletion of *JAG1* in mice results in striking deficits of vascular smooth muscle[269]).	Mutations in *JAG1* cause Alagille syndrome, which has cholestatic, skeletal, cardiac, ocular, and facial characteristics and includes renal involvement with hypertension[270,271].	-
GNAS-EDN3	**rs6015450** chr20:57,184,512 20q13.2 in intergenic region near *EDN3* & 10kb from *ZNF831* and *MRPS16P*	The endothelin 3 (*EDN3*) gene is a strong candidate for blood pressure regulation. The endothelins are widely expressed vasoactive peptides that exert proliferative, inflammatory, and fibrotic changes in blood vessels and other organs involved in the regulation of vascular tone and blood pressure. The endothelin receptors are being investigated as treatments for hypertension and vascular disease[182]. The SNP is near *ZNF831* and the pseudogene *MRPS16P* with no known function. *GNAS* encodes the alpha subunit of the heterotrimeric G-protein, which mediates signal transduction at the β_1 and β_2 adrenergic receptors, influencing heart rate and smooth muscle tone[272].	Yes[91].	In a rat model, the EDN3 gene has been closely linked to a QTL for blood pressure and heart weight[273]. In a mouse model, EDN3 was involved in axonal direction of developing sympathetic neurons toward intermediate vascular targets and selected end-organs[274].	Mutations in *EDN3* and the EDN3 receptor B are associated with Hirschsprung disease and Waardenburg syndrome, congenital disorders involving neural crest-derived cells[275]. Altered EDN3 gene expression reported in breast cancer[276]. Response to treatment in schizophrenia In patients with pulmonary arterial hypertension, EDN3 levels were lower (and endothelin 1 higher) compared with controls correlating with hemodynamic markers of disease severity[277].	Hypertension[91], Response to antipsychotic therapy - extrapyramidal side effects[278]

17. Supplementary Figures

See the following 10 pages.
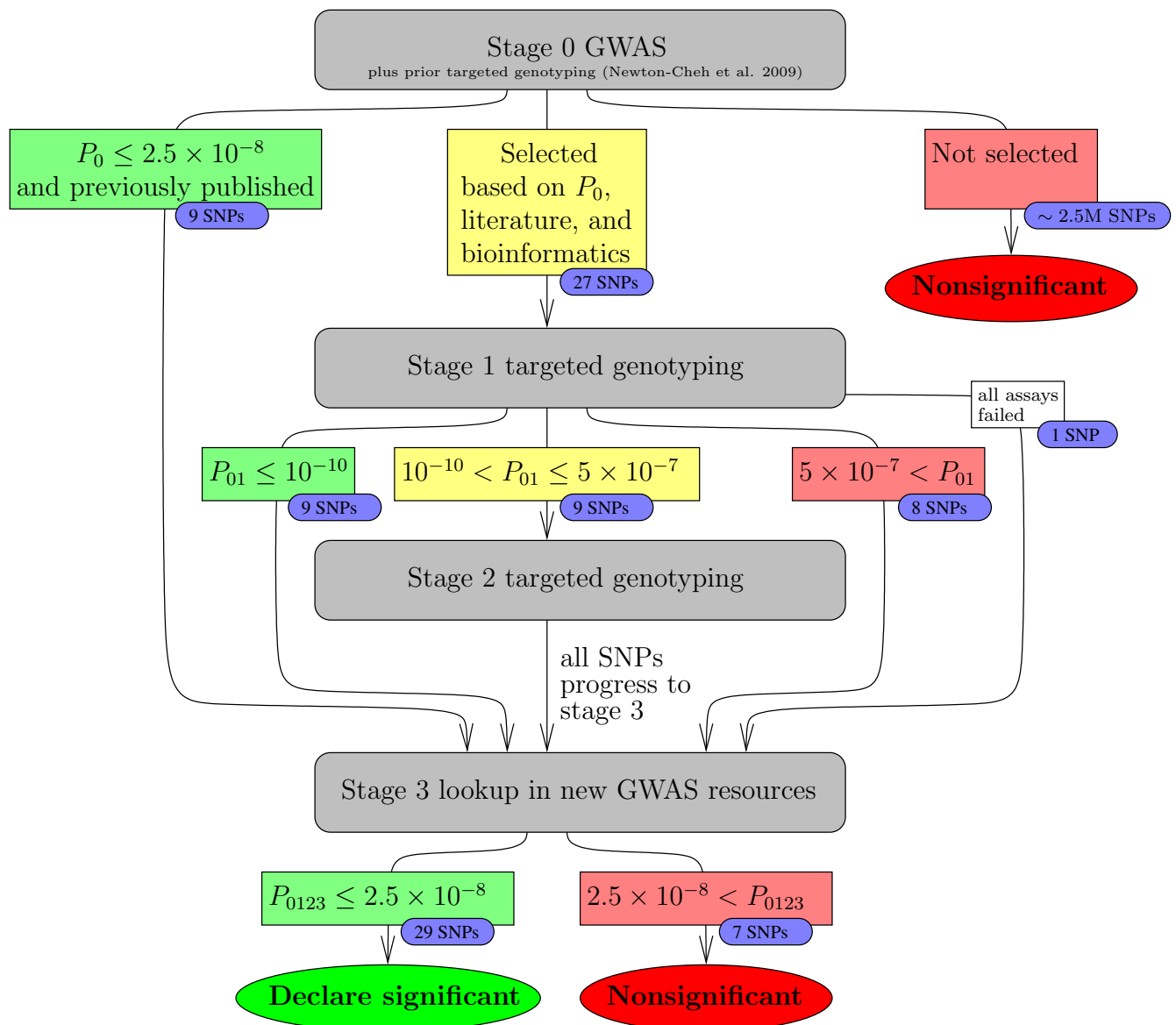
Suppl. Fig. 1: Sequential design used for the ICBP GWAS discovery and validation study. Decision about the progression of each SNP (arrows and rectangles) depended on the association P-value calculated using data from all previous stages, denoted P_A where $A \subseteq \{0, 1, 2, 3\}$. Additional targeted genotyping was considered not to be cost effective both for SNPs with sufficiently strong evidence (green rectangles on left, indicated by stars in Supplementary Table 4) and also for SNPs with sufficiently weak evidence (red rectangles on right, indicated by daggers in Supplementary Table 4).
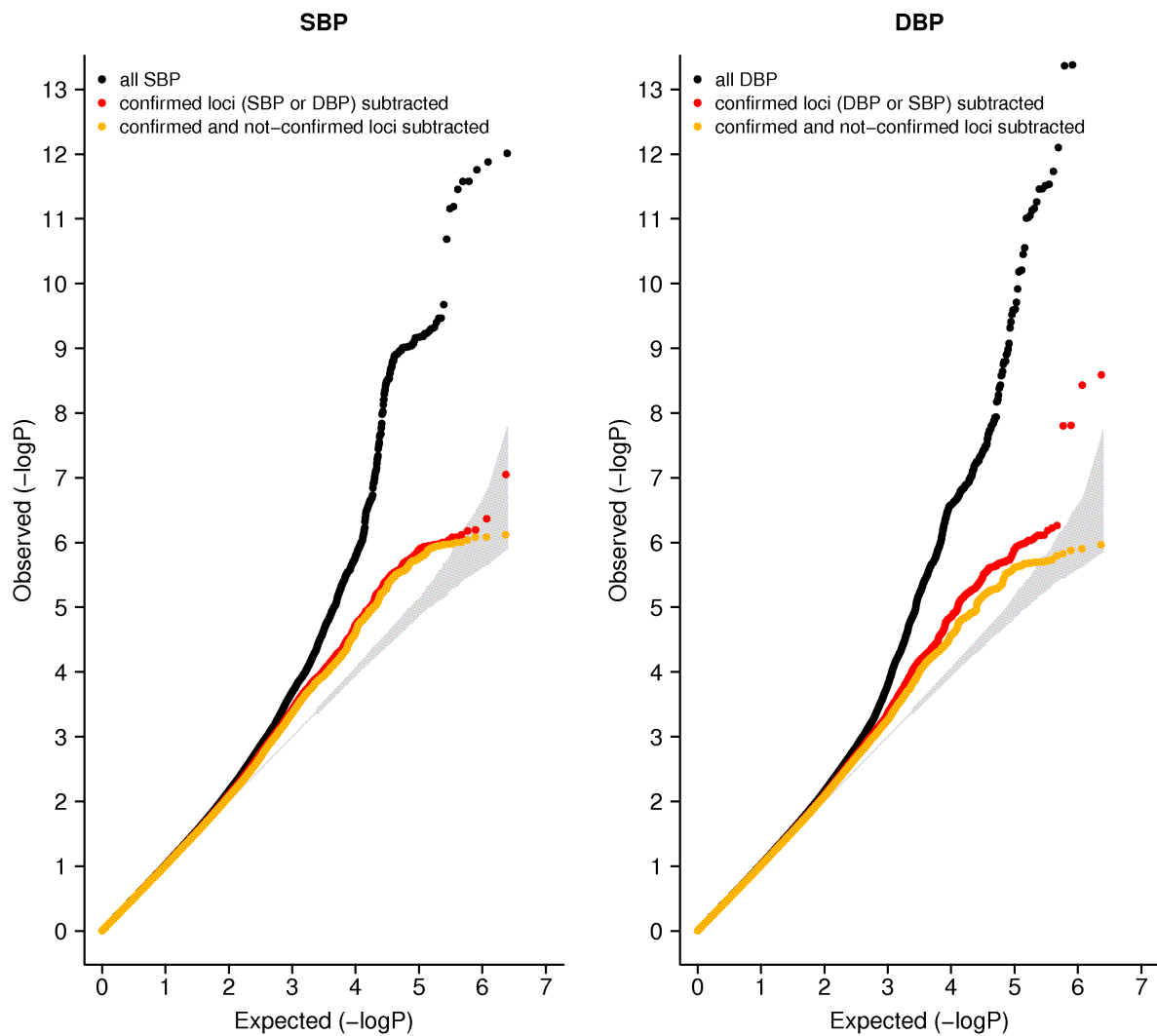
Suppl. Fig. 2: QQ plots for meta-analysis associations statistics. Results for all $\sim$ 2.5M SNPs with $N_{\text{effective}} \geq 17,500$ are plotted in black. Results excluding all SNPs within 1Mb of 29 confirmed signals are plotted in red, and results excluding SNPs within 1Mb of 36 signals for which validation was attempted (see Supplementary Table 4) are plotted in orange.
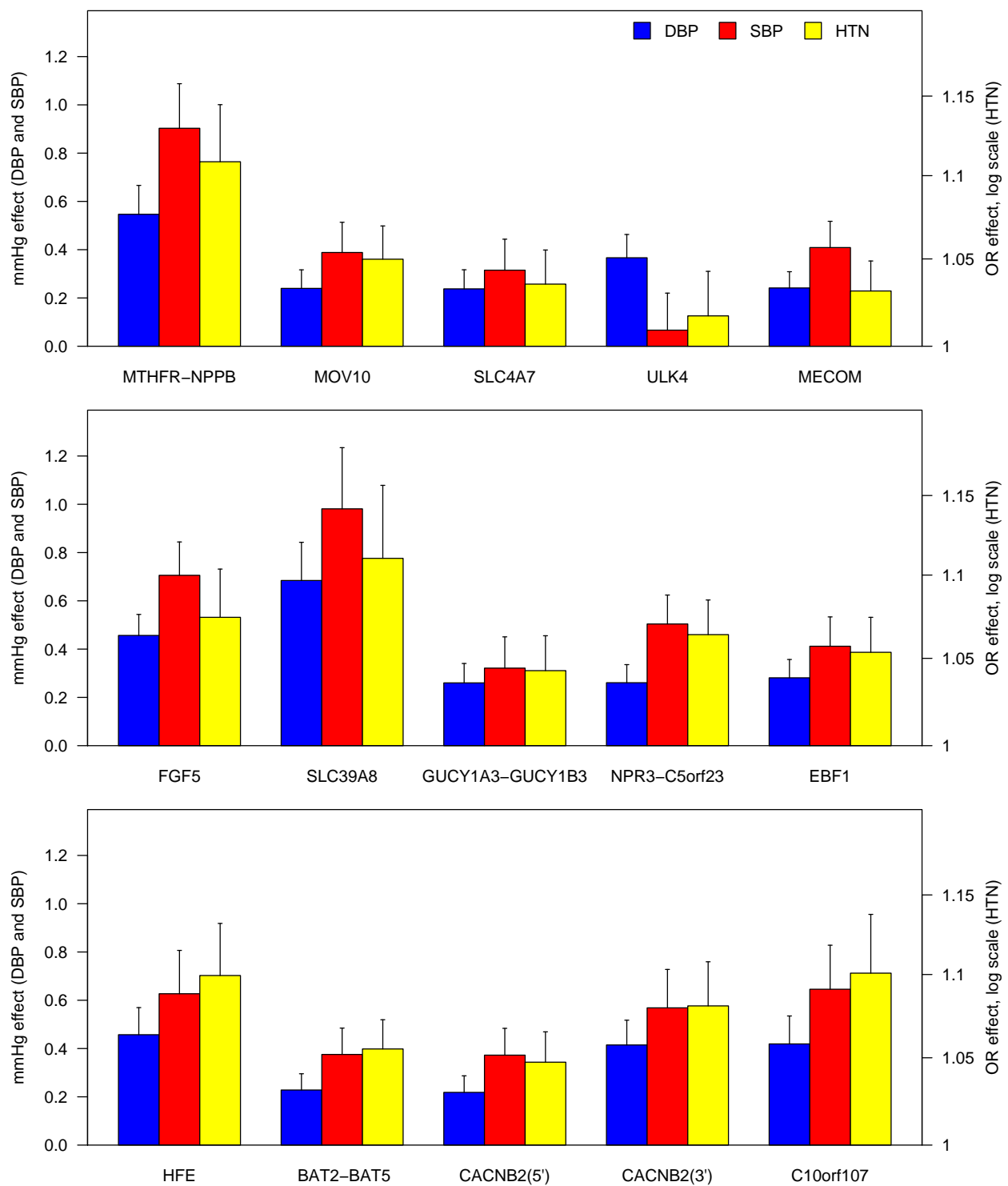
Suppl. Fig. 3: Concordance of effects on DBP, SBP, and hypertension. The per-allele mmHg and odds ratio (OR) effect sizes are scaled relative to each other such that the median effect size (over all SNPs) is a bar of the same height.
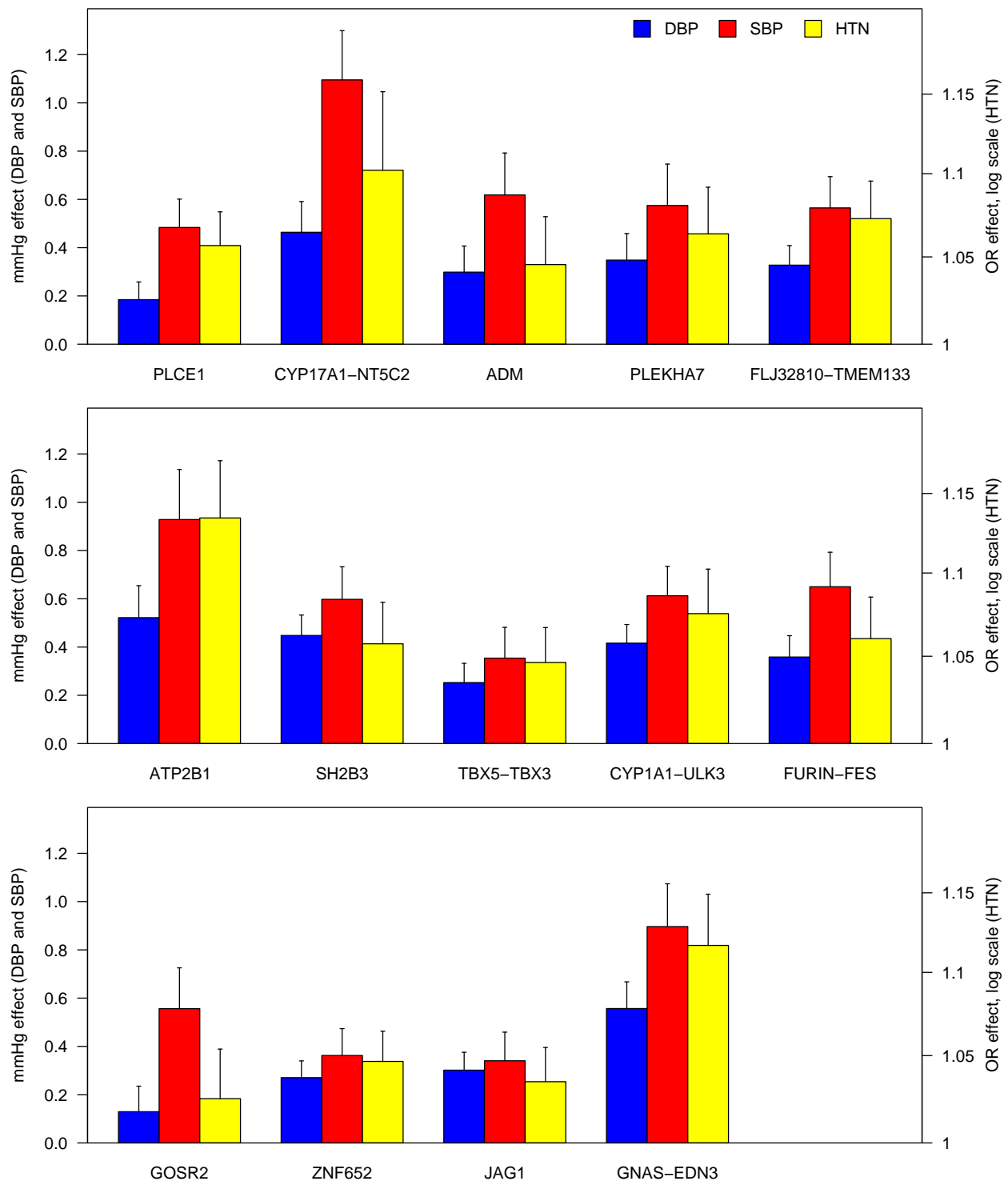
Suppl. Fig. 3: (continued).
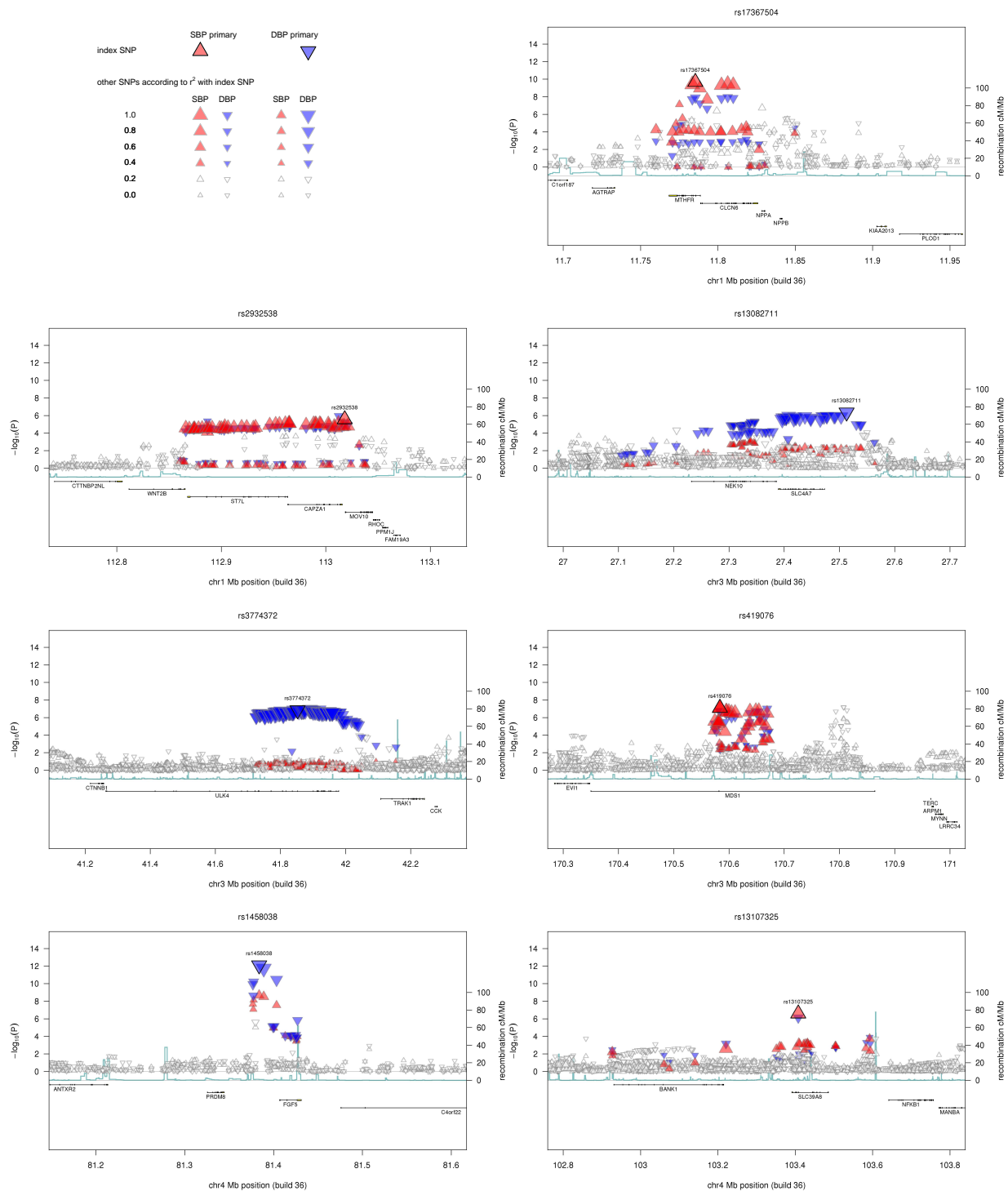
Suppl. Fig. 4: Regional association plots (continued over three more pages)
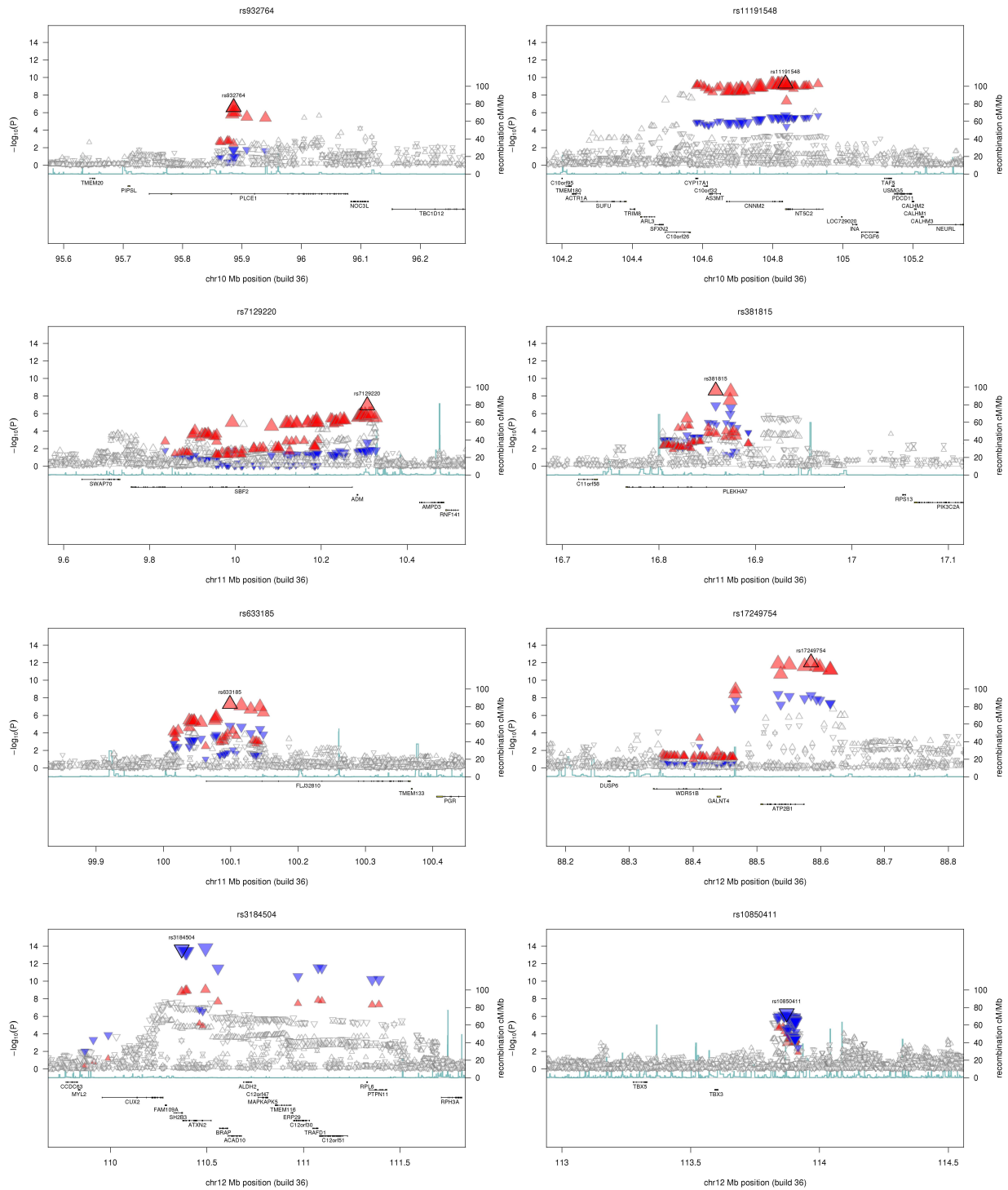
Suppl. Fig. 4: (continued) Regional association plots

