doi:10.1038/nature10659

New gene functions in megakaryopoiesis and platelet formation

Supplementary Information

Corresponding authors:

Christian Gieger
Institute of Genetic Epidemiology
Helmholtz Zentrum München, German Research Center for Environmental Health
85764 Neuherberg, Germany
Email: christian.gieger@helmholtz-muenchen.de

Willem H Ouwehand
Department of Haematology
University of Cambridge & NHS Blood and Transplant
Cambridge, UK
Email: who1000@cam.ac.uk

Nicole Soranzo
Human and Medical Genetics Department
Wellcome Trust Sanger Institute
Hinxton, UK
Email: ns6@sanger.ac.uk

Table of Contents

Population samples

Discovery

ARIC

Sample. The Atherosclerosis Risk in Communities (ARIC) Study recruited by probability sampling 15,792 African American and European American adults aged 45 to 64 years in 1987 through 1989 from Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis, Minnesota; and Washington County, Maryland [1]. The Jackson sample comprised African Americans only; the other three samples included African Americans, European Americans, as well as a small proportion of participants of other ethnicity. The ARIC study was approved by the institutional review board of each field center institutes and participants gave informed consent including consent for genetic testing. In this study we included only European American adults.

Blood count measurements. Fasting blood was drawn at baseline and processed following a standard protocol. Platelet count was determined in hospital-based, independent laboratories within 24 hours after venepuncture, after storage at 4^0C. The measurement was conducted using automated particle counters (Coulter Diagnostics, Hialeah, Florida, in three field centers; Technicon H-6000, Technicon Corporation, Tarrytown, New York, in one field center).

CHS

Sample. The Cardiovascular Health Study (CHS) recruited participants 65 years of age and older from 4 US communities in 2 waves: 5,201 participants in 1989-1990, and an additional 687 African Americans in 1992-1993 [2]. The human subjects committee approved the study and written informed consent to use genetic information was provided by study participants included in the analysis. In this study we included European American adults.

Blood count measurements. Plasma measures of platelet counts were obtained at the time of cohort entry for CHS. Venous blood was collected into a 4.5 mmol/L EDTA tube [3]. Platelet counts were measured at field center laboratories by Coulter counters.

EPIC - Norfolk

Sample. The European Prospective Investigation into Cancer and Nutrition (EPIC) Obesity study used a case-cohort design which included 1284 participants whose body mass index was above 30 kg/m^2 and a random sample of 2,566 participants from the EPIC-Norfolk Study, a population-based cohort study of 25,663 men and women of European descent aged 39-79 years recruited in Norfolk between 1993 and 1997. Only individuals of the population-based random sample were analyzed and included in the meta-analysis.

Blood count measurements. Blood sample was taken during the day in the GPs' surgeries or EPIC clinic, were held overnight. Early the following morning, samples were collected from GP surgeries by technicians. Some assays were performed on fresh blood samples and the remaining blood was stored in straws. A 1 x 2ml EDTA sample provided blood for full blood count. Two x 10 ml citrated samples provided

twelve straws of plasma, four straws of red cells plus preservation buffer and four straws of buffy coat and saline. A Coulter MD18 hematology analyser was used for the measurement of full blood counts. Quality controls were carried out on the Coulter scheme daily. In addition, the Hematology Department of Addenbrooke's Hospital included the EPIC Laboratory in a monthly quality control scheme.

HVH

Sample. The Heart and Vascular (HVH) Study is a population-based, case-control study conducted at Group Health (GH), a large integrated health care system in western Washington State [4-6]. Subjects were women and hypertensive men 30-79 years of age. The human subjects committee at GH approved the study, and all study participants provided written informed consent. For the meta-analysis, only control subjects were included.

Blood count measurements. Plasma measures of platelet counts were obtained at the time of phlebotomy. Venous blood was collected into a 4.5 mmol/L EDTA tube. Platelet counts were measured as part of a complete blood count panel.