Suppl. Fig. 4: (continued) Regional association plots
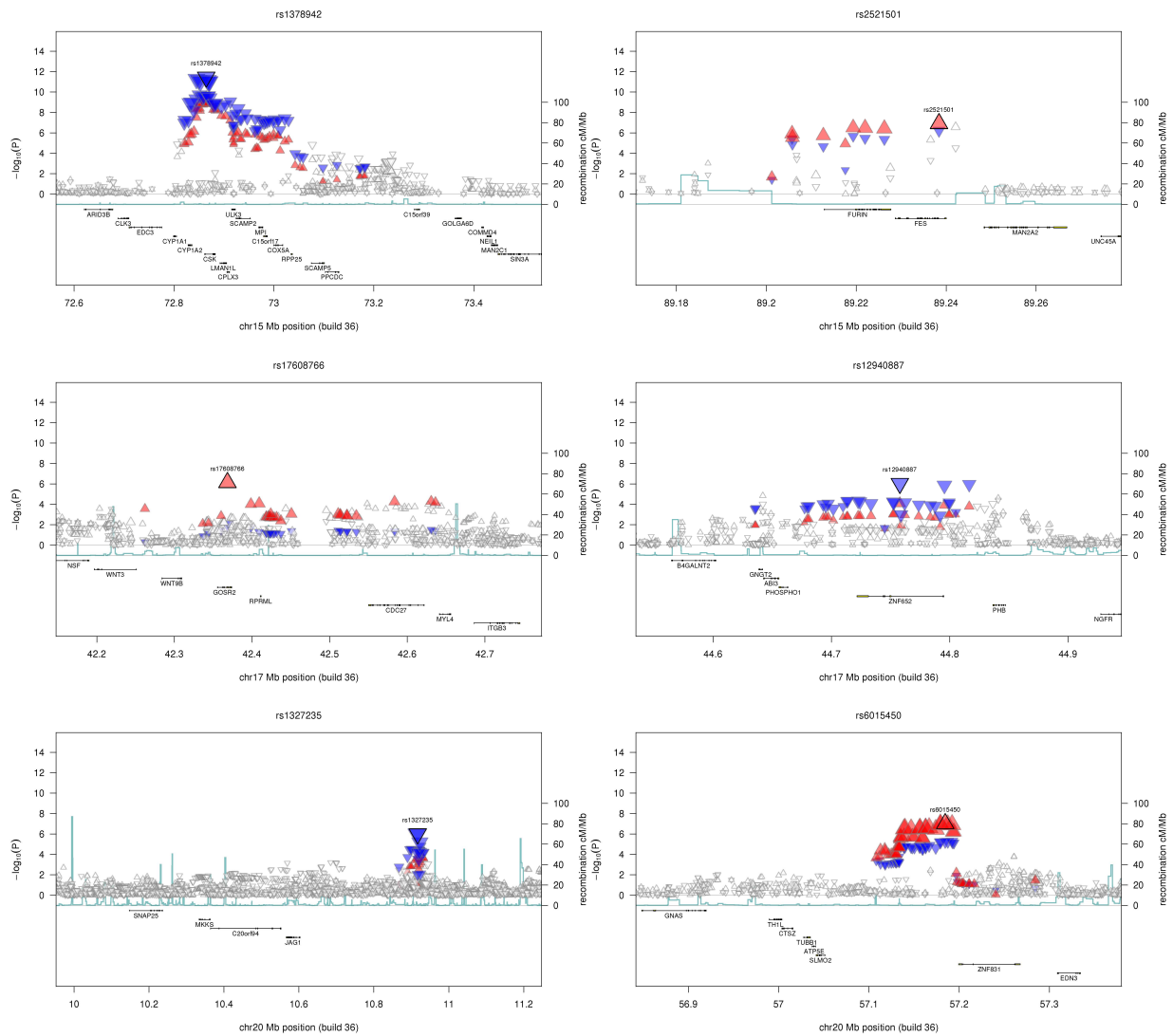
Suppl. Fig. 4: (continued) Regional association plots
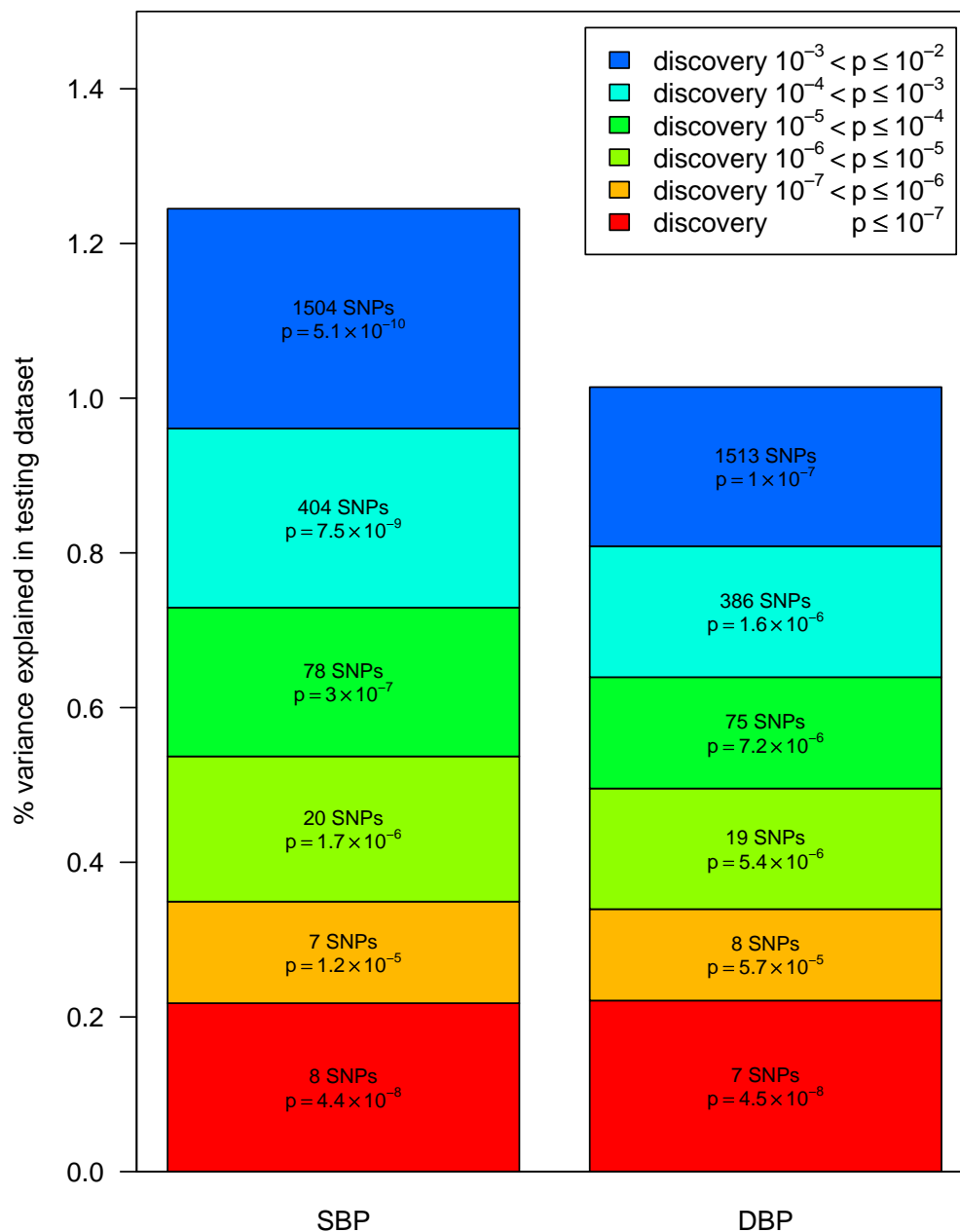
Suppl. Fig. 5: The contributions of additional common variants of modest effect were studied in 50 distinct cohortwise partitions of our primary (stage 0 GWAS) dataset. For each partition, $\sim 80\%$ of the N=69,899 samples were used for "discovery" and the nonoverlapping $\sim 20\%$ were used for "testing". Each block in the figure represents SNPs with discovery P-values in a given bin. We selected SNPs that were effectively independent of each other and independent of all SNPs with more significant discovery P-values using a pairwise $r^2 < 0.05$ threshold. The percentage of phenotypic variance explained in the testing subset (mean over 50 partitions, plotted on the y-axis) is therefore approximately additive across bins. Each block is annotated with the number of selected SNPs, and the P-value for a 1 d.f. test of association with the resulting multi-SNP risk score in the testing subset (both are medians over 50 partitions).
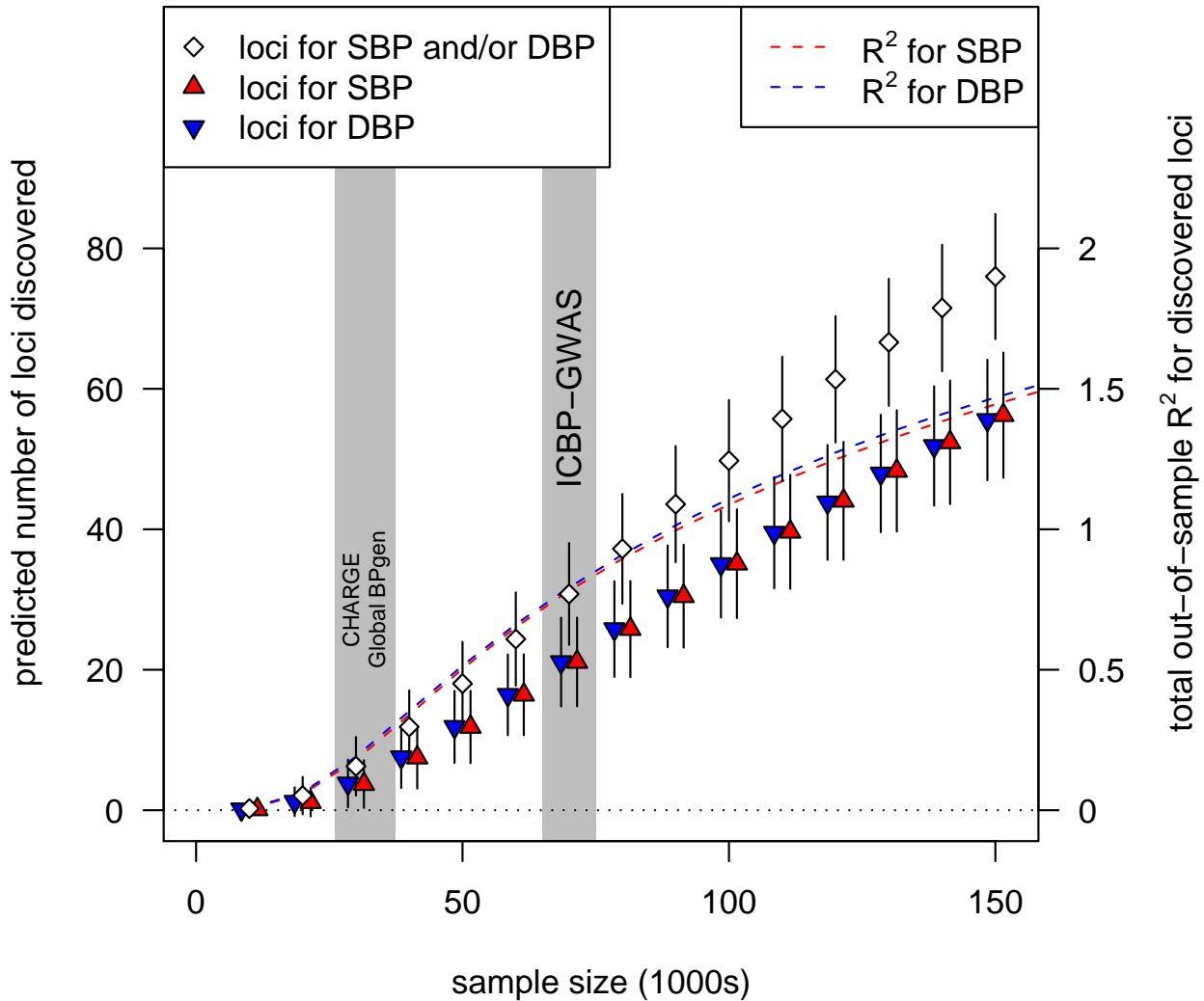
Suppl. Fig. 6: We used the estimated distribution of effect sizes for the total 29 independent associations validated here and previously, and standard power calculations, to estimate the total number of independent associations that could be detected as a function of sample size (see Supplementary Appendix A for details). The error bars represent variability across replicate discovery experiments with the same sample size, assuming the point estimate of 116 variants in total is fixed.

18. References

1   Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet* **3.** e114, (2007).

2   Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat.Genet* **39.** 906-913, (2007).

3   Li, Y. & Abecasis, G. R. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum.Genet.* **S79.** 2290, (2006).

4   Project, T. I. H. The International HapMap Project. *Nature* **426.** 789-796, (2003).

5   Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10.** 387-406, (2009).

6   Guan, Y. & Stephens, M. Practical issues in imputation-based association mapping. *PLoS Genet* **4.** e1000279, (2008).

7   Harris, T. B. *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* **165.** 1076-1087, (2007).

8   Rose, G. A., Blackburn, H. & . Cardiovascular population studies:Methods. (World Health Organisation, Geneva, 1966).

9   Mitchell, B. D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am Heart J* **155.** 823-828, (2008).

10   Rampersaud, E. *et al.* The association of coronary artery calcification and carotid artery intima-media thickness with distinct, traditional coronary artery disease risk factors in asymptomatic adults. *Am J Epidemiol* **168.** 1016-1023, (2008).

11   Investigators, A. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129.** 687-702, (1989).

12   Oren, A. *et al.* The Atherosclerosis Risk in Young Adults (ARYA) study: rationale and design. *Eur J Epidemiol* **18.** 715-727, (2003).

13   Adams, C., Burke, V. & Beilin, L. J. Accuracy of blood pressure measurement and anthropometry among volunteer observers in a large community survey. *J Clin Epidemiol* **55.** 338-344, (2002).