INGI Carlantino

Sample. The INGI Carlantino cohort comprised of about 900 samples from an isolated village of southern Italy (Carlantino). Genotyping data for 679 people is available, of these 521 had also phenotypic data. Ethics approval was obtained from the Ethics Committee of the Burlo Garofolo children hospital in Trieste. Written informed consent was obtained from every participant to the study.

Blood count measurements. Venous blood was anticoagulated with EDTA and full blood counts (FBCs) were performed within few hours using an automated Coulter.

INGI FVG

Sample. The INGI Friuli Venezia Giulia (FVG) cohort comprised of about 1300 samples from 5 isolated villages of Friuli Venezia Giulia a region of Northern Italy. The villages are: Resia, Illegio, San Martino del Carso, Erto/Casso, Clauzetto. Genotyping data for 1,266 people is available, of these 1,046 had also phenotypic data. Ethics approval was obtained from the Ethics Committee of the Burlo Garofolo children hospital in Trieste. Written informed consent was obtained from every participant to the study.

Blood count measurements. Venous blood was anticoagulated with EDTA and FBCs were performed within few hours using an automated Coulter.

INGI Val Borbera

Sample. The INGI-Val Borbera population is a collection of 1,664 genotyped samples collected in the Val Borbera Valley, a geographically isolated valley located within the Appennine Mountains in Northwest Italy. The valley is inhabited by about 3,000 descendants from the original population, living in 7 villages along the valley and in the mountains. The valley was inhabited by about 10,000 people in the 19th century when endogamy was >80%. Around 1930, the population started to decrease due to emigration to South America. Participants were healthy people between 18 and 102 years of age that had at least one grandfather living in the valley.

Blood count measurements. Venous blood was performed using either an SF3000

hematology analyzer or a XE2100 hematology analyzer. The two instruments displayed no significant statistical differences in measurements range so association analyses were not adjusted for instrument type.

KORA F3

Sample. The study population was recruited from the KORA S3 survey (4,856 subjects, response 75%). This is an independent population-based sample from the general population living in the region of Augsburg, Southern Germany, examined in the years 1994/95. The standardized examinations have been described in detail elsewhere [7]. A total of 3,006 subjects participated in a follow-up examination of S3 in 2004/05 (KORA F3). For the genome-wide KORA F3 study we selected 1,644 subjects of these participants then aged 35 to 79 years. Informed consent has been given. The local ethical committee has approved the study.

Blood count measurements. DNA was extracted from fresh blood, and was stored at -80°C. FBCs were performed on fresh venous EDTA-anticoagulated blood using an automatic blood counter (Beckman Coulter STKS).

KORA F4

Sample. The KORA S4 survey, an independent population-based sample from the general population living in the region of Augsburg, Southern Germany, was conducted in 1999/2001. The standardized examinations applied in the survey (4,261 participants, response 67%) have been described in detail elsewhere [8]. A total of 3,080 subjects participated in a follow-up examination of S4 in 2006/08 (KORA F4). For the genome-wide KORA F4 study we selected 1,814 subjects of these participants. Informed consent has been given. The local ethical committee has approved the study. The KORA S3 and S4 samples do not overlap.

Blood count measurements. DNA was extracted from fresh blood, and was stored at -80°C. FBCs were performed on fresh venous EDTA-anticoagulated blood using an automatic blood counter (Beckman Coulter LH 750).

Lolipop

Sample. London Life Sciences Population (LOLIPOP) study is an ongoing population based cohort study of ~30,000 Indian Asian and European white men and women, aged 35-75 years, recruited from the lists of 58 General Practitioners in West London, United Kingdom. Response rates have averaged 62%; there are no material differences between responders and non-responders with respect to age, sex, co-morbidity and available risk factors. Five studies were included in the analysis:

1. LOLIPOP_EW_A: European whites from the general population
2. LOLIPOP_EW_P: European whites from the general population
3. LOLIPOP_IA300: Indian Asians (enriched for insulin resistance and component phenotypes [9].
4. LOLIPOP_IA610: Indian Asians, CHD case-controls, so corrected for CHD and cohort (for possibly different time of recruitment)
5. LOLIPOP_IA_P: Indian Asians from the general population

Blood count measurements. Venous blood was anticoagulated with K2 EDTA and transferred in 4 ml BD Vacutainer Rapid Serum Tubes. Full blood counts were performed using a XE2100 automated hematology analyser (Sysmex, Kobe, Japan).