14   Shock, N. *et al.*   Vol. 84-2450 *The Baltimore Longitudinal Study of Aging* (ed National Institutes of Health) (Government Printing Office, Washington DC, 1980).

15   Musunuru, K. *et al.* Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet* **3.** 267-275, (2010).

16   Fried, L. P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1.** 263-276, (1991).

17   Adair, L. S. *et al.* Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey. *Int J Epidemiol*, (2010).

18   Perneger, T. V. *et al.* A prospective study of blood pressure and serum creatinine. Results from the 'Clue' Study and the ARIC Study. *Jama* **269.** 488-493, (1993).

19   Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* **8.** 6, (2008).

20   Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316.** 1331-1336, (2007).

21   Schwarz, P. E. *et al.* Hypoadiponectinemia is associated with progression toward type 2 diabetes and genetic variation in the ADIPOQ gene promoter. *Diabetes Care* **29.** 1645-1650, (2006).

22   Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1.** 95-103, (1999).

23   Palli, D. *et al.* A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori* **89.** 586-593, (2003).

24   FBPP. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension* **39.** 3-9, (2002).

25   Dawber, T. R., Meadors, G. F. & Moore, F. E., Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* **41.** 279-281, (1951).

26   Feinleib, M., Kannel, W. B., Garrison, R. J., McNamara, P. M. & Castelli, W. P. The Framingham Offspring Study. Design and preliminary data. *Prev Med* **4.** 518-525, (1975).

27   Splansky, G. L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* **165.** 1328-1335, (2007).

28   Kastarinen, M. J. *et al.* Trends in blood pressure levels and control of hypertension in Finland from 1982 to 1997. *J Hypertens* **16.** 1379-1387, (1998).

29   Li, Y. *et al.* Angiotensin-converting enzyme I/D and alpha-adducin Gly460Trp polymorphisms: from angiotensin-converting enzyme activity to cardiovascular outcome. *Hypertension* **49.** 1291-1297, (2007).

30   Valle, T. *et al.* Mapping genes for NIDDM. Design of the Finland-United States Investigation of NIDDM Genetics (FUSION) Study. *Diabetes Care* **21.** 949-958, (1998).

31   Tobin, M. D. *et al.* Common variants in genes underlying monogenic hypertension and hypotension and blood pressure in the general population. *Hypertension* **51.** 1658-1664, (2008).

32   Adeyemo, A. *et al.* A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet* **5.** e1000564, (2009).

33   Holmen, J. The Nørd-Trondelag Health Study 1995-97 (HUNT 2) Objectives, contents, methods and participation. . *Norwegian Journal of Epidemiology* **13.** 19-32, (2003).

34   Org, E. *et al.* Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Hum Mol Genet* **18.** 2288-2296, (2009).

35   Ferrucci, L. *et al.* Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc* **48.** 1618-1625, (2000).

36   Berg, C. *et al.* Eating patterns and portion size associated with obesity in a Swedish population. *Appetite* **52.** 21-26, (2009).

37   Strandhagen, E. *et al.* Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. *Eur J Epidemiol* **25.** 163-172, (2010).

38   Golara, M., Jones, C., Randhawa, M. & Shennan, A. H. Inflationary oscillometric blood pressure monitoring: validation of the OMRON-MIT. *Blood Press Monit* **7.** 325-328, (2002).

39   Henkin, L. *et al.* Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS Family Study design and methods. *Ann Epidemiol* **13.** 211-217, (2003).

40   Cooper, R. *et al.* Evaluation of an electronic blood pressure device for epidemiological studies. *Blood Press Monit* **2.** 35-40, (1997).

41   Ataman, S. L. *et al.* Standardization of blood pressure measurement in an international comparative study. *J Clin Epidemiol* **49.** 869-877, (1996).

42   Takeuchi, F. *et al.* Blood pressure and hypertension are associated with 7 loci in the Japanese population. *Circulation* **121.** 2302-2309, (2010).

43   Imai, Y. *et al.* Ambulatory blood pressure of adults in Ohasama, Japan. *Hypertension* **22.** 900-912, (1993).

44   Tamaki, S. *et al.* Polymorphism of the angiotensin converting enzyme gene and blood pressure in a Japanese general population (the Shigaraki Study). *Hypertens Res* **25.** 843-848, (2002).

45   Yoshida, M. *et al.* Incidence of acute myocardial infarction in Takashima, Shiga, Japan. *Circ J* **69.** 404-408, (2005).

46   Mannami, T., Konishi, M., Baba, S., Nishi, N. & Terao, A. Prevalence of asymptomatic carotid atherosclerotic lesions detected by high-resolution ultrasonography and its relation to cardiovascular risk factors in the general population of a Japanese city: the Suita study. *Stroke* **28.** 518-525, (1997).

47   Tabara, Y. *et al.* Replication study of candidate genes associated with type 2 diabetes based on genome-wide screening. *Diabetes* **58.** 493-498, (2009).

48   Jin, J. J. *et al.* Association of endothelin-1 gene variant with hypertension. *Hypertension* **41.** 163-167, (2003).

49   Hong, K. W. *et al.* Genetic variations in ATP2B1, CSK, ARSG and CSMD1 loci are related to blood pressure and/or hypertension in two Korean cohorts. *J Hum Hypertens* **24.** 367-372, (2010).

50   Heid, I. M. *et al.* Association of the 103I MC4R allele with decreased body mass in 7937 participants of two population based surveys. *J Med Genet* **42.** e21, (2005).

51   Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1.** S26-30, (2005).

52   Polasek, O. *et al.* Genome-wide association study of anthropometric traits in Korcula Island, Croatia. *Croat Med J* **50.** 7-16, (2009).

53   Chambers, J. C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* **40.** 716-718, (2008).

54   Kooner, J. S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* **40.** 149-151, (2008).

55   Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41.** 666-676, (2009).

56   Zia, E. *et al.* Blood pressure in relation to the incidence of cerebral infarction and intracerebral hemorrhage. Hypertensive hemorrhage: debated nomenclature is still relevant. *Stroke* **38.** 2681-2685, (2007).

57   Kramer, H. *et al.* Racial/ethnic differences in hypertension and hypertension treatment and control in the multi-ethnic study of atherosclerosis (MESA). *Am J Hypertens* **17.** 963-970, (2004).

58   Bild, D. E. *et al.* Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* **156.** 871-881, (2002).

59   Stancakova, A. *et al.* Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58.** 1212-1221, (2009).

60   Berglund, G. *et al.* Cardiovascular risk groups and mortality in an urban swedish male population: the Malmo Preventive Project. *J Intern Med* **239.** 489-497, (1996).

61   Wadsworth, M., Kuh, D., Richards, M. & Hardy, R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int J Epidemiol* **35.** 49-54, (2006).

62   Jarvelin, M. R. *et al.* Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension* **44.** 838-846, (2004).

63   Cooper, R. *et al.* The prevalence of hypertension in seven populations of west African origin. *Am J Public Health* **87.** 160-168, (1997).

64   Cooper, J. A. *et al.* Comparison of novel hemostatic factors and conventional risk factors for prediction of coronary heart disease. *Circulation* **102.** 2816-2822, (2000).

65   Pinto-Sietsma, S. J. *et al.* Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol* **11.** 1882-1888, (2000).

66   Hillege, H. L. *et al.* Urinary albumin excretion predicts cardiovascular and noncardiovascular mortality in general population. *Circulation* **106.** 1777-1782, (2002).

67   Broadbent, H. M. *et al.* Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet* **17.** 806-814, (2008).

68   Boker, L. K. *et al.* Prospect-EPIC Utrecht: study design and characteristics of the cohort population. European Prospective Investigation into Cancer and Nutrition. *Eur J Epidemiol* **17.** 1047-1053, (2001).

69   Hofman, A. *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol* **22.** 819-829, (2007).

70   Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* **2.** e132, (2006).

71   John, U. *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Praventivmed* **46.** 186-194, (2001).

72   Foong, A. W. *et al.* Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol* **14.** 25-35, (2007).

73   Lim, L. S. *et al.* Distribution and determinants of ocular biometric parameters in an Asian population: the Singapore Malay eye study. *Invest Ophthalmol Vis Sci* **51.** 103-109, (2010).

74   Nang, E. E. *et al.* Is there a clear threshold for fasting plasma glucose that differentiates between those with and without neuropathy and chronic kidney disease?: the Singapore Prospective Study Program. *Am J Epidemiol* **169.** 1454-1462, (2009).

75   Rudan, I. *et al.* "10001 Dalmatians:" Croatia launches its national biobank. *Croat Med J* **50.** 4-6, (2009).

76   Hercberg, S. *et al.* Background and rationale behind the SU.VI.MAX Study, a prevention trial using nutritional doses of a combination of antioxidant vitamins and minerals to reduce cardiovascular diseases and cancers. SUpplementation en VItamines et Mineraux AntioXydants Study. *Int J Vitam Nutr Res* **68.** 3-20, (1998).

77   Vitart, V. *et al.* 3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia. *Eur J Hum Genet* **14.** 478-487, (2006).

78   Lewington, S., Clarke, R., Qizilbash, N., Peto, R. & Collins, R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360.** 1903-1913, (2002).

79   Colditz, G. A. *et al.* Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol* **123.** 894-900, (1986).

80   Conen, D., Ridker, P. M., Buring, J. E. & Glynn, R. J. Risk of cardiovascular events among women with high normal blood pressure or blood pressure progression: prospective cohort study. *BMJ* **335.** 432, (2007).

81   Cui, J. S., Hopper, J. L. & Harrap, S. B. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension* **41.** 207-210, (2003).

82   Charchar, F. J. *et al.* Association of the human Y chromosome with cholesterol levels in the general population. *Arterioscler Thromb Vasc Biol* **24.** 308-312, (2004).

83   Tobin, M. D., Sheehan, N. A., Scurrah, K. J. & Burton, P. R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* **24.** 2911-2935, (2005).

84   Xie, R. & Stram, D. O. Asymptotic equivalence between two score tests for haplotype-specific risk in general linear models. *Genet Epidemiol* **29.** 166-170, (2005).

85   Kutalik, Z. *et al.* Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics*, (2010).

86   Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38.** 904-909, (2006).

87   Loos, R. J. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* **40.** 768-775, (2008).

88   Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316.** 889-894, (2007).

89   Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41.** 25-34, (2009).

90   Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55.** 997-1004, (1999).

91   Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41.** 677-687, (2009).

92   Johnson, A. D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24.** 2938-2939, (2008).

93   Goring, H. H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet* **39.** 1208-1216, (2007).

94   Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat Genet* **42.** 62-67, (2010).

95   Heap, G. A. *et al.* Complex nature of SNP genotype effects on gene expression in primary human leucocytes. *BMC Med Genomics* **2.** 1, (2009).

96   Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat Genet* **39.** 1202-1207, (2007).

97   Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat Genet* **39.** 1217-1224, (2007).

98   Kwan, T. *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* **40.** 225-231, (2008).

99   Heinzen, E. L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* **6.** e1, (2008).

100   Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452.** 423-428, (2008).

101   Inouye, M. *et al.* Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* **6.** 441, (2010).

102   Webster, J. A. *et al.* Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* **84.** 445-458, (2009).

103   Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6.** e107, (2008).

104   Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res* **19.** 1942-1952, (2009).

105   Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325.** 1246-1250, (2009).

106   Freedman, J. E. *et al.* Relation of platelet and leukocyte inflammatory transcripts to body mass index in the Framingham heart study. *Circulation* **122.** 119-129, (2010).

107   Newton-Cheh, C. *et al.* Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. *Nat Genet* **41.** 348-353, (2009).

108   Smith, E. N. *et al.* Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet* **6.** (2010).

109   Inouye, M. *et al.* An immune response network associated with blood lipid levels. *PLoS Genet* **6.** (2010).

110   Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134.** 1781-1785, (2009).

111   Liebisch, G., Drobnik, W., Lieser, B. & Schmitz, G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin Chem* **48.** 2217-2224, (2002).

112   Liebisch, G., Lieser, B., Rathenberg, J., Drobnik, W. & Schmitz, G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim Biophys Acta* **1686.** 108-117, (2004).

113   Brugger, B., Erben, G., Sandhoff, R., Wieland, F. T. & Lehmann, W. D. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc Natl Acad Sci U S A* **94.** 2339-2344, (1997).

114   Zemski Berry, K. A. & Murphy, R. C. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J Am Soc Mass Spectrom* **15.** 1499-1508, (2004).

115   Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23.** 1294-1296, (2007).

116   Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30.** 97-101, (2002).

117   Amin, N., van Duijn, C. M. & Aulchenko, Y. S. A genomic background based method for association analysis in related individuals. *PLoS One* **2.** e1274, (2007).

118   Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26.** 2190-2191, (2010).

119   McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* **40.** 1166-1174, (2008).

120   Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464.** 704-712, (2010).

121   Craddock, N. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464.** 713-720, (2010).

122   Segre, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6.** (2010).

123   Caulfield, M. *et al.* Genome-wide mapping of human loci for essential hypertension. *Lancet* **361.** 2118-2123, (2003).

124   Smith, N. L. *et al.* Association of genome-wide variation with the risk of incident heart failure in adults of European and African ancestry: a prospective meta-analysis from the cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* **3.** 256-266, (2010).

125   Vasan, R. S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* **302.** 168-178, (2009).

126   Ikram, M. A. *et al.* Genomewide association studies of stroke. *N Engl J Med* **360.** 1718-1728, (2009).

127   Yadav, S. *et al.* Bio-Repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet* **12.** 34, (2011).

128   Meschia, J. F. *et al.* The Siblings With Ischemic Stroke Study (SWISS) protocol. *BMC Med Genet* **3.** 1, (2002).

129   Meschia, J. F. *et al.* The Ischemic Stroke Genetics Study (ISGS) Protocol. *BMC Neurol* **3.** 4, (2003).

130   Wheeler, H. E. *et al.* Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. *PLoS Genet* **5.** e1000685, (2009).

131   Samani, N. J. *et al.* Large scale association analysis of novel genetic loci for coronary artery disease. *Arterioscler Thromb Vasc Biol* **29.** 774-780, (2009).

132   Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43.** 333-338, (2011).

133   Consortium, C. A. D. C. D. G. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* **43.** 339-344, (2011).

134   Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42.** 376-384, (2010).

135   Chambers, J. C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* **42.** 373-375, (2010).

136   Psaty, B. M. *et al.* The risk of myocardial infarction associated with the combined use of estrogens and progestins in postmenopausal women. *Arch Intern Med* **154.** 1333-1339, (1994).

137   Psaty, B. M. *et al.* The risk of myocardial infarction associated with antihypertensive drug therapies. *Jama* **274.** 620-625, (1995).

138   Klungel, O. H. *et al.* Antihypertensive drug therapies and the risk of ischemic stroke. *Arch Intern Med* **161.** 37-43, (2001).

139   Ridker, P. M. *et al.* Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin Chem* **54.** 249-255, (2008).

140   Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* **41.** 56-65, (2009).

141   Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42.** 570-575, (2010).

142   Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106.** 9362-9367, (2009).

143   Johnson, A. D. & O'Donnell, C. J. An open access database of genome-wide association results. *BMC Med Genet* **10.** 6, (2009).

144   Frosst, P. *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* **10.** 111-113, (1995).

145   Yan, T. T. *et al.* Homocysteine impaired endothelial function via compromised VEGF/Akt/eNOS signaling (Revised 1). *Clin Exp Pharmacol Physiol,* (2010).

146   Qian, X., Lu, Z., Tan, M., Liu, H. & Lu, D. A meta-analysis of association between C677T polymorphism in the methylenetetrahydrofolate reductase gene and hypertension. *Eur J Hum Genet* **15.** 1239-1245, (2007).

147   Ho, J. E. *et al.* Discovery and replication of novel blood pressure genetic loci in the Women's Genome Health Study. *J Hypertens* **29.** 62-69, (2011).

148   Tomaszewski, M. *et al.* Genetic architecture of ambulatory blood pressure in the general population: insights from cardiovascular gene-centric array. *Hypertension* **56.** 1069-1076, (2010).

149   Virdis, A. *et al.* Effect of hyperhomocystinemia and hypertension on endothelial function in methylenetetrahydrofolate reductase-deficient mice. *Arterioscler Thromb Vasc Biol* **23.** 1352-1357, (2003).

150   John, S. W. *et al.* Genetic decreases in atrial natriuretic peptide and salt-sensitive hypertension. *Science* **267.** 679-681, (1995).

151   Tamura, N. *et al.* Cardiac fibrosis in mice lacking brain natriuretic peptide. *Proc Natl Acad Sci U S A* **97.** 4239-4244, (2000).

152   Ariyaratnam, R. *et al.* Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32,500 individuals. *PLoS Med* **4.** e131, (2007).

153   Casas, J. P., Hingorani, A. D., Bautista, L. E. & Sharma, P. Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Archives of neurology* **61.** 1652-1661, (2004).

154   Greene, N. D., Stanier, P. & Copp, A. J. Genetics of human neural tube defects. *Hum Mol Genet* **18.** R113-129, (2009).

155   Dong, L. M. *et al.* Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *Jama* **299.** 2423-2436, (2008).

156   Benjamin, E. J. *et al.* Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet* **41.** 879-881, (2009).

157   Hazra, A. *et al.* Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. *Hum Mol Genet* **18.** 4677-4687, (2009).

158   Tanaka, T. *et al.* Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet* **84.** 477-482, (2009).

159   Pyle, W. G. *et al.* Actin capping protein: an essential element in protein kinase signaling to the myofilaments. *Circ Res* **90.** 1299-1306, (2002).

160   Koshikawa, S., Nishikimi, T., Inaba, C., Akimoto, K. & Matsuoka, H. Fasudil, a Rho-kinase inhibitor, reverses L-NAME exacerbated severe nephrosclerosis in spontaneously hypertensive rats. *J Hypertens* **26.** 1837-1848, (2008).

161   Pushkin, A. *et al.* Cloning, tissue distribution, genomic organization, and functional characterization of NBC3, a new member of the sodium bicarbonate cotransporter family. *J Biol Chem* **274.** 16569-16575, (1999).

162   Aalkjaer, C., Frische, S., Leipziger, J., Nielsen, S. & Praetorius, J. Sodium coupled bicarbonate transporters in the kidney, an update. *Acta Physiol Scand* **181.** 505-512, (2004).

163   Boedtkjer, E., Praetorius, J. & Aalkjaer, C. NBCn1 (slc4a7) mediates the Na+-dependent bicarbonate transport important for regulation of intracellular pH in mouse vascular smooth muscle cells. *Circ Res* **98.** 515-523, (2006).

164   Bok, D. *et al.* Blindness and auditory impairment caused by loss of the sodium bicarbonate cotransporter NBC3. *Nat Genet* **34.** 313-319, (2003).

165   Whitfield, J. B. *et al.* Genetic Effects on Toxic and Essential Elements in Humans: Arsenic, Cadmium, Copper, Lead, Mercury, Selenium and Zinc in Erythrocytes. *Environ Health Perspect*, (2010).

166   Ahmed, S. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41.** 585-590, (2009).

167   Fears, S. *et al.* Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. *Proc Natl Acad Sci U S A* **93.** 1642-1647, (1996).

168   Soderholm, J., Kobayashi, H., Mathieu, C., Rowley, J. D. & Nucifora, G. The leukemia-associated gene MDS1/EVI1 is a new type of GATA-binding transactivator. *Leukemia* **11.** 352-358, (1997).

169   Vinatzer, U., Taplick, J., Seiser, C., Fonatsch, C. & Wieser, R. The leukaemia-associated transcription factors EVI-1 and MDS1/EVI1 repress transcription and interact with histone deacetylase. *Br J Haematol* **114.** 566-573, (2001).

170   Hoyt, P. R. *et al.* The Evi1 proto-oncogene is required at midgestation for neural, heart, and paraxial mesenchyme development. *Mech Dev* **65.** 55-70, (1997).

171   WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447.** 661-678, (2007).

172   Bei, J. X. *et al.* A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet* **42.** 599-603, (2010).

173   Meyer, T. E. *et al.* Genome-wide association studies of serum magnesium, potassium, and sodium concentrations identify six Loci influencing serum magnesium levels. *PLoS Genet* **6.** (2010).

174   Tabara, Y. *et al.* Common variants in the ATP2B1 gene are associated with susceptibility to hypertension: the Japanese Millennium Genome Project. *Hypertension* **56.** 973-980, (2010).

175   Niu, W. *et al.* Confirmation of top polymorphisms in hypertension genome wide association study among Han Chinese. *Clin Chim Acta* **411.** 1491-1495, (2010).

176   He, L., Wang, B., Hay, E. B. & Nebert, D. W. Discovery of ZIP transporters that participate in cadmium damage to testis and kidney. *Toxicology and applied pharmacology* **238.** 250-257, (2009).

177   Dalton, T. P. *et al.* Identification of mouse SLC39A8 as the transporter responsible for cadmium-induced toxicity in the testis. *Proc Natl Acad Sci U S A* **102.** 3401-3406, (2005).

178   Satarug, S., Nishijo, M., Lasker, J. M., Edwards, R. J. & Moore, M. R. Kidney dysfunction and hypertension: role for cadmium, p450 and heme oxygenases? *Tohoku J Exp Med* **208.** 179-202, (2006).

179   Waterworth, D. M. *et al.* Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* **30.** 2264-2276, (2010).

180   Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466.** 707-713, (2010).

181   Kloss, S., Bouloumie, A. & Mulsch, A. Aging and chronic hypertension decrease expression of rat aortic soluble guanylyl cyclase. *Hypertension* **35.** 43-47, (2000).

182   Prasad, V. S., Palaniswamy, C. & Frishman, W. H. Endothelin as a clinical target in the treatment of systemic hypertension. *Cardiol Rev* **17.** 181-191, (2009).

183   Ruetten, H., Zabel, U., Linz, W. & Schmidt, H. H. Downregulation of soluble guanylyl cyclase in young and aging spontaneously hypertensive rats. *Circ Res* **85.** 534-541, (1999).

184   Friebe, A., Mergia, E., Dangel, O., Lange, A. & Koesling, D. Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylyl cyclase. *Proc Natl Acad Sci U S A* **104.** 7699-7704, (2007).

185   Groneberg, D. *et al.* Smooth muscle-specific deletion of nitric oxide-sensitive guanylyl cyclase is sufficient to induce hypertension in mice. *Circulation* **121.** 401-409, (2010).

186   Schenk, D. B. *et al.* Purification and subunit composition of atrial natriuretic peptide receptor. *Proc Natl Acad Sci U S A* **84.** 1521-1525, (1987).

187   Matsukawa, N. *et al.* The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc Natl Acad Sci U S A* **96.** 7403-7408, (1999).

188   Ahluwalia, A. & Hobbs, A. J. Endothelium-derived C-type natriuretic peptide: more than just a hyperpolarizing factor. *Trends Pharmacol Sci* **26.** 162-167, (2005).

189   Estrada, K. *et al.* A genome-wide association study of northwestern Europeans involves the C-type natriuretic peptide signaling pathway in the etiology of human height variation. *Hum Mol Genet* **18.** 3516-3524, (2009).

190   Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* **5.** e1000445, (2009).

191   Moffatt, P. *et al.* Osteocrin is a specific ligand of the natriuretic Peptide clearance receptor that modulates bone growth. *J Biol Chem* **282.** 36454-36462, (2007).

192   Fox, A. A. *et al.* Natriuretic peptide system gene variants are associated with ventricular dysfunction after coronary artery bypass grafting. *Anesthesiology* **110.** 738-747, (2009).

193   Hesslein, D. G. *et al.* Ebf1-dependent control of the osteoblast and adipocyte lineages. *Bone* **44.** 537-546, (2009).

194   Osaki, E. *et al.* Identification of a novel Sry-related gene and its germ cell-specific expression. *Nucleic Acids Res* **27.** 2503-2510, (1999).

195   Neuner, R., Cousin, H., McCusker, C., Coyne, M. & Alfandari, D. Xenopus ADAM19 is involved in neural, neural crest and muscle development. *Mech Dev* **126.** 240-255, (2009).

196   Johnson, A. D. *et al.* Genome-wide association meta-analysis for total serum bilirubin levels. *Hum Mol Genet* **18.** 2700-2710, (2009).

197   Benyamin, B. *et al.* Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nat Genet* **41.** 1173-1175, (2009).

198   Benyamin, B. *et al.* Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. *Am J Hum Genet* **84.** 60-65, (2009).

199   Chambers, J. C. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet* **41.** 1170-1172, (2009).

200   Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* **41.** 1191-1198, (2009).

201   Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* **13.** 399-408, (1996).

202   Ellervik, C., Tybjaerg-Hansen, A., Appleyard, M., Ibsen, H. & Nordestgaard, B. G. Haemochromatosis genotype and iron overload: association with hypertension and left ventricular hypertrophy. *J Intern Med*, (2010).

203   Levy, J. E., Montross, L. K. & Andrews, N. C. Genes that modify the hemochromatosis phenotype in mice. *J Clin Invest* **105.** 1209-1216, (2000).

204   Moczulski, D. K., Grzeszczak, W. & Gawlik, B. Role of hemochromatosis C282Y and H63D mutations in HFE gene in development of type 2 diabetes and diabetic nephropathy. *Diabetes Care* **24.** 1187-1191, (2001).

205   Dorak, M. T. *et al.* Conserved extended haplotypes of the major histocompatibility complex: further characterization. *Genes Immun* **7.** 450-467, (2006).

206   Kwak, J. H., Kim, S. I., Kim, J. K. & Choi, M. E. BAT3 interacts with transforming growth factor-beta (TGF-beta) receptors and enhances TGF-beta1-induced type I collagen expression in mesangial cells. *J Biol Chem* **283.** 19816-19825, (2008).

207   Wang, Y. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* **40.** 1407-1409, (2008).

208   Harney, S. M. *et al.* Fine mapping of the MHC Class III region demonstrates association of AIF1 and rheumatoid arthritis. *Rheumatology (Oxford)* **47.** 1761-1767, (2008).

209   Diakite, M. *et al.* A genetic association study in the Gambia using tagging polymorphisms in the major histocompatibility complex class III region implicates a HLA-B associated transcript 2 polymorphism in severe malaria susceptibility. *Hum Genet* **125.** 105-109, (2009).

210   Luke, M. M. *et al.* Gene variants associated with ischemic stroke: the cardiovascular health study. *Stroke* **40.** 363-368, (2009).

211   Hakonarson, H. *et al.* A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448.** 591-594, (2007).

212   Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* **41.** 18-24, (2009).

213   Landi, M. T. *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* **85.** 679-691, (2009).

214   Broderick, P. *et al.* Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res* **69.** 6633-6641, (2009).

215   Hafler, D. A. *et al.* Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* **357.** 851-862, (2007).

216   Plenge, R. M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* **39.** 1477-1482, (2007).

217   Lao, Q. Z., Kobrinsky, E., Harry, J. B., Ravindran, A. & Soldatov, N. M. New Determinant for the CaVbeta2 subunit modulation of the CaV1.2 calcium channel. *J Biol Chem* **283.** 15577-15588, (2008).

218   Neef, J. *et al.* The Ca2+ channel subunit beta2 regulates Ca2+ channel abundance and function in inner hair cells and is required for hearing. *J Neurosci* **29.** 10730-10740, (2009).

219   Antzelevitch, C. *et al.* Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. *Circulation* **115.** 442-449, (2007).

220   Melzer, D. *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* **4.** e1000072, (2008).

221   Zhou, W. & Hildebrandt, F. Molecular cloning and expression of phospholipase C epsilon 1 in zebrafish. *Gene Expr Patterns* **9.** 282-288, (2009).

222   Hinkes, B. *et al.* Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nat Genet* **38.** 1397-1405, (2006).

223   Gbadegesin, R. *et al.* Mutations in PLCE1 are a major cause of isolated diffuse mesangial sclerosis (IDMS). *Nephrol Dial Transplant* **23.** 1291-1297, (2008).

224   Wang, L. D. *et al.* Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat Genet* **42.** 759-763, (2010).

225   Abnet, C. C. *et al.* A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* **42.** 764-767, (2010).

226   Martin, R. M. *et al.* P450c17 deficiency in Brazilian patients: biochemical diagnosis through progesterone levels confirmed by CYP17 genotyping. *J Clin Endocrinol Metab* **88.** 5739-5746, (2003).

227   Geller, D. H., Auchus, R. J., Mendonca, B. B. & Miller, W. L. The genetic and functional basis of isolated 17,20-lyase deficiency. *Nat Genet* **17.** 201-205, (1997).

228   Schlawicke Engstrom, K. *et al.* Arsenic metabolism is influenced by polymorphisms in genes involved in one-carbon metabolism and reduction reactions. *Mutation research* **667.** 4-14, (2009).

229   Cifuentes, F. *et al.* Chronic exposure to arsenic in tap water reduces acetylcholine-induced relaxation in the aorta and increases oxidative stress in female rats. *International journal of toxicology* **28.** 534-541, (2009).

230   Yokohira, M. *et al.* Severe systemic toxicity and urinary bladder cytotoxicity and regenerative hyperplasia induced by arsenite in arsenic (+3 oxidation state) methyltransferase knockout mice. A preliminary report. *Toxicology and applied pharmacology* **246.** 1-7, (2010).

231   Yasuno, K. *et al.* Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat Genet* **42.** 420-425, (2010).

232   Levy, D. *et al.* Associations of plasma natriuretic peptide, adrenomedullin, and homocysteine levels with alterations in arterial stiffness: the Framingham Heart Study. *Circulation* **115.** 3079-3085, (2007).

233   Al-Omari, M. A. *et al.* Mid-regional pro-adrenomedullin is associated with pulse pressure, left ventricular mass, and albuminuria in African Americans with hypertension. *Am J Hypertens* **22.** 860-866, (2009).

234   Shindo, T. *et al.* Vascular abnormalities and elevated blood pressure in mice lacking adrenomedullin gene. *Circulation* **104.** 1964-1971, (2001).

235   Caron, K., Hagaman, J., Nishikimi, T., Kim, H. S. & Smithies, O. Adrenomedullin gene expression differences in mice do not affect blood pressure but modulate hypertension-induced pathology in males. *Proc Natl Acad Sci U S A* **104.** 3420-3425, (2007).

236　Chasman, D. I. *et al.* Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet* **5.** e1000730, (2009).

237　Lei, S. F. *et al.* Genome-wide association study identifies two novel loci containing FLNB and SBF2 genes underlying stature variation. *Hum Mol Genet* **18.** 1661-1669, (2009).

238　Huang, J. *et al.* Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry* **167.** 1254-1263, (2010).

239　Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* **41.** 527-534, (2009).

240　Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*, (2009).

241　Gudbjartsson, D. F. *et al.* Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* **41.** 342-347, (2009).

242　Hunt, K. A. *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* **40.** 395-402, (2008).

243　Ikram, M. K. *et al.* Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. *PLoS Genet* **6.** e1001184, (2010).

244　Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106.** 709-721, (2001).

245　Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15.** 21-29, (1997).

246　Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nat Genet* **42.** 153-159, (2010).

247　Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* **42.** 1068-1076, (2010).

248　Buch, S. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet* **39.** 995-999, (2007).

249　Palatini, P. *et al.* CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. *J Hypertens* **27.** 1594-1601, (2009).

250　Lowry, W. E. *et al.* Csk, a critical link of g protein signals to actin cytoskeletal reorganization. *Developmental cell* **2.** 733-744, (2002).

251   Tripodi, G. *et al.* Hypertension-associated point mutations in the adducin alpha and beta subunits affect actin cytoskeleton and ion transport. *J Clin Invest* **97.** 2815-2822, (1996).

252   Kopf, P. G. *et al.* Cytochrome P4501A1 is Required for Vascular Dysfunction and Hypertension Induced by 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicol Sci*, (2010).

253   Takebe, A. *et al.* Microarray analysis of PDGFR alpha+ populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including ARID3b that is essential for development of embryonic mesenchymal cells. *Developmental biology* **293.** 25-37, (2006).

254   Corchero, J., Pimprale, S., Kimura, S. & Gonzalez, F. J. Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. *Pharmacogenetics* **11.** 1-6, (2001).

255   Cornelis, M. C., El-Sohemy, A., Kabagambe, E. K. & Campos, H. Coffee, CYP1A2 genotype, and risk of myocardial infarction. *Jama* **295.** 1135-1141, (2006).

256   Gagliardo, B. *et al.* Pro-hepcidin is unable to degrade the iron exporter ferroportin unless maturated by a furin-dependent process. *J Hepatol* **50.** 394-401, (2009).

257   Cousin, C. *et al.* Soluble form of the (pro)renin receptor generated by intracellular cleavage by furin is secreted in plasma. *Hypertension* **53.** 1077-1082, (2009).

258   Harris, M., Garcia-Caballero, A., Stutts, M. J., Firsov, D. & Rossier, B. C. Preferential assembly of epithelial sodium channel (ENaC) subunits in Xenopus oocytes: role of furin-mediated endogenous proteolysis. *J Biol Chem* **283.** 7455-7463, (2008).

259   Li, N. *et al.* Associations between genetic variations in the FURIN gene and hypertension. *BMC Med Genet* **11.** 124, (2010).

260   Lesch, K. P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm* **115.** 1573-1585, (2008).

261   Bui, T. D., Levy, E. R., Subramaniam, V. N., Lowe, S. L. & Hong, W. cDNA characterization and chromosomal mapping of human golgi SNARE GS27 and GS28 to chromosome 17. *Genomics* **57.** 285-288, (1999).

262   Levy, D. *et al.* Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension* **36.** 477-483, (2000).

263   Julier, C. *et al.* Genetic susceptibility for human familial essential hypertension in a region of homology with blood pressure linkage on rat chromosome 10. *Hum Mol Genet* **6.** 2077-2085, (1997).

264   Meyer, T. E. *et al.* GOSR2 Lys67Arg is associated with hypertension in whites. *Am J Hypertens* **22.** 163-168, (2009).

265   Hilbert, P. *et al.* Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature* **353.** 521-529, (1991).

266   Jacob, H. J. *et al.* Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **67.** 213-224, (1991).

267   Kumar, R. *et al.* ZNF652, a novel zinc finger protein, interacts with the putative breast tumor suppressor CBFA2T3 to repress transcription. *Mol Cancer Res* **4.** 655-665, (2006).

268   Kimura, K. *et al.* Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* **16.** 55-65, (2006).

269   High, F. A. *et al.* Endothelial expression of the Notch ligand Jagged1 is required for vascular smooth muscle development. *Proc Natl Acad Sci U S A* **105.** 1955-1959, (2008).

270   LaBrecque, D. R. *et al.* Four generations of arteriohepatic dysplasia. *Hepatology* **2.** 467-474, (1982).

271   Harendza, S. *et al.* Renal failure and hypertension in Alagille syndrome with a novel JAG1 mutation. *J Nephrol* **18.** 312-317, (2005).

272   Wettschureck, N. & Offermanns, S. Mammalian G proteins and their cell type specific functions. *Physiol Rev* **85.** 1159-1204, (2005).

273   Cicila, G. T., Choi, C., Dene, H., Lee, S. J. & Rapp, J. P. Two blood pressure/cardiac mass quantitative trait loci on chromosome 3 in Dahl rats. *Mamm Genome* **10.** 112-116, (1999).

274   Makita, T., Sucov, H. M., Gariepy, C. E., Yanagisawa, M. & Ginty, D. D. Endothelins are vascular-derived axonal guidance cues for developing sympathetic neurons. *Nature* **452.** 759-763, (2008).

275   Amiel, J. *et al.* Hirschsprung disease, associated syndromes and genetics: a review. *J Med Genet* **45.** 1-14, (2008).

276   Wiesmann, F. *et al.* Frequent loss of endothelin-3 (EDN3) expression due to epigenetic inactivation in human breast cancer. *Breast Cancer Res* **11.** R34, (2009).

277   Montani, D. *et al.* Endothelin-1/endothelin-3 ratio: a potential prognostic factor of pulmonary arterial hypertension. *Chest* **131.** 101-108, (2007).

278   Aberg, K. *et al.* Genomewide association study of movement-related adverse antipsychotic effects. *Biological psychiatry* **67.** 279-282, (2010).

19. Author contributions

ICBP-GWAS Steering Committee
Gonçalo R Abecasis, Murielle Bochud, Michael Boehnke, Mark J. Caulfield (co-chair), Aravinda Chakravarti, Georg B. Ehret, Paul Elliott, Tamara B. Harris, Marjo Riitta Järvelin, Andrew D. Johnson, Toby Johnson (co-chair), Martin G. Larson, Lenore Launer, Daniel Levy (co-chair), Patricia B. Munroe (co-chair), Christopher Newton-Cheh (co-chair), Bruce M. Psaty, Kenneth M. Rice, Albert V. Smith, Martin D. Tobin, Cornelia M. van Duijn, Germaine C. Verwoert

ICBP-GWAS Writing Group (drafted and edited manuscript)
Gonçalo R Abecasis, Mark J. Caulfield (co-chair), Aravinda Chakravarti, Georg B. Ehret, Paul Elliott, Andrew D. Johnson, Toby Johnson (co-chair), Daniel Levy (co-chair), Patricia B. Munroe (co-chair), Christopher Newton- Cheh (co-chair), Kenneth M. Rice, Albert V. Smith, Cornelia M. van Duijn

Analyses
Central data QA&QC and meta-analyses: Georg B. Ehret, Toby Johnson
Risk scores and total number of variants: Toby Johnson
Central analysis of non-European ancestry cohort data: Georg B. Ehret, Toby Johnson
Gene expression (eSNP) analyses: Andrew D. Johnson, Daniel Levy
Non-synonymous SNP lookups: Gonçalo R. Abecasis
Sex- and BMI-interaction analyses: Georg B. Ehret, Toby Johnson, Vasyl Pihur, Nick RG Shrine
Copy number variant (CNV) analyses: Toby Johnson
MAGENTA analyses: Albert V. Smith
Figures and Tables: Georg B. Ehret, Toby Johnson, Patricia B. Munroe
Literature review and database lookups: Murielle Bochud, Mark J. Caulfield, Aravinda Chakravarti, Georg B. Ehret, Paul Elliott, Andrew D. Johnson, Toby Johnson, Daniel Levy, Patricia B. Munroe, Christopher Newton-Cheh, Louise V. Wain
Statistical review: Gonçalo R. Abecasis, Aravinda Chakravarti, Martin G. Larson, Christopher Newton-Cheh, Ken M. Rice

Cohort contributions

Age, Gene/Environment Susceptibility study.   Study concept/design: A.V.S., L.L., T.B.H., T.A., V.G. Phenotype data acquisition/QC: L.L., T.B.H., T.A., V.G. Genotype data acquisition/QC: A.V.S., M.G. Data analysis: A.V.S., T.A.

Amish studies.   Study concept/design: A.R.S. Phenotype data acquisition/QC: A.Parsa, A.R.S., N.I.S., Y.-P.C. Genotype data acquisition/QC: J.R.O'C. Data analysis: A.Parsa, J.R.O'C.

ARYA study.   Study concept/design: C.S.P.M.U., D.E.G., M.L.B. Phenotype data acquisition/QC: C.S.P.M.U., D.E.G., M.L.B. Data analysis: C.S.P.M.U., D.E.G., M.L.B.

Atherosclerosis Risk In Communities study.   Study concept/design: A.C., G.B.E., E.Boerwinkle, J.C., S.K.G. Phenotype data acquisition/QC: A.C., E.Boerwinkle, G.B.E., A.C.M., A.K., J.C., S.K.G., W.H.L.K. Genotype data acquisition/QC: A.C., G.B.E., D.E.A., E.Boerwinkle, A.K., J.C., S.K.G., W.H.L.K. Data analysis: G.B.E., A.C.M., A.C., D.C.R., E.Boerwinkle, G.S., J.C., W.H.L.K., Y.V.S.