Samples were transferred at room temperature averaging (26.85 °C) with measurements performed within 24 hours from venesection.

LBC1921

Sample. The Lothian Birth Cohort 1921 (LBC1921) cohort consists of 550 relatively healthy individuals, 316 females and 234 males, assessed on cognitive and medical traits at 79 years of age. They were born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) in Scotland. A full description of participant recruitment and testing can be found elsewhere [10]. Ethics permission for the study was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LREC/1998/4/183). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent.

Blood count measurements. Venous blood was collected in 2.7ml Sarstedt tubes and anticoagulated with EDTA. Full blood counts were performed on the same day using a Coulter LH 750 Haematology Analyser (Beckman Coulter Inc, Milton, UK).

LBC1936

Sample. The Lothian Birth Cohort 1936 (LBC1936) consists of 1,091 (548 men and 543 women) relatively healthy individuals assessed on cognitive and medical traits at 70 years of age. They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland. A full description of participant recruitment and testing can be found elsewhere [11]. Ethics permission for the study was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and from Lothian Research Ethics Committee (LREC/2003/2/29). The research was carried out in compliance with the Helsinki Declaration. All subjects gave written, informed consent.

Blood count measurements. Venous blood was collected in 2.7ml Sarstedt tubes and anticoagulated with EDTA. Full blood counts were performed on the same day using a Coulter LH 750 Haematology Analyser (Beckman Coulter Inc, Milton, UK).

MICROS / South Tyrol

Sample. The MICROS study 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. An extensive description of the study is available elsewhere [12]. Briefly, 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. Information on the health status of participants was collected through a standardized questionnaire. Laboratory data were obtained from standard blood analyses. Genotyping was performed on just under 1,400 participants with 1,334 available for analysis after data cleaning.

Blood count measurements. Venous blood was anticoagulated with EDTA and FBCs were performed using an HMX IL (Beckmann Coulter) on average <5 hours from venesection.

NFBC1966

Sample. The North Finland Birth Cohort of 1966 (NFBC1966, n=12,058 live born) 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 as is the final study sample in the present analyses. A total of 4,763 genotyped samples were available from the NFBC1966.

Blood count measurements. FBCs were performed on fresh blood immediately or within 24 hours using a Coulter.

NTR

Sample. Participants were registered with the Netherlands Twin Register (NTR) and drawn from the GAIN-MDD study, which is a case-control study of major depressive disorder in unrelated individuals aged 18-77 years. Participants gave informed consent to participation, and the study was approved by an ethic committee.

Blood count measurements. Biological samples were taken at the respondents' home between 07.00 and 10.00 h. FBC were performed within 6h of blood collection using a Coulter instrument.

POPGEN

Sample. DNA samples of 1,228 unrelated individuals were obtained from the PopGen biobank [13]. All individuals had white skin colour and were of German descent, i.e. declared that they, their parents and their grandparents were born in Germany. Written, informed consent was obtained from all study participants and the institutional ethics committee approved all protocols.

Blood count measurements. Venous EDTA blood samples were analyzed immediately after venesection by use of the hematology automated analyser Sysmex XE-2100.