Baltimore Longitudinal Study of Ageing.   Study concept/design: L.F. Phenotype data acquisition/QC: S.N. Genotype data acquisition/QC: D.Hernandez Data analysis: T.Tanaka

British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium. Study concept/design: D.P.S. Phenotype data acquisition/QC: D.P.S. Genotype data acquisition/QC: S.Heath, W.L.McA. Data analysis: D.P.S.

British 1958 Birth Cohort – Wellcome Trust Case Control Consortium. Study concept/design: D.P.S. Phenotype data acquisition/QC: D.P.S. Genotype data acquisition/QC: W.L.McA. Data analysis: D.Hadley, D.P.S.

BRItish Genetics of HyperTension study. Study concept/design: A.F.D., J.M.C., M.J.C., M.F., M.J.B., N.J.S., P.B.M. Phenotype data acquisition/QC: A.F.D., J.M.C., M.J.C., M.J.B., N.J.S., P.B.M. Genotype data acquisition/QC: M.J.C., P.B.M., P.H., S.S.-H., T.J. Data analysis: P.B.M., T.J.

British Regional Heart Study. Study concept/design: P.H.W., R.W.M. Phenotype data acquisition/QC: P.H.W., R.W.M. Genotype data acquisition/QC: A.D.H., P.H.W., R.W.M. Data analysis: R.W.M.

British Women's Heart and Health Study. Study concept/design: D.A.L., J.P.C. Phenotype data acquisition/QC: D.A.L., J.P.C. Genotype data acquisition/QC: D.A.L., I.N.M.D., T.R.G. Data analysis: D.A.L., T.R.G.

Busselton Health Study. Study concept/design: L.J.P. Phenotype data acquisition/QC: J.P.B. Genotype data acquisition/QC: J.P.B., R.W.L. Data analysis: M.N.C., R.W.L.

C4D consortium. Study concept/design: R.Clarke, R.Collins Phenotype data acquisition/QC: J.C.H., R.Clarke, R.Collins Genotype data acquisition/QC: H.O., J.C.H. Data analysis: H.O., J.C.H.

Candidate-gene Association REsource. Phenotype data acquisition/QC: A.W.D., E.R.F., J.H.Y., X.Z., Y.Li Data analysis: X.Z., Y.Li

Cardiovascular Health Study. Study concept/design: B.M.P., J.I.R., J.C.B., K.M.R., N.L.G. Phenotype data acquisition/QC: B.M.P. Genotype data acquisition/QC: J.I.R., J.C.B., N.L.G. Data analysis: J.C.B., K.M.R., N.L.G., X.G.

Cardiovascular risk in Young Finns Study. Study concept/design: J.V., M.Kähönen, O.R., T.L. Phenotype data acquisition/QC: J.V., M.Kähönen, O.R., T.L. Genotype data acquisition/QC: J.V., M.Kähönen, O.R., T.L. Data analysis: O.R., T.L.

Cebu Longitudinal Health and Nutrition Survey. Study concept/design: K.L.M., L.S.A. Phenotype data acquisition/QC: L.S.A., N.R.L. Genotype data acquisition/QC: K.L.M. Data analysis: Y.W.

CHARGE-Heart Failure Working Group. For complete acknowledgments see Smith NL et al. Circ Cardiovasc Genet. 2010 Jun 1;3(3):256-66.

CKDGen consortium. Study concept/design: CKDGen Consortium Phenotype data acquisition/QC: CKDGen Consortium Genotype data acquisition/QC: CKDGen Consortium Data analysis: A.Teumer, C.P., M.Olden, M.H.C., CKDGen Consortium

CLUE study. Study concept/design: A.C., J.C., K.-D.H.N., S.K.G., W.H.L.K. Phenotype data acquisition/QC: A.K., A.C., J.C., J.H.B., K.-D.H.N., S.K.G., W.H.L.K. Genotype data acquisition/QC: A.K., A.C., G.B.E., G.H., J.C., J.H.B., K.-D.H.N., S.K.G., W.H.L.K. Data analysis: A.C., G.B.E., K.-D.H.N., S.K.G., W.H.L.K.

Cohorte Lausannoise. Study concept/design: P.Vollenweider, V.M. Phenotype data acquisition/QC: M.Bochud, P.Vollenweider, V.M. Genotype data acquisition/QC: P.Vollenweider, V.M. Data analysis: J.S.B., M.Bochud, S.Bergmann, T.J.

Control of Blood Pressure and Risk Attenuation study. Phenotype data acquisition/QC: N.C., T.H.J. Genotype data acquisition/QC: T.H.J., T.M.F. Data analysis: M.I., T.H.J.

Coronary ARtery Disease Genome-wide Replication And Meta-analysis consortium. Study concept/design: CARDIoGRAM consortium Phenotype data acquisition/QC: CARDIoGRAM

consortium Genotype data acquisition/QC: CARDIoGRAM consortium Data analysis: J.E., CARDIoGRAM consortium

Coronary Risk of Insulin Sensitivity in Indian Subjects study.   Phenotype data acquisition/QC: S.R.K. Genotype data acquisition/QC: G.R.C. Data analysis: G.R.C.

DIAbetes GENetic study.   Study concept/design: B.L., J.G., P.E.H.S., S.R.B. Phenotype data acquisition/QC: P.E.H.S. Data analysis: P.E.H.S.

Diabetes Genetics Initiative.   Study concept/design: C.N.-C., D.A., L.G. Phenotype data acquisition/QC: C.N.-C., L.G. Genotype data acquisition/QC: B.F.V., D.A. Data analysis: B.F.V., C.N.-C., P.A.

DILGOM expression/metabonomics.   Phenotype data acquisition/QC: J.Kettunen Data analysis: J.Kettunen

EchoGen consortium.   For complete acknowledgments see Vasan RS et al. JAMA. 2009 Jul 8;302(2):168-78.

Edinburgh Artery Study.   Study concept/design: F.G.R.F. Phenotype data acquisition/QC: F.G.R.F. Genotype data acquisition/QC: F.G.R.F., P.H. Data analysis: I.T., P.H.

English Longitudinal Study of Ageing.   Phenotype data acquisition/QC: M.Kumari, M.G.M. Genotype data acquisition/QC: A.Taylor, M.Kumari

EPIC Turin study.   Study concept/design: G.M., P.Vineis Phenotype data acquisition/QC: F.Ricceri Genotype data acquisition/QC: S.G. Data analysis: F.Ricceri

eQTL lookup.   Study concept/design: A.Plump, V.E. Phenotype data acquisition/QC: D.L. Genotype data acquisition/QC: A.Plump, V.E. Data analysis: A.Plump, V.E. Database design and cross-dataset aggregation: A.D.J.

ERF study (EUROSPAN).   Study concept/design: A.Hofman, B.A.O., C.M.vanD., J.C.M.W. Phenotype data acquisition/QC: C.M.vanD., E.J.G.S., F.M.-R., G.C.V., J.C.M.W. Genotype data acquisition/QC: A.G.U., F.Rivadeneira Data analysis: A.Dehghan, C.M.vanD., G.C.V., N.A., Y.A.

European Prospective Investigation of Cancer Norfolk study.   Study concept/design: K.-T.K., N.J.W. Phenotype data acquisition/QC: K.-T.K., N.J.W. Genotype data acquisition/QC: I.B., J.H.Z., N.J.W., R.J.F.L. Data analysis: J.H.Z., R.J.F.L.

Family Blood Pressure Program, GenNet study.   Study concept/design: A.B.W., A.C., G.B.E. Phenotype data acquisition/QC: A.B.W., A.C., G.B.E., K.-D.H.N., S.K.G. Genotype data acquisition/QC: A.C., G.B.E., G.H., K.-D.H.N., S.K.G. Data analysis: G.B.E., A.C., K.-D.H.N., S.K.G.

Family Blood Pressure Program, Hypertension Genetic Epidemiology Network.   Study concept/design: A.C., D.C.R., G.B.E., S.C.H. Phenotype data acquisition/QC: A.C., D.C.R., G.B.E., K.-D.H.N., S.K.G., S.C.H. Genotype data acquisition/QC: A.C., D.C.R., G.B.E., K.-D.H.N., S.K.G. Data analysis: G.B.E., A.C., D.C.R., G.S., K.-D.H.N., S.K.G., Y.V.S.

Fenland study.   Study concept/design: N.J.W. Phenotype data acquisition/QC: N.J.W. Genotype data acquisition/QC: J.a.L., N.J.W., R.J.F.L. Data analysis: J.a.L., R.J.F.L.

Finland-United States Investigation of NIDDM Genetics study.   Study concept/design: F.S.C., J.Tuomilehto, K.L.M., M.Boehnke, R.N.B. Phenotype data acquisition/QC: J.Tuomilehto Genotype data acquisition/QC: F.S.C. Data analysis: A.U.J.

Finland-United States Investigation of NIDDM Genetics study 2.   Study concept/design: K.L.M., L.J.S. Phenotype data acquisition/QC: H.M.S. Genotype data acquisition/QC: L.L.B.

FINRISK 1997 Study.   Study concept/design: L.P., V.S. Phenotype data acquisition/QC: E.V., V.S. Genotype data acquisition/QC: C.N.-C., L.P., M.P. Data analysis: C.N.-C.

Flemish Study on Environment, Genes and Health Outcomes.   Study concept/design: E.Brand, J.A.S., S.-M.B.-H., T.K. Phenotype data acquisition/QC: J.A.S., T.K. Genotype data acquisition/QC: E.Brand, S.-M.B.-H. Data analysis: J.A.S., T.K.

Framingham Heart Study. Phenotype data acquisition/QC: D.L., M.G.L., R.S.V., S.-J.H., T.J.W. Genotype data acquisition/QC: M.G.L., S.-J.H. Data analysis: M.G.L., S.-J.H.

Genetic Regulation of Arterial Pressure of Humans in the Community study. Study concept/design: M.D.T., N.J.S., P.R.B., P.S.B. Phenotype data acquisition/QC: M.D.T., N.J.S., P.R.B. Genotype data acquisition/QC: M.T., M.S.A., M.D.T., N.J.S., P.R.B., P.S.B. Data analysis: M.T., M.S.A., M.D.T., N.J.S.

Georgia Cohorts. Study concept/design: H.S. Phenotype data acquisition/QC: X.W., H.S. Genotype data acquisition/QC: H.Z., Y.D. Data analysis: X.W.

Health ABC study. Study concept/design: T.B.H. Phenotype data acquisition/QC: K.K.L., T.B.H., Y.Liu Genotype data acquisition/QC: K.K.L., M.E.R., T.B.H., Y.Liu Data analysis: K.K.L., Y.Liu

Heart and Vascular Health Study. Study concept/design: N.L.S., S.R.H. Phenotype data acquisition/QC: K.L.W., N.L.S., S.R.H. Genotype data acquisition/QC: N.L.S., S.R.H.

Howard University Family Study. Study concept/design: A.A., C.R. Phenotype data acquisition/QC: A.A., A.Doumatey, C.R. Genotype data acquisition/QC: A.A., A.Doumatey, C.R., D.Shriner Data analysis: A.A., C.R., D.Shriner

HYPertension in ESTonia study. Study concept/design: E.O., M.V., M.Laan Phenotype data acquisition/QC: E.O., G.V., M.V., M.Laan Genotype data acquisition/QC: E.O., M.Laan Data analysis: S.Sõber

Indian Migration Study. Study concept/design: D.P., G.D.S., S.Kinra Phenotype data acquisition/QC: D.P., S.Kinra Genotype data acquisition/QC: D.G.V., G.R.C. Data analysis: G.R.C., V.T.

Insulin Resistance Atherosclerosis Family Study. Study concept/design: C.D.L., J.I.R., L.E.W. Phenotype data acquisition/QC: C.D.L., J.I.R., L.E.W. Genotype data acquisition/QC: C.D.L., N.D.P. Data analysis: C.D.L.

Intergene study. Study concept/design: A.Rosengren, D.S.T. Phenotype data acquisition/QC: A.Rosengren, D.S.T. Genotype data acquisition/QC: F.N., S.D. Data analysis: F.N., S.D.

Invecchiare in Chianti study. Study concept/design: S.Bandinelli, Y.M. Phenotype data acquisition/QC: A.M.C. Genotype data acquisition/QC: A.Singleton

Jamaican individuals from Kingston and Spanish Town & Nigerian study. Study concept/design: B.T., T.F., C.A.McK., R.S.C. Phenotype data acquisition/QC: B.T., T.F., C.A.McK., T.S. Genotype data acquisition/QC: B.T., C.A.McK., C.N.-C., G.H. Data analysis: C.N.-C., B.T.

Japanese Millenium Genome Project. Study concept/design: H.U., N.I., S.U., T.Ohkubo, T.Miki, T.Okamura, T.Ogihara, Y.T., Y.K. Phenotype data acquisition/QC: T.Ohkubo, T.Okamura, Y.T., Y.K. Genotype data acquisition/QC: N.I., T.Miki, Y.T. Data analysis: Y.T.

KidneyGen consortium. For complete acknowledgments see Chambers JC Nat Genet. 2010 May;42(5):373-5

KOoperative Gesundheitsforschung in der Region Augsburg. Study concept/design: C.G., E.O., H.-E.W., M.Laan, T.Meitinger Phenotype data acquisition/QC: H.-E.W. Genotype data acquisition/QC: C.G., E.O., H.-E.W., M.Laan, T.Meitinger Data analysis: S.Sõber, S.E.

Korea Association REsource. Study concept/design: B.G.H., H.-L.K. Phenotype data acquisition/QC: J.-Y.L., Y.S.C. Genotype data acquisition/QC: J.-Y.L., Y.S.C. Data analysis: M.J.G., Y.S.C.

KORKULA study. Study concept/design: A.F.W., C.H., H.C., I.R. Phenotype data acquisition/QC: C.H., I.R., O.P. Genotype data acquisition/QC: C.H., H.C., I.R., O.P., V.V. Data analysis: C.H., O.P., V.V.

LOndon LIfe Sciences POPulation study. Phenotype data acquisition/QC: J.S., J.S.K., J.S.S., J.C.C., P.E. Genotype data acquisition/QC: J.S., J.S.K., J.C.C., P.E., W.Z. Data analysis: J.S., J.S.K., J.C.C., W.Z.

Malmö Diet and Cancer study. Study concept/design: O.M. Phenotype data acquisition/QC: B.H., O.M., P.N. Genotype data acquisition/QC: C.N.-C., C.F., M.Sjögren, O.M. Data analysis: C.N.-C., C.F., M.Sjögren, O.M.

Malmö Preventive Project. Study concept/design: O.M. Phenotype data acquisition/QC: B.H., O.M., P.N. Genotype data acquisition/QC: C.N.-C., C.F., M.Sjögren, O.M. Data analysis: C.N.-C., C.F., M.Sjögren, O.M.

Medical Research Council National Survey of Health and Development. Study concept/design: D.K., R.H. Phenotype data acquisition/QC: D.K., R.H. Genotype data acquisition/QC: A.W., D.K. Data analysis: A.W., R.H.

Metabolic Syndrome in Men study. Study concept/design: M.Laakso Phenotype data acquisition/QC: J.Kuusisto, M.Laakso Genotype data acquisition/QC: A.Stančáková, N.

MICROS study (EUROSPAN). Study concept/design: P.P.P. Phenotype data acquisition/QC: P.P.P. Genotype data acquisition/QC: A.A.H. Data analysis: A.A.H.

Multi-Ethnic Study of Atherosclerosis. Study concept/design: J.I.R., L.J.R., W.P., X.G. Phenotype data acquisition/QC: J.Y., L.J.R., W.P., X.G. Genotype data acquisition/QC: J.I.R., J.Y., L.J.R., X.G. Data analysis: J.Y., W.P., X.G.

Myocardial Infarction Genetics Consortium. Study concept/design: C.O'D., D.A., O.M., R.E., S.Kathiresan, S.M.S., V.S. Phenotype data acquisition/QC: C.O'D., D.A., G.L., O.M., R.E., S.Kathiresan, S.M.S., V.S. Genotype data acquisition/QC: B.F.V., C.O'D., D.A., G.L., O.M., R.E., S.Kathiresan, S.M.S., V.S. Data analysis: G.L.

NEURO-CHARGE consortium. Study concept/design: M.A.I., S.Seshadri, T.H.M., W.T.L. Phenotype data acquisition/QC: M.A.I., S.Seshadri, T.H.M., W.T.L. Data analysis: M.A.I., S.Seshadri

Nord-Trøndelag Health study. Study concept/design: C.G.P.P., K.H. Phenotype data acquisition/QC: C.G.P.P., K.H. Genotype data acquisition/QC: A.J.S.

North Finland Birth Cohort of 1966. Study concept/design: M.R.J., P.Z. Phenotype data acquisition/QC: A.-L.H., J.L., M.R.J., P.Z., P.E. Genotype data acquisition/QC: A.-L.H., I.P., M.R.J., P.Z., P.E. Data analysis: P.F.O'R.

Northern Swedish Population Health Study (EUROSPAN). Study concept/design: U.B.G. Phenotype data acquisition/QC: U.B.G. Data analysis: W.I.

Northwick Park Heart Study II. Study concept/design: S.E.H. Phenotype data acquisition/QC: J.A.C., P.J.T., S.E.H. Genotype data acquisition/QC: J.P., P.J.T. Data analysis: J.A.C.

Orkney Complex Disease Study (EUROSPAN). Study concept/design: J.F.W. Phenotype data acquisition/QC: J.F.W., S.H.W. Genotype data acquisition/QC: J.F.W. Data analysis: J.F.W., S.H.W.

Pakistan Risk of Myocardial Infarction Study. Study concept/design: A.Rasheed, D.Saleheen, J.D. Phenotype data acquisition/QC: A.Rasheed, D.Saleheen, J.D., M.A., P.D. Genotype data acquisition/QC: D.Saleheen, J.D., M.A., P.D. Data analysis: D.Saleheen, J.D., M.A.

Precocious Coronary Artery Disease study. Study concept/design: A.Hamsten, H.W., J.F.P., M.F. Phenotype data acquisition/QC: A.Hamsten, J.F.P. Genotype data acquisition/QC: A.G., J.F.P. Data analysis: A.G., J.F.P., M.F.

Prevention of REnal and Vascular End-stage Disease study. Study concept/design: G.N., P.v.d.H., S.J.L.B., W.H.v.G. Phenotype data acquisition/QC: G.N., S.J.L.B., W.H.v.G. Genotype data acquisition/QC: P.v.d.H. Data analysis: P.v.d.H., S.J.L.B.

Prospect-EPIC. Phenotype data acquisition/QC: N.C.O-M., Y.T.vanderS.

Pune Maternal Nutritional Study.   Study concept/design: C.S.Y. Phenotype data acquisition/QC: C.S.Y. Genotype data acquisition/QC: C.S.J., G.R.C., K.R.M. Data analysis: C.S.J., G.R.C.

Rotterdam Extension Study.   Study concept/design: A.Hofman, B.A.O., C.M.vanD., J.C.M.W. Phenotype data acquisition/QC: C.M.vanD., E.J.G.S., F.M.-R., G.C.V., J.C.M.W. Genotype data acquisition/QC: A.G.U., F.Rivadeneira Data analysis: A.Dehghan, C.M.vanD., G.C.V., N.A., Y.A.

Rotterdam lipidomics.   Phenotype data acquisition/QC: A.I., A.Demirkan Data analysis: A.I., A.Demirkan

Rotterdam Study.   Study concept/design: A.Hofman, B.A.O., C.M.vanD., J.C.M.W. Phenotype data acquisition/QC: C.M.vanD., E.J.G.S., F.M.-R., G.C.V., J.C.M.W. Genotype data acquisition/QC: A.G.U., F.Rivadeneira. Data analysis: A.Dehghan, C.M.vanD., G.C.V., N.A., Y.A.

SardiNIA study.   Study concept/design: E.G.L., G.R.A. Phenotype data acquisition/QC: A.Scuteri, M.Orru Genotype data acquisition/QC: G.R.A., M.U. Data analysis: J.L.B.-G.

Serum NMR metabonomics.   Phenotype data acquisition/QC: M.A.K., P.Soininen, A.J.K., T.Tukiainen. Data analysis: M.A.K., P.Soininen, A.J.K., T.Tukiainen, L.P.L., P.W.

Singapore Malay Eye Study.   Phenotype data acquisition/QC: T.Y.W. Genotype data acquisition/QC: M.Seielstad, R.T-H.O., T.Y.W., X.S. Data analysis: X.S.

Singapore Prospective Study Program.   Study concept/design: E.S.T. Phenotype data acquisition/QC: E.S.T. Genotype data acquisition/QC: M.Seielstad, R.T-H.O., X.S. Data analysis: X.S.

SPLIT study.   Study concept/design: A.F.W., C.H., H.C., I.R. Phenotype data acquisition/QC: C.H., I.R., J.Terzic, O.P. Genotype data acquisition/QC: C.H., H.C., I.R., O.P., V.V. Data analysis: C.H., O.P., V.V.

Study of Health in Pomerania.   Study concept/design: H.V., H.K.K., M.Dörr, R.R., U.V. Phenotype data acquisition/QC: H.V., M.Dörr, R.R. Genotype data acquisition/QC: H.V., H.K.K., U.V. Data analysis: A.Teumer, U.V.

Supplémentation en Vitamines et Minéraux Antioxydants study.   Study concept/design: P.M., P.G., S.Hercberg Phenotype data acquisition/QC: P.G., S.Hercberg Genotype data acquisition/QC: D.Z., M.Lathrop, S.Heath Data analysis: P.M., T.J.

TwinsUK study.   Study concept/design: F.Z., T.D.S. Phenotype data acquisition/QC: F.Z., N.S., P.D., T.D.S. Genotype data acquisition/QC: F.Z., G.Z., M.M., N.S., P.D. Data analysis: F.Z.

UK-US Stroke Collaborative Group.   Study concept/design: L.F., P.Sharma Phenotype data acquisition/QC: J.F.M., P.Sharma Data analysis: J.F.M., L.F., M.A.N., P.Sharma

VIS study (EUROSPAN).   Study concept/design: A.F.W., H.C., I.R. Phenotype data acquisition/QC: C.H., I.R., J.Terzic, O.P. Genotype data acquisition/QC: A.F.W., H.C., I.R.

Wellcome Genetic study.   Genotype data acquisition/QC: G.R.C., M.J.K.K. Data analysis: G.R.C.

Whitehall II study.   Phenotype data acquisition/QC: M.Kumari, M.G.M. Genotype data acquisition/QC: M.Kumari, M.Kivimaki

Women's Genome Health Study.   Study concept/design: D.I.C., P.M.R. Phenotype data acquisition/QC: D.I.C., P.M.R. Genotype data acquisition/QC: A.N.P., D.I.C., P.M.R. Data analysis: D.I.C.

Young Men Cardiovascular Association study.   Study concept/design: E.Z.-S., F.J.C., M.T. Phenotype data acquisition/QC: E.Z.-S., F.J.C., M.T. Genotype data acquisition/QC: M.T., M.Denniff Data analysis: M.T., M.Denniff

20. Consortium members

CARDIoGRAM consortium

Hua Tang, Joshua Knowles, Mark Hlatky, Stephen Fortmann, Themistocles L. Assimes, Thomas Quertermous, Alan Go, Carlos Iribarren, Devin Absher, Neil Risch, Richard Myers, Steven Sidney, Andreas Ziegler, Arne Schillert, Christoph Bickel, Christoph Sinning, Hans J. Rupprecht, Karl Lackner, Philipp Wild, Renate Schnabel, Stefan Blankenberg, Tanja Zeller, Thomas Münzel, Claire Perret, Francois Cambien, Laurence Tiret, Viviane Nicaud, Carole Proust, Abbas Dehghan, Albert Hofman, Andre Uitterlinden, Cornelia van Duijn, Daniel Levy, Jaqueline Whitteman, L. Adrienne Cupples, Serkalem Demissie-Banjaw, Vasan Ramachandran, Albert Smith, Vilmundur Gudnason, Eric Boerwinkle, Aaron Folsom, Alanna Morrison, Bruce M. Psaty, Ida Yii-Der Chen, Jerome I. Rotter, Joshua Bis, Kelly Volcik, Kenneth Rice, Kent D. Taylor, Kristin Marciante, Nicholas Smith, Nicole Glazer, Susan Heckbert, Tamara Harris, Thomas Lumley, Augustine Kong, Gudmar Thorleifsson, Gudmundur Thorgeirsson, Hilma Holm, Jeffrey R. Gulcher, Kari Stefansson, Karl Andersen, Solveig Gretarsdottir, Unnur Thorsteinsdottir, Michael Preuss, Stefan Schreiber, Thomas Meitinger, Inke R. König, Wolfgang Lieb, Christian Hengstenberg, Heribert Schunkert, Jeanette Erdmann, Marcus Fischer, Anika Grosshennig, Anja Medack, Klaus Stark, Patrick Linsel-Nitschke, Petra Bruse, Zouhair Aherrahrou, Annette Peters, Christina Loley, Christina Willenborg, Janja Nahrstedt, Jennifer Freyer, Stephanie Gulde, Angela Doering, Christina Meisinger, H.-Erich Wichmann, Norman Klopp, Thomas Illig, Andreas Meinitzer, Andreas Tomaschitz, Eran Halperin, Harald Dobnig, Hubert Scharnagl, Marcus Kleber, Reijo Laaksonen, Stefan Pilz, Tanja B. Grammer, Tatjana Stojakovic, Wilfried Renner, Winfried März, Bernhard O. Böhm, Bernhard R. Winkelmann, Karl Winkler, Michael M. Hoffmann, Christopher J. O´Donnell, Benjamin F. Voight, David Altshuler, David S. Siscovick, Kiran Musunuru, Leena Peltonen, Maja Barbalic, Olle Melander, Roberto Elosua, Sekar Kathiresan, Stephen M. Schwartz, Veikko Salomaa, Candace Guiducci, Noel Burtt, Stacey B. Gabriel, Alexandre F. R. Stewart, George A. Wells, Li Chen, Olga Jarinova, Robert Roberts, Ruth McPherson, Sonny Dandona, Augusto D. Pichard, Daniel J. Rader, Joe Devaney, Joseph M. Lindsay, Kenneth M. Kent, Liming Qu, Lowell Satler, Mary Susan Burnett, Mingyao Li, Muredach P. Reilly, Robert Wilensky, Ron Waksman, Stephen Epstein, William Matthai, Christopher W. Knouff, Dawn M. Waterworth, Hakon H. Hakonarson, Max C. Walker, Vincent Mooser, Alistair S. Hall, Anthony J. Balmforth, Benjamin J. Wright, Chris Nelson, John R. Thompson, Nilesh J. Samani, Peter S. Braund, Stephen G. Ball

CHARGE-Heart Failure Working Group

Nicholas L.Smith, Janine F.Felix, Alanna C.Morrison, Serkalem Demissie, Nicole L.Glazer, Laura R.Loehr, L.Adrienne Cupples, Abbas Dehghan, Thomas Lumley, Wayne D.Rosamond, Wolfgang Lieb, Fernando Rivadeneira, Joshua C.Bis, Aaron R.Folsom, Emelia Benjamin, Yurii S.Aulchenko, Talin Haritunians, David Couper, Joanne Murabito, Ying A.Wang, Bruno H.Stricker, John S.Gottdiener, Patricia P.Chang, Thomas J.Wang, Kenneth

M.Rice, Albert Hofman, Susan R.Heckbert, Ervin R.Fox, Christopher J.O'Donnell, Andre G.Uitterlinden, Jerome I.Rotter, James T.Willerson, Daniel Levy, Cornelia M.van Duijn, Bruce M.Psaty, Jacqueline C.M.Witteman, Eric Boerwinkle, Ramachandran S.Vasan

CKDGen consortium

Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O'Connell JR, Li M, Schmidt H, Tanaka T, Isaacs A, Ketkar S, Hwang SJ, Johnson AD, Dehghan A, Teumer A, Paré G, Atkinson EJ, Zeller T, Lohman K, Cornelis MC, Probst-Hensch NM, Kronenberg F, Tönjes A, Hayward C, Aspelund T, Eiriksdottir G, Launer LJ, Harris TB, Rampersaud E, Mitchell BD, Arking DE, Boerwinkle E, Struchalin M, Cavalieri M, Singleton A, Giallauria F, Metter J, de Boer IH, Haritunians T, Lumley T, Siscovick D, Psaty BM, Zillikens MC, Oostra BA, Feitosa M, Province M, de Andrade M, Turner ST, Schillert A, Ziegler A, Wild PS, Schnabel RB, Wilde S, Munzel TF, Leak TS, Illig T, Klopp N, Meisinger C, Wichmann HE, Koenig W, Zgaga L, Zemunik T, Kolcic I, Minelli C, Hu FB, Johansson A, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Schreiber S, Aulchenko YS, Felix JF, Rivadeneira F, Uitterlinden AG, Hofman A, Imboden M, Nitsch D, Brandstätter A, Kollerits B, Kedenko L, Mägi R, Stumvoll M, Kovacs P, Boban M, Campbell S, Endlich K, Völzke H, Kroemer HK, Nauck M, Völker U, Polasek O, Vitart V, Badola S, Parker AN, Ridker PM, Kardia SL, Blankenberg S, Liu Y, Curhan GC, Franke A, Rochat T, Paulweber B, Prokopenko I, Wang W, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Shlipak MG, van Duijn CM, Borecki I, Krämer BK, Rudan I, Gyllensten U, Wilson JF, Witteman JC, Pramstaller PP, Rettig R, Hastie N, Chasman DI, Kao WH, Heid IM, Fox CS

EchoGen consortium

Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, Watzinger N, Larson MG, Smith NL, Dehghan A, Grosshennig A, Schillert A, Teumer A, Schmidt R, Kathiresan S, Lumley T, Aulchenko YS, König IR, Zeller T, Homuth G, Struchalin M, Aragam J, Bis JC, Rivadeneira F, Erdmann J, Schnabel RB, Dörr M, Zweiker R, Lind L, Rodeheffer RJ, Greiser KH, Levy D, Haritunians T, Deckers JW, Stritzke J, Lackner KJ, Völker U, Ingelsson E, Kullo I, Haerting J, O'Donnell CJ, Heckbert SR, Stricker BH, Ziegler A, Reffelmann T, Redfield MM, Werdan K, Mitchell GF, Rice K, Arnett DK, Hofman A, Gottdiener JS, Uitterlinden AG, Meitinger T, Blettner M, Friedrich N, Wang TJ, Psaty BM, van Duijn CM, Wichmann HE, Munzel TF, Kroemer HK, Benjamin EJ, Rotter JI, Witteman JC, Schunkert H, Schmidt H, Völzke H, Blankenberg S.

KidneyGen consortium

Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, Beckmann J, Bilo HJ, Bochud M, Brown MJ, Caulfield MJ, Connell JM, Cook HT, Cotlarciuc I, Davey Smith G, de Silva R, Deng G, Devuyst O, Dikkeschei LD, Dimkovic N, Dockrell M, Dominiczak A, Ebrahim S, Eggermann T, Farrall M, Ferrucci L, Floege J, Forouhi NG, Gansevoort RT, Han X, Hedblad B, Homan van der Heide JJ, Hepkema BG, Hernandez-Fuentes M, Hypponen E, Johnson T, de Jong PE, Kleefstra N, Lagou V, Lapsley M, Li Y, Loos RJ, Luan J, Luttropp K, Maréchal C, Melander

O, Munroe PB, Nordfors L, Parsa A, Peltonen L, Penninx BW, Perucha E, Pouta A, Prokopenko I, Roderick PJ, Ruokonen A, Samani NJ, Sanna S, Schalling M, Schlessinger D, Schlieper G, Seelen MA, Shuldiner AR, Sjögren M, Smit JH, Snieder H, Soranzo N, Spector TD, Stenvinkel P, Sternberg MJ, Swaminathan R, Tanaka T, Ubink-Veltmaat LJ, Uda M, Vollenweider P, Wallace C, Waterworth D, Zerres K, Waeber G, Wareham NJ, Maxwell PH, McCarthy MI, Jarvelin MR, Mooser V, Abecasis GR, Lightstone L, Scott J, Navis G, Elliott P, Kooner JS.

21. Acknowledgments

Appendix A. Estimating number of as-yet-undiscovered signals

See the following 16 pages.

Appendix A: Estimating number of as-yet-undiscovered signals

A.1 Introduction

Our approach to estimating the total number of independent signals associated with continuous blood pressure phenotypes, and the total fraction variance explained, is based on the approach described recently by Park *et al.* [2010]. The underlying principle is as follows: For a set of signals detected in a discovery GWAS, data independent of the data used for discovery is free of bias caused by the winners' curse effect, and therefore can be used to consistently estimate the true effect sizes. This allows consistent estimates of the power of the discovery GWAS to detect each of the signals that was detected. For the i-th signal detected, knowing that there was discovery power B_i implies a natural estimate that there are $M_i = 1/B_i$ signals with similar effect sizes, of which by chance one was detected and $M_i - 1$ were not detected. Summing over the set of signals detected gives $M = \sum_i M_i = \sum_i 1/B_i$, which is an estimate of the total number of associated signals with effect sizes similar to those detected.

The estimate $M = \sum_i 1/B_i$ is extremely sensitive to both the number of signals detected for which B_i is small, and also to the numerical estimates of B_i for such signals. Consider, for example, a set of discoveries where the true detection power for the first signal is $B_1 = 0.02$. Then, either (i) using an underestimate $B_1 = 0.01$ for this signal, or (ii) including this signal when it was not in fact discovered by the process that has detection power $B_1 = 0.02$, but was in fact discovered by analysis of e.g. a different phenotype, would both result in an upward bias in M such that an incorrect additional 50 signals would be estimated to exist.

Park *et al.* [2010] did not give any explicit expressions for discovery power B_i, remarking only that the use of standard power calculation tools is sufficient. This applies for a relatively idealised scenario, where a discovery GWAS was followed by a validation study with extremely large sample size, in which validation was attempted for all independent sig-

page A.1 (Appendix A)

nals attaining a pre-specified P-value threshold. More precisely, standard power calculation tools would be adequate for calculating the B_i when it can be assumed that (i) validation genotyping was undertaken if and only if the association P-value, for a single phenotype studied, was smaller than some prespecified constant α in the discovery GWAS, and (ii) the validation genotyping establishes without error the true effect size (and therefore also true/false status and true discovery power) for each signal discovered. Although applying the principle underlying the method of Park *et al.* [2010] does not require these assumptions, it is nonetheless clear that to minimise the source of bias in M described above, it is necessary to undertake realistic modelling of the discovery process that gave rise to the data being analysed. In the following sections, we describe how we estimated discovery power for the ICBP-GWAS study, which (i) analysed two phenotypes, SBP and DBP, in parallel, (ii) did not have a strict P-value threshold below which all signals were taken into the validation stages of the study, and (iii) acquired relatively little validation data for some previously reported signals.

Although both non-parametric and parametric extensions to the basic estimation methodology were proposed by Park *et al.* [2010], these extensions are essentially smoothing procedures that are intended to improve the estimation of the shape of the distribution of effect sizes, and do not directly address the problem of bias in M caused by underestimation of discovery power.

A.2 Effect size estimates free of winners' curse bias

Our estimate of the total number of independent signals associated with blood pressure is based essentially only on the observation that the ICBP-GWAS discovery analysis (stage 0, see Supplementary Figure 1) led to the discovery of 29 associated signals[1] (see Table 1). Discoveries made previously in subsets of the ICBP-GWAS discovery dataset, namely the discoveries of Newton-Cheh *et al.* [2009] and of Levy *et al.* [2009], are relevant only insofar as they allow larger datasets to be declared free of winners' curse bias, and thus allow more precise estimates of true effect size and discovery power.

For signals discovered for the first time in the present study, we excluded all data used in the discovery GWAS meta-analysis (stage 0), and estimated effect sizes free of bias from the winners' curse effect by using data from the validation stages 1–3 only.

A total of thirteen independent signals associated with blood pressure were discovered in previous analyses [Newton-Cheh *et al.* 2009, Levy *et al.* 2009], of which we exclude one (PLCD3; rs12946454) that is likely to have been a false positive. The strength of evidence

---

[1]We define a signal here as a statistically independent associated variant, and thus by this definition there are two associated signals in the CACNB2 gene region.

from previous studies meant that the additional validation data for these signals that was acquired during stages 1–3 of the present study comprised relatively small sample sizes. Using these small sample sizes alone would have yielded effect size estimates with an unnecessary lack of precision. We therefore took into account the historic discovery process, and separately for each association we estimated an effect size using the largest data set that we considered to be free of the winners' curse effect. Specifically, we these were as follows:

Four signals were discovered previously by the Global BPgen consortium, namely MTHFR-NPPB (rs17367504), FGF5 (rs1458038), C10orf107 (rs4590817), and ZNF652 (rs12940887) [Newton-Cheh *et al.* 2009], and we reasoned that a meta-analysis excluding the discovery data used by Global BPgen would be relatively free of winners' curse effects and would yield reasonably unbiased effect size estimates. Likewise, five signals were discovered previously by the CHARGE consortium, namely ULK4 (rs3774372), CACNB2(3') (rs1813353), PLEKHA7 (rs381815), ATP2B1 (rs17249754) and TBX5-TBX3 (rs10850411) [Levy *et al.* 2009], and therefore we meta-analysed excluding the discovery data used by CHARGE to obtain reasonably unbiased effect size estimates.

Three signals were discovered previously by both Global BPgen and CHARGE consortia, namely CYP17A1-NT5C2 (rs11191548), SH2B3 (rs3184504), and CYP1A1-ULK3 (rs1378942) [Newton-Cheh *et al.* 2009, Levy *et al.* 2009]. Discovery by Global BPgen implies that the CHARGE data should provide an estimate of effect size free of winners' curse bias, and discovery by CHARGE implies that the Global BPgen data should also provide an estimate of effect size free of winners' curse bias. By this logic, both discovery datasets provide unbiased effect size estimates, and for these signals we therefore estimated effect sizes from a meta-analysis of all available data.

The phenotypic correlation between SBP and DBP means that test statistics and effect size estimates for the two phenotypes are correlated, and therefore we used the same dataset exclusions to estimate effect sizes for both phenotypes, regardless of which phenotype was analysed when each signal was first discovered.