QIMR

Sample. FBC were obtained from 2,538 adolescent twins from 1,089 Australian families ascertained from the general population. Twins were enlisted through primary schools, media appeals and by word of mouth and tested longitudinally as close as possible to their twelfth, fourteenth and sixteenth birthdays in the context of an ongoing study of melanocytic naevi as described in detail elsewhere [14,15]. Participants (and where appropriate their parents or guardians) gave informed consent to participation, and appropriate ethics committees approved all studies

Blood count measurements. The clinical protocol used for blood collection and processing has been described in detail previously [16]. Briefly, venous blood samples from the twins and, where possible, from their parents and siblings, were collected for hematological assessment (twins and sibs only) and DNA genotyping. FBC were

obtained within 24h after venesection using a Coulter (Model STKS) instrument. For each trait, outlier observations (6 SD above the mean) at each time of assessment (ages 12, 14 and 16) were excluded from analysis and the average across all available time points was computed.

SardiNIA

Sample. We recruited and phenotyped 6,148 individuals, males and females, ages 14–102 yr, from a cluster of four towns in the Ogliastra province of Sardinia [17]. The local ethical committee approved the study protocol and all participants provided a written informed consent.

Blood count measurements. During physical examination, a blood sample was collected from each individual, and divided into two aliquots; one was used for genomic DNA extraction and the second aliquot to characterize several blood phenotypes, including evaluation of platelet count by using the Beckman COULTER LH 750 Series hematology systems according to manufacturer's instruction.

SHIP

Sample. The Study of Health in Pomerania (SHIP) is a longitudinal population-based cohort study conducted in West Pomerania, the north-east area of Germany [18]. For the baseline examinations, a sample of 6,265 eligible subjects aged 20 to 79 years was drawn from population registries. Only individuals with German citizenship and main residency in the study area were included. Selected persons received a maximum of three written invitations. In case of non-response, letters were followed by a phone call or by home visits if contact by phone was not possible. The SHIP population finally comprised 4,308 participants (response 68.8%). Baseline examinations were conducted between 1997 and 2001.

Blood count measurements. Non-fasting blood samples were taken in the supine position. The blood count was measured within 60 minutes: erythrocytes, hemoglobin, hematocrit, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), platelet count (PLT), mean platelet volume (MPV) and leukocytes. Samples were analyzed either at the hospital laboratory in Greifswald with a Coulter Max M analyzer (Coulter Electronics, Miami, USA) or at the hospital laboratory in Stralsund with a Coulter T660 analyzer (Coulter Electronics, Miami, USA). Both analyzers were calibrated and maintained according to the manufacturers' instructions. Quality control was performed internally as well as externally by participating in external proficiency testing programs.

SORBS

Sample. All subjects are part of a sample from an extensively phenotyped self-contained population from Eastern Germany, the Sorbs [19,20]. At present, about 1000 Sorbian individuals are enrolled in the study. Sampling comprised unrelated subjects as well as families. Extensive phenotyping included standardised questionnaires for past medical history and family history. 890 subjects were available for the present study. The ethics committee of the University of Leipzig approved the study and all subjects gave written informed consent before taking part in the study.

Blood count measurements. Venous EDTA blood samples were analyzed by use of the haematology automated analyser Sysmex XE-2100.

TwinsUK 317k and 610k

Sample. The TwinsUK cohort is an adult twin British registry shown to be representative of singleton populations and the United Kingdom population [21]. The 317k set was used for Stage 1 and included 1,460 (PLT, 100% females) and 1,082 (MPV, 100% females) individuals; the 610k set was used for Stage 2 and included 1,553 (PLT) and 945 (MPV) individuals respectively (82% females). Ethics approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant to the study.

Blood count measurements. Venous blood was anticoagulated with EDTA and FBCs were performed using either an ADVIA 2120 Hematology System (Siemens Healthcare Diagnostics, Deerfield, IL, US) or a XE2100 automated hematology analyser (Sysmex, Kobe, Japan) on average within 24 hours from venesection (range 20 - 30 hrs). The two instruments displayed differences in measurements range, with means (SD) of 9.69 (0.96) and 11.17 (1.03) respectively. Hence association analyses were adjusted for instrument type.

UKBS-CC1 and UKBS-CC2

Sample. The UK Blood Services (UKBS) Common Controls Panel 1 and 2 (UKBS