A.3 Variance explained

We assume Hardy–Weinberg proportions, so that the additive genetic variance for a given phenotype contributed by the i-th signal is $2p_i(1-p_i)\beta_i^2$, where p_i is the minor allele frequency p_i and β_i is the per-allele additive effect on the given phenotype. The fraction of phenotypic variance explained is then simply $R_i^2 = 2p_i(1-p_i)\beta_i^2/V$, where V is the phenotypic variance for the given phenotype (after adjustment for medication and correction for other covariates). This can be summed over independent signals, so the total variance explained by the inferred number of (discovered and as-yet-undiscovered) signals is simply $R_M^2 = \sum_i M_i R_i^2$.

page A.3 (Appendix A)

Table A.1: Estimates of effect sizes on DBP, based on all data (not used for this analysis but shown for comparison) and on data free from winners' curse bias. Discovery power depends on true effect size and also on the standard error for the discovery analysis (σ_D). All sample sizes (N) are effective sample sizes.

index SNP	All data			Excluding winners' curse			Discovery	
	N	$\hat{\beta}_\mathrm{D}^{(\mathrm{all})}$	SE	N	$\hat{\beta}_\mathrm{D}$	SE (s_D)	N	σ_D
rs17367504	125369	-0.547	0.0611	91262[G]	-0.561	0.0691	69718	0.0861
rs2932538	194654	0.24	0.0392	129872[I]	0.203	0.0466	64782	0.0725
rs13082711	198084	-0.238	0.0403	132696[I]	-0.18	0.0479	65388	0.0746
rs3774372	162073	-0.367	0.0492	131889[C]	-0.338	0.0526	69728	0.0797
rs419076	193714	0.241	0.0344	124412[I]	0.206	0.0415	69302	0.0614
rs1458038	139670	0.457	0.0444	108820[G]	0.434	0.0489	63514	0.0702
rs13107325	150918	-0.684	0.0807	92000[I]	-0.707	0.1025	58917	0.1308
rs13139571	185408	0.26	0.041	118270[I]	0.208	0.0496	67137	0.0726
rs1173771	158658	0.261	0.0382	89572[I]	0.281	0.0483	69086	0.0625
rs11953630	161073	-0.281	0.0387	91938[I]	-0.278	0.0488	69135	0.0635
rs1799945	143761	0.457	0.0573	76902[I]	0.44	0.0754	66860	0.0882
rs805303	201727	0.228	0.0343	132352[I]	0.219	0.0409	69376	0.0632
rs4373814	188373	-0.218	0.035	121006[I]	-0.195	0.0422	67368	0.0625
rs1813353	101835	0.415	0.0523	72548[C]	0.394	0.0589	67747	0.0668
rs4590817	111034	0.419	0.0591	77088[G]	0.372	0.0666	68502	0.0842
rs932764	160871	0.185	0.0374	91718[I]	0.207	0.047	69154	0.0618
rs11191548	161702	0.464	0.065	161702[n]	0.464	0.065	67829	0.1101
rs7129220	182867	-0.299	0.0552	118800[I]	-0.294	0.0663	64067	0.0999
rs381815	97144	0.348	0.0561	67312[C]	0.283	0.0637	68819	0.0697
rs633185	160448	-0.328	0.0412	91972[I]	-0.348	0.0519	68477	0.068
rs17249754	96416	0.522	0.0676	67619[C]	0.456	0.0774	68085	0.0842
rs3184504	120604	0.448	0.0432	120604[n]	0.448	0.0432	65005	0.0629
rs10850411	161136	0.253	0.0407	132844[C]	0.216	0.0435	65825	0.0683
rs1378942	163120	0.416	0.0392	163120[n]	0.416	0.0392	69386	0.064
rs2521501	127201	0.359	0.0451	88498[I]	0.332	0.053	38703	0.0858
rs17608766	151818	-0.129	0.054	92348[I]	-0.071	0.0671	59470	0.091
rs12940887	188169	0.27	0.0354	154411[G]	0.246	0.0382	68414	0.0636
rs1327235	158466	0.302	0.0378	89616[I]	0.302	0.0479	68850	0.0615
rs6015450	159170	0.557	0.0564	92294[I]	0.623	0.0699	66877	0.0955

Notes:

[I] excluding all ICBP GWAS discovery (stage 0) data

[C] excluding CHARGE discovery GWAS of Levy *et al.* [2009]

[G] excluding Global BPgen discovery GWAS of Newton-Cheh *et al.* [2009]

[n] no exclusions

page A.4 (Appendix A)

Table A.2: Estimates of effect sizes on SBP, based on all data (not used for this analysis but shown for comparison) and on data free from winners' curse bias. Discovery power depends on true effect size and also on the standard error for the discovery analysis (σ_S). All sample sizes (N) are effective sample sizes.

index SNP	All data			Excluding winners' curse			Discovery	
	N	$\hat{\beta}_S^{(all)}$	SE	N	$\hat{\beta}_S$	SE (s_S)	N	σ_S
rs17367504	125369	-0.903	0.0942	91262[G]	-0.95	0.1079	69718	0.1356
rs2932538	194654	0.388	0.0638	129872[I]	0.321	0.0768	64782	0.1147
rs13082711	198084	-0.315	0.0655	132696[I]	-0.318	0.0787	65388	0.1182
rs3774372	162073	-0.067	0.0783	131889[C]	-0.073	0.083	69728	0.1247
rs419076	193714	0.409	0.0555	124412[I]	0.355	0.0678	69302	0.0967
rs1458038	139670	0.706	0.0705	108820[G]	0.732	0.0792	63514	0.1106
rs13107325	150918	-0.981	0.1293	92000[I]	-0.923	0.1649	58917	0.2084
rs13139571	185408	0.321	0.0661	118270[I]	0.266	0.0809	67137	0.1144
rs1173771	158658	0.504	0.0612	89572[I]	0.495	0.0781	69086	0.0986
rs11953630	161073	-0.412	0.062	91938[I]	-0.357	0.0789	69135	0.1002
rs1799945	143761	0.627	0.0916	76902[I]	0.649	0.1214	66860	0.1396
rs805303	201727	0.376	0.0556	132352[I]	0.327	0.0671	69376	0.0996
rs4373814	188373	-0.373	0.0567	121006[I]	-0.318	0.0692	67368	0.0986
rs1813353	101835	0.569	0.0812	72548[C]	0.489	0.0895	67747	0.1051
rs4590817	111034	0.646	0.0931	77088[G]	0.626	0.1066	68502	0.1332
rs932764	160871	0.484	0.0599	91718[I]	0.471	0.0759	69154	0.0976
rs11191548	161702	1.095	0.1041	161702[n]	1.095	0.1041	67829	0.1742
rs7129220	182867	-0.619	0.0886	118800[I]	-0.52	0.1079	64067	0.1555
rs381815	97144	0.575	0.0876	67312[C]	0.485	0.0972	68819	0.1098
rs633185	160448	-0.565	0.066	91972[I]	-0.553	0.0838	68477	0.1073
rs17249754	96416	0.928	0.1059	67619[C]	0.763	0.119	68085	0.1339
rs3184504	120604	0.598	0.0688	120604[n]	0.598	0.0688	65005	0.0993
rs10850411	161136	0.354	0.0651	132844[C]	0.322	0.069	65825	0.1076
rs1378942	163120	0.613	0.0621	163120[n]	0.613	0.0621	69386	0.1006
rs2521501	127201	0.65	0.073	88498[I]	0.62	0.0862	38703	0.1369
rs17608766	151818	-0.556	0.0863	92348[I]	-0.47	0.1084	59470	0.1424
rs12940887	188169	0.362	0.0568	154411[G]	0.354	0.0621	68414	0.1002
rs1327235	158466	0.34	0.0605	89616[I]	0.329	0.0774	68850	0.0971
rs6015450	159170	0.896	0.0905	92294[I]	0.951	0.1134	66877	0.1502

Notes:

[I] excluding all ICBP GWAS discovery (stage 0) data

[C] excluding CHARGE discovery GWAS of Levy *et al.* [2009]

[G] excluding Global BPgen discovery GWAS of Newton-Cheh *et al.* [2009]

[n] no exclusions

page A.5 (Appendix A)

We obtained residual phenotypic variances applicable for the combined ICBP-GWAS discovery sample from the discovery stage meta-analysis association standard errors (σ_i for the i-th signal), effective sample sizes (N_i for the i-th signal), and allele frequencies, using

$$V = \text{mean}_i\big(2p_i(1 - p_i) \times N \times \sigma^2\big) \tag{1}$$

where the mean is over all signals (truly associated or not). We obtained $V_\text{D} = (11.25\text{mmHg})^2$ for DBP and $V_\text{S} = (17.89\text{mmHg})^2$ for SBP. (These are similar to the values obtained by taking weighted averages of residual SDs, after adjustment for medication and correction for covariates, over all cohorts in the discovery dataset.)

Using the unbiased effect size estimates from Tables A.1 and A.2, we estimate that the total variance explained by the 29 discovered signals (only) is 0.943% for DBP and 0.919% for SBP.

A.4 Estimating power for discovered associations

It is important to take into account uncertainty about the true effect sizes of the discovered signals, because discovery power is a non-linear function of effect size. A reasonable method for estimating discovery power is to express power as a function of true effect size, and integrate with respect to a probability distribution for true effect size. Here, for each signal we integrate with respect to a joint distribution for the true effect sizes on DBP (β_D) and on SBP (β_S) that is the Bayesian posterior distribution, assuming a locally uniform prior. That is, we assume

$$\begin{bmatrix} \beta_\text{D} \\ \beta_\text{S} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\beta}_\text{D} \\ \hat{\beta}_\text{S} \end{bmatrix}, \begin{bmatrix} s_\text{D}^2 & rs_\text{D}s_\text{S} \\ rs_\text{D}s_\text{S} & s_\text{S}^2 \end{bmatrix}\right) \tag{2}$$

where the parameters of the normal distribution are the mean and the variance-covariance matrix. Here, $\hat{\beta}_\text{D}$ and $\hat{\beta}_\text{S}$ are the estimates (from Tables A.1 and A.2) excluding winners' curse, s_D and s_S the corresponding standard errors, and r the phenotypic correlation between DBP and SBP (after adjustment for covariates; this is asymptotically equal to the posterior correlation).

Given true effect sizes β_D and β_S, the joint distribution of beta estimates in the ICBP-GWAS discovery experiment is bivariate normal

$$\begin{bmatrix} \tilde{\beta}_\text{D} \\ \tilde{\beta}_\text{S} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_\text{D} \\ \beta_\text{S} \end{bmatrix}, \begin{bmatrix} \sigma_\text{D}^2 & r\sigma_\text{D}\sigma_\text{S} \\ r\sigma_\text{D}\sigma_\text{S} & \sigma_\text{S}^2 \end{bmatrix}\right) \tag{3}$$

page A.6 (Appendix A)

where σ_D and σ_S are the standard errors for the ICBP-GWAS discovery experiment (given in Tables A.1 and A.2), which for small effect sizes depend only on the phenotypic variances (after adjustment for covariates), effective sample size, and allele frequency, all of which we assume are known in advance and do not depend on β_D and β_S.

Estimating discovery power by integrating with respect to a probability distribution for the true effect sizes β_D and β_S is equivalent to assuming that the true effect sizes are random. Marginal to the true effect sizes, the joint distribution of the estimates in a discovery experiment is

$$\begin{bmatrix} \tilde{\beta}_D \\ \tilde{\beta}_S \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\beta}_D \\ \hat{\beta}_S \end{bmatrix}, \begin{bmatrix} s_D^2 + \sigma_D^2 & r(s_D s_S + \sigma_D \sigma_S) \\ r(s_D s_S + \sigma_D \sigma_S) & s_S^2 + \sigma_S^2 \end{bmatrix} \right) \tag{4}$$

and the joint distribution of the corresponding signed t-statistics ($T_D \equiv \tilde{\beta}_D/\sigma_D$ and $T_S \equiv \tilde{\beta}_S/\sigma_S$) is

$$\begin{bmatrix} T_D \\ T_S \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\beta}_D/\sigma_D \\ \hat{\beta}_S/\sigma_S \end{bmatrix}, \begin{bmatrix} 1 + s_D^2/\sigma_D^2 & r(1 + s_D s_S/\sigma_D \sigma_S) \\ r(1 + s_D s_S/\sigma_D \sigma_S) & 1 + s_S^2/\sigma_S^2 \end{bmatrix} \right) \tag{5}$$

Following Park *et al.* [2010], we assume that signals with truely non-zero effects will be discovered by ICBP-GWAS if and only if they are selected for entry into (stage 1 of) the ICBP-GWAS validation experiment, on the basis of (stage 0) ICBP-GWAS meta-analysis results. Hence, discovery power is equivalent to probability of selection for (stage 1) validation. However, for the ICBP-GWAS study, selection for validation was not based on a strict P-value threshold, but was based on P-value along with additional bioinformatic and literature information. Here, we make the simplying assumption that this additional information was independent of the true association status for each signal, and therefore that selection for validation can be modelled as a random selection process depending only on stage 0 association P-values for DBP and SBP. Specifically, we assume that selection depended on the smaller of the assocation P-values for SBP and DBP, according to a binomial regression model with probit link function and dependence on ($|T_D| \wedge |T_S|$), the greater of the absolute values of the t-statistics for DBP and SBP. This regression model has the technically advantageous feature that "random" selection (with probability of selection increasing for larger absolute t-statistics) is well approximated by deterministic selection at a fixed threshold, after the t-statistics have been perturbed by normal random "noise". We assume that selection is made independently for each signal, so that this noise is independent across signals. Because selection is a decision made on the basis of the test statistics for both DBP and SBP, the noise is perfectly correlated for the DBP and SBP test statistics for a given signal.

page A.7 (Appendix A)

Specifically,

$$\Pr(\text{select}) = \Phi(d_1 + d_2(|T_{\mathrm{D}}| \wedge |T_{\mathrm{S}}|)) \tag{6}$$

$$= \Pr(Z \le d_1 + d_2(|T_{\mathrm{D}}| \wedge |T_{\mathrm{S}}|)) \tag{7}$$

$$\simeq \Pr((|T_{\mathrm{D}} + \epsilon Z| \ge k) \cup (|T_{\mathrm{S}} + \epsilon Z| \ge k)) \tag{8}$$

where d_1 and d_2 are coefficients in the regression model, Z is an independent standard normal random variable, $\epsilon = -1/d_2$ measures the "noisiness" of the selection process, and $k = -d_1/d_2$ is the "average" threshold for selection, and the approximation assumes ϵ is small relative to k, so that e.g. $|T_{\mathrm{D}}| + \epsilon Z \simeq |T_{\mathrm{D}} + \epsilon Z|$ for signals that have non-neglible probability of selection. By retrospectively fitting this regression model to the discovery data (T_{D} and T_{S} at stage 0 each independent SNP) and yes/no selection decisions made, we estimate $\epsilon = -0.173$ and $k = 4.83$, which we treat as fixed in the following.

The joint distribution is these "perturbed" t-statistics for SBP and DBP, perturbed by the *same* random noise ϵZ, is bivariate normal

$$\begin{bmatrix} T_{\mathrm{D}} + \epsilon Z \\ T_{\mathrm{S}} + \epsilon Z \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \hat{\beta}_{\mathrm{D}}/\sigma_{\mathrm{D}} \\ \hat{\beta}_{\mathrm{S}}/\sigma_{\mathrm{S}} \end{bmatrix}, \begin{bmatrix} 1 + s_{\mathrm{D}}^2/\sigma_{\mathrm{D}}^2 + \epsilon^2 & r(1 + s_{\mathrm{D}}s_{\mathrm{S}}/\sigma_{\mathrm{D}}\sigma_{\mathrm{S}}) + \epsilon^2 \\ r(1 + s_{\mathrm{D}}s_{\mathrm{S}}/\sigma_{\mathrm{D}}\sigma_{\mathrm{S}}) + \epsilon^2 & 1 + s_{\mathrm{S}}^2/\sigma_{\mathrm{S}}^2 + \epsilon^2 \end{bmatrix} \right) \tag{9}$$

We calculate the discovery probability by integrating the bivariate normal density (9) over the region satisfying the inequality in expression (8). We performed the integration using the standard normal distribution function to obtain $\Pr(|T_{\mathrm{D}} + \epsilon Z| \ge k)$, and to obtain $\Pr(T_{\mathrm{S}} + \epsilon Z \ge k \mid T_{\mathrm{D}})$ over a grid of values for T_{D}:

$$\Pr(\text{select}) \simeq \Pr((|T_{\mathrm{D}} + \epsilon Z| \ge k) \cup (|T_{\mathrm{S}} + \epsilon Z| \ge k)) \tag{10}$$

$$\simeq \Pr(|T_{\mathrm{D}} + \epsilon Z| \ge k)$$

$$+ \sum_i w_i \Pr(T_{\mathrm{D}} + \epsilon Z = z_i) \Pr(|T_{\mathrm{S}} + \epsilon Z| \ge k \mid T_{\mathrm{D}} + \epsilon Z = z_i) \tag{11}$$

$$\simeq \Pr(|T_{\mathrm{S}} + \epsilon Z| \ge k)$$

$$+ \sum_i w_i \Pr(T_{\mathrm{S}} + \epsilon Z = z_i) \Pr(|T_{\mathrm{D}} + \epsilon Z| \ge k \mid T_{\mathrm{S}} + \epsilon Z = z_i) \tag{12}$$

where the z_i are quadrature points on $[-k, k]$ and w_i are quadrature weights. The implementation and accuracy of the numerical integration were checked by numerically comparing (11) and (12). The estimated power for each detected signal, and resulting estimate of total number of signals, are given in Table A.3. The shape of the "visible tail" of the effect size distributions can be visualised using a nonparametric density estimation method, as shown in Figures A.1 and A.2.

page A.8 (Appendix A)

Table A.3: Estimates of discovery power for 29 associated signals, and resulting estimate of total number of signals and total variance that would be explained.

index SNP	signal	power (B_i)	estimated signals (M_i)	R^2 (%) DBP	R^2 (%) SBP
rs17367504	MTHFR-NPPB	0.986	1	0.062	0.071
rs2932538	MOV10	0.084	11.9	0.012	0.012
rs13082711	SLC4A7	0.056	17.8	0.009	0.011
rs3774372	ULK4	0.313	3.2	0.027	0
rs419076	MECOM	0.239	4.2	0.017	0.02
rs1458038	FGF5	0.973	1	0.06	0.068
rs13107325	SLC39A8	0.743	1.3	0.051	0.035
rs13139571	GUCY1A3-GUCY1B3	0.071	14.2	0.013	0.008
rs1173771	NPR3-C5orf23	0.673	1.5	0.03	0.037
rs11953630	EBF1	0.424	2.4	0.029	0.019
rs1799945	HFE	0.691	1.4	0.038	0.033
rs805303	BAT2-BAT5	0.194	5.1	0.018	0.016
rs4373814	CACNB2(5')	0.154	6.5	0.015	0.016
rs1813353	CACNB2(3')	0.842	1.2	0.053	0.032
rs4590817	C10orf107	0.604	1.7	0.029	0.033
rs932764	PLCE1	0.529	1.9	0.017	0.034
rs11191548	CYP17A1-NT5C2	0.9	1.1	0.027	0.059
rs7129220	ADM	0.152	6.6	0.013	0.016
rs381815	PLEKHA7	0.51	2	0.024	0.028
rs633185	FLJ32810-TMEM133	0.775	1.3	0.039	0.039
rs17249754	ATP2B1	0.87	1.1	0.044	0.049
rs3184504	SH2B3	0.985	1	0.079	0.056
rs10850411	TBX5-TBX3	0.127	7.9	0.015	0.013
rs1378942	CYP1A1-ULK3	0.968	1	0.062	0.053
rs2521501	FES	0.473	2.1	0.037	0.051
rs17608766	GOSR2	0.114	8.8	0.001	0.017
rs12940887	ZNF652	0.282	3.5	0.022	0.018
rs1327235	JAG1	0.554	1.8	0.036	0.017
rs6015450	GNAS-EDN3	0.973	1	0.063	0.058
Total signals $M = \sum_i M_i$			115.6		
Total % variance explained $R^2 = \sum_i M_i R_i^2$				2.162	2.183

page A.9 (Appendix A)

Figure A.1: Distribution of effect sizes on the R^2 scale. Solid lines show estimates using a kernel density method applied to the unbiased R^2 estimates for each signal (from Table A.3, shown as ticks on the x-axis), with weights proportional to the estimated number of signals (the M_i from Table A.3).
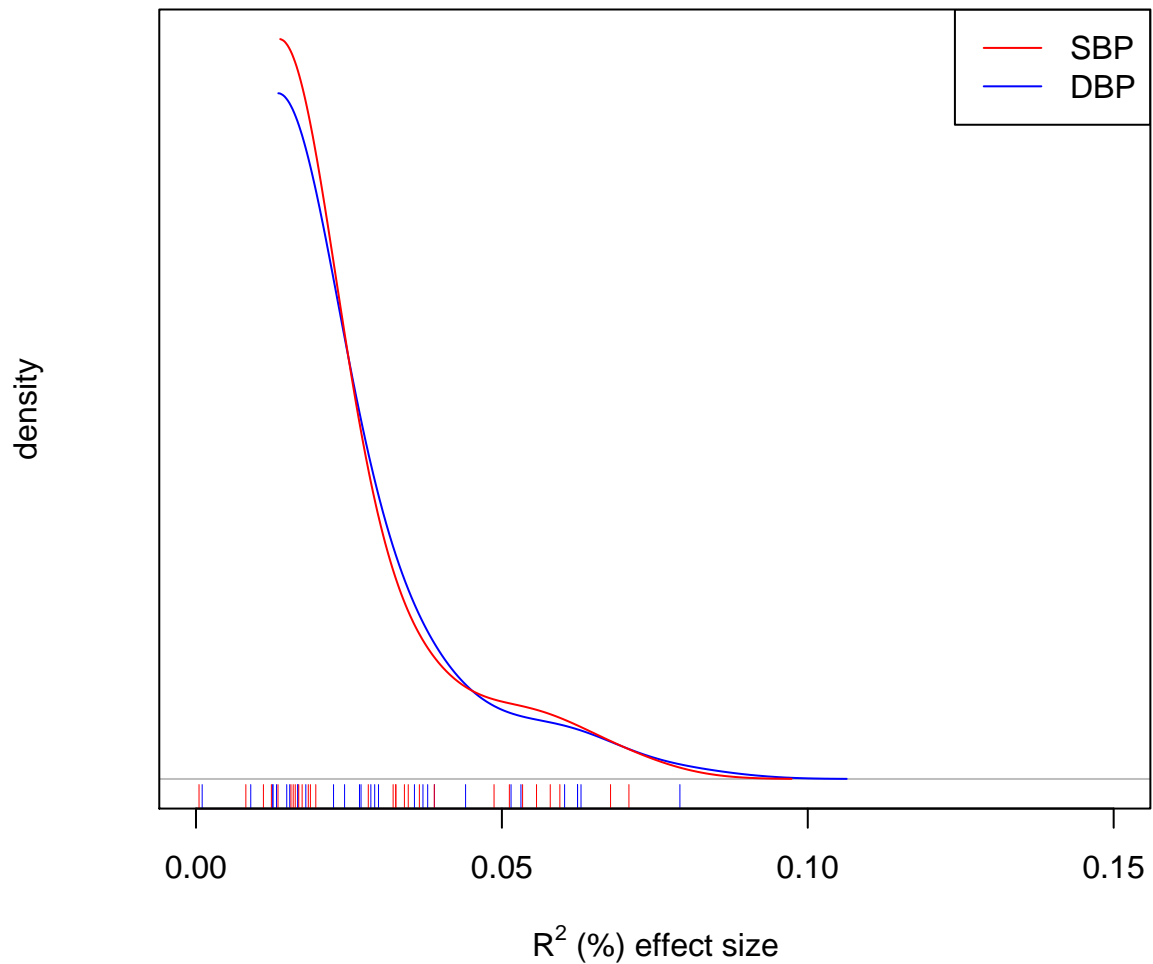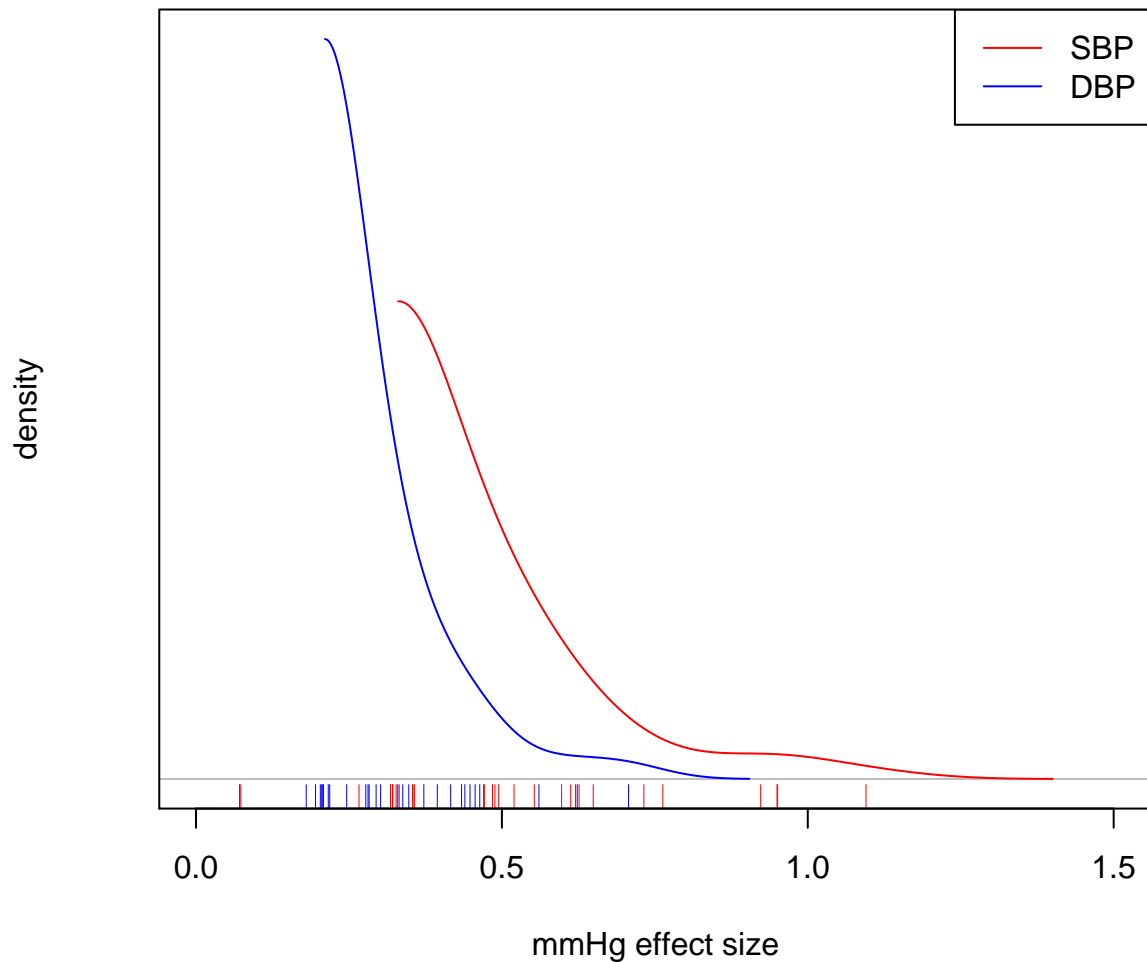
page A.10 (Appendix A)

Figure A.2: Distribution of effect sizes on the mmHg scale. Solid lines show estimates using a kernel density method applied to the unbiased mmHg effect size estimates for each signal (from Tables A.1 and A.2, shown as ticks on the x-axis), with weights proportional to the estimated number of signals (the M_i from Table A.3).

A.5 Bootstrap confidence interval

We followed Park *et al.* [2010] to obtain an approximate confidence interval on the estimate M, by parametric bootstrapping. Since the estimated contribution to the total number of signals for each row in Table A.3, M_i, is non-integer, for each bootstrap simulation we simulated a discovery dataset as follows: For each i, the number of discovered signals L_i was simulated from a binomial distribution with parameters B_i and $\lceil M_i \rceil$. We then adjusted the contribution to the total estimate M by an appropriate inverse factor, so that

$$M_{\text{boot}} = \sum_i L_i \times 1/B_i \times M_i/\lceil M_i \rceil \tag{13}$$

Using 10000 bootstrap simulations we obtained a bootstrap distribution for M with mean 115.8, 2.5-th percentile 63.8 and 97.5-th percentile 179.0, which suggests that (under the parametric assumptions) unbiased estimation of M with 95% confidence interval 52–168.

The variance of the bootstrap distribution can in fact be calculated analytically and is

$$\text{Var}(M_{\text{boot}}) = \sum_i M_i^2/\lceil M_i \rceil \, (1 - B_i)/B_i \simeq \sum_i M_i(1 - B_i)/B_i \tag{14}$$

which for our data is 876.6 and agrees closely with the observed variance 861.0. Inspection of equation (14) shows that, like the estimate M itself, the the bootstrap variance is also very sensitive to the number of signals with small B_i.

A.6 Prediction of novel discoveries

We predicted the number of signals that would be discovered as a function of sample size (main text figure 2b) as follows: We modelled discovery power as a function of (effective) discovery sample size N, assuming that standard errors (σ_i) in a discovery experiment depend on N according to

$$\sigma_i = \sqrt{\frac{V}{2p_i(1 - p_i) \times N}} \tag{15}$$

where the phenotypic variance V was as estimated using Equation (1) and p_i is the allele frequency. For each observed signal in Table A.3, we assumed there were M_i similar true signals that could potentially be discovered, with allele frequencies (p_i) equal to those observed, and effect size of each distributed as estimated from the winners' curse free data (Tables A.1 and A.2; Equation (2)). For each true signal assumed to exist, we estimated discovery power $B_i^{(N)}$ for each value of N using Equations (9)–(12), for a strict discovery

<center>page A.12 (Appendix A)</center>

P-value threshold $P < \alpha = 10^{-6}$ (corresponding to $\epsilon = 0$ and $k = \Phi^{-1}(1 - \alpha/2) = 4.89$). We assume each true signal is detected with independent probability $B_i^{(N)}$, and hence the total variance in predicted number of discovered signals is a sum of Bernouilli variances.

We predicted the number of novel signals that would be discovered (excluding the 29 signals already discovered) if further independent signals from our discovery experiment were to be were followed up in a future large scale validation study. We calculated power for the ICBP-GWAS stage 0 discovery experiment, but for a range of less stringent significance thresholds $\alpha = 10^{-5}$, $\alpha = 10^{-4}$, $\alpha = 10^{-3}$ and $\alpha = 10^{-2}$ for the more significant of the test statistics for DBP and SBP, assuming deterministic selection of signals for validation. That is, we repeated the power calculations described above, but used $\epsilon = 0$ and $k = \Phi^{-1}(1 - \alpha/2)$, to obtain power estimates $B_i^{(\alpha)}$ for each significance threshold α.

The expected number of novel discoveries (with effect size equal to that of the i-th observed discovery) to be made in a validation experiment following up all signals with $P \leq \alpha$ is then given by

$$D_i^{(\alpha)} = (M_i - 1)B_i^{(\alpha)} \tag{16}$$

The total expected number of novel discoveries for each α are given in Table A.4.

We note that, because power increases most slowly for the effect sizes with the largest estimated numbers of signals, very large GWAS sample sizes would be needed to identify the majority of the as-yet-undiscovered signals within the spectrum of effect sizes observed for the signals that were discovered by ICBP-GWAS. For example (as shown in the main text Figure 2b), in an enlarged GWAS with $N = 150,000$, we expect $(76 - 29)/(115.6 - 29) = 54\%$ of as-yet-undiscovered signals to reach $P < 10^{-6}$. For the same reason (as shown in Table A.4), it would be necessary to follow up an extremely large number of signals (e.g. all independent signals with discovery $P < 0.01$ in ICBP-GWAS) in a large independent validation experiment, in order to discover $68.3/(115.6 - 29) = 79\%$ of as-yet-undiscovered signals.

A.7 Robustness

We assumed $r = 0.6$ for all results presented. Assuming $r = 0.4$ or $r = 0.8$ made relatively little difference to the overall estimate ($M = 113$ and $M = 120$ respectively).

We explored the effect of not taking into account uncertainty in the true effect sizes β, which results in smaller estimates of discovery power B for signals with small B. Assuming no uncertainty ($s_D = 0$ and $s_S = 0$ for all signals) resulted in a substantially larger estimate of $M = 202$, illustrating the importance of properly taking into account uncertainty in effect

page A.13 (Appendix A)

Table A.4: Discovery power $B_i^{(\alpha)}$ for the ICBP-GWAS discovery experiment at a range of less stringent significance thresholds α. We estimated the total number of truely associated novel signals that would be discovered if all signals reaching these thresholds were taken forward into sufficiently large validation experiments.

index SNP	estimated signals (M_i)	$\alpha = 10^{-2}$		$\alpha = 10^{-3}$		$\alpha = 10^{-4}$		$\alpha = 10^{-5}$	
		$B_i^{(\alpha)}$	$D_i^{(\alpha)}$	$B_i^{(\alpha)}$	$D_i^{(\alpha)}$	$B_i^{(\alpha)}$	$D_i^{(\alpha)}$	$B_i^{(\alpha)}$	$D_i^{(\alpha)}$
rs17367504	1	1	0	1	0	0.999	0	0.995	0
rs2932538	11.9	0.752	8	0.506	5	0.295	2.5	0.154	0.8
rs13082711	17.8	0.676	11	0.418	6.4	0.225	3	0.109	0.9
rs3774372	3.2	0.92	1.9	0.787	1.5	0.616	1	0.442	0.4
rs419076	4.2	0.911	2.8	0.754	2.1	0.554	1.3	0.366	0.5
rs1458038	1	1	0	1	0	0.998	0	0.99	0
rs13107325	1.3	0.996	0.3	0.978	0.3	0.931	0.3	0.846	0.1
rs13139571	14.2	0.703	9	0.454	5.4	0.256	2.6	0.131	0.9
rs1173771	1.5	0.993	0.5	0.966	0.4	0.901	0.3	0.794	0.2
rs11953630	2.4	0.964	1.3	0.876	1.1	0.735	0.7	0.566	0.3
rs1799945	1.4	0.993	0.4	0.967	0.4	0.906	0.3	0.805	0.2
rs805303	5.1	0.889	3.6	0.709	2.6	0.497	1.6	0.311	0.6
rs4373814	6.5	0.849	4.5	0.645	3.2	0.429	1.8	0.255	0.7
rs1813353	1.2	0.998	0.2	0.99	0.2	0.965	0.1	0.913	0.1
rs4590817	1.7	0.989	0.6	0.949	0.6	0.864	0.4	0.736	0.2
rs932764	1.9	0.978	0.8	0.917	0.7	0.808	0.5	0.663	0.3
rs11191548	1.1	1	0.1	0.996	0.1	0.984	0.1	0.952	0.1
rs7129220	6.6	0.845	4.6	0.639	3.2	0.424	1.8	0.252	0.7
rs381815	2	0.974	0.9	0.909	0.8	0.794	0.6	0.645	0.3
rs633185	1.3	0.998	0.3	0.985	0.3	0.947	0.2	0.872	0.1
rs17249754	1.1	0.999	0.1	0.993	0.1	0.974	0.1	0.932	0.1
rs3184504	1	1	0	1	0	0.999	0	0.995	0
rs10850411	7.9	0.827	5.5	0.606	3.8	0.386	2	0.219	0.7
rs1378942	1	1	0	1	0	0.997	0	0.989	0
rs2521501	2.1	0.98	1.1	0.915	0.9	0.79	0.7	0.623	0.3
rs17608766	8.8	0.726	5.4	0.507	3.4	0.321	1.8	0.188	0.6
rs12940887	3.5	0.94	2.3	0.807	1.9	0.617	1.2	0.422	0.5
rs1327235	1.8	0.981	0.8	0.926	0.7	0.824	0.5	0.686	0.2
rs6015450	1	1	0	1	0	0.998	0	0.99	0
Total	115.6		68.3		50.2		33.6		21.0

page A.14 (Appendix A)

sizes to avoid upward bias in M.

References

Levy, D., G. B. Ehret, K. Rice, G. C. Verwoert, L. J. Launer, A. Dehghan, N. L. Glazer, A. C. Morrison, A. D. Johnson, T. Aspelund, Y. Aulchenko, T. Lumley, A. Kottgen, R. S. Vasan, F. Rivadeneira, G. Eiriksdottir, X. Guo, D. E. Arking, G. F. Mitchell, F. U. S. Mattace-Raso, A. V. Smith, K. Taylor, R. B. Scharpf, S.-J. Hwang, E. J. G. Sijbrands, J. Bis, T. B. Harris, S. K. Ganesh, C. J. O'Donnell, A. Hofman, J. I. Rotter, J. Coresh, E. J. Benjamin, A. G. Uitterlinden, G. Heiss, C. S. Fox, J. C. M. Witteman, E. Boerwinkle, T. J. Wang, V. Gudnason, M. G. Larson, A. Chakravarti, B. M. Psaty and C. M. van Duijn (2009) Genome-wide association study of blood pressure and hypertension. *Nature Genetics* **41**:677–687. doi:10.1038/ng.384.

Newton-Cheh, C., T. Johnson, V. Gateva, M. D. Tobin, M. Bochud, L. Coin, S. S. Najjar, J. H. Zhao, S. C. Heath, S. Eyheramendy, K. Papadakis, B. F. Voight, L. J. Scott, F. Zhang, M. Farrall, T. Tanaka, C. Wallace, J. C. Chambers, K.-T. Khaw, P. Nilsson, P. van der Harst, S. Polidoro, D. E. Grobbee, N. C. Onland-Moret, M. L. Bots, L. V. Wain, K. S. Elliott, A. Teumer, J. Luan, G. Lucas, J. Kuusisto, P. R. Burton, D. Hadley, W. L. McArdle, M. Brown, A. Dominiczak, S. J. Newhouse, N. J. Samani, J. Webster, E. Zeggini, J. S. Beckmann, S. Bergmann, N. Lim, K. Song, P. Vollenweider, G. Waeber, D. M. Waterworth, X. Yuan, L. Groop, M. Orho-Melander, A. Allione, A. D. Gregorio, S. Guarrera, S. Panico, F. Ricceri, V. Romanazzi, C. Sacerdote, P. Vineis, I. Barroso, M. S. Sandhu, R. N. Luben, G. J. Crawford, P. Jousilahti, M. Perola, M. Boehnke, L. L. Bonnycastle, F. S. Collins, A. U. Jackson, K. L. Mohlke, H. M. Stringham, T. T. Valle, C. J. Willer, R. N. Bergman, M. A. Morken, A. Doring, C. Gieger, T. Illig, T. Meitinger, E. Org, A. Pfeufer, H. E. Wichmann, S. Kathiresan, J. Marrugat, C. J. O'Donnell, S. M. Schwartz, D. S. Siscovick, I. Subirana, N. B. Freimer, A.-L. Hartikainen, M. I. McCarthy, P. F. O'Reilly, L. Peltonen, A. Pouta, P. E. de Jong, H. Snieder, W. H. van Gilst, R. Clarke, A. Goel, A. Hamsten, J. F. Peden, U. Seedorf, A.-C. Syvanen, G. Tognoni, E. G. Lakatta, S. Sanna, P. Scheet, D. Schlessinger, A. Scuteri, M. Dorr, F. Ernst, S. B. Felix, G. Homuth, R. Lorbeer, T. Reffelmann, R. Rettig, U. Volker, P. Galan, I. G. Gut, S. Hercberg, G. M. Lathrop, D. Zelenika, P. Deloukas, N. Soranzo, F. M. Williams, G. Zhai, V. Salomaa, M. Laakso, R. Elosua, N. G. Forouhi, H. Volzke, C. S. Uiterwaal, Y. T. van der Schouw, M. E. Numans, G. Matullo, G. Navis, G. Berglund, S. A. Bingham, J. S. Kooner, J. M. Connell, S. Bandinelli, L. Ferrucci, H. Watkins, T. D. Spector, J. Tuomilehto, D. Altshuler, D. P. Strachan, M. Laan, P. Meneton, N. J. Wareham, M. Uda, M.-R. Jarvelin,

V. Mooser, O. Melander, R. J. Loos, P. Elliott, G. R. Abecasis, M. Caulfield and P. B. Munroe (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics* **41**:666–676. doi:10.1038/ng.361.

Park, J.-H., S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs, S. J. Chanock and N. Chatterjee (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genetics* **42**:570–575. doi:10.1038/ng.610.

page A.16 (Appendix A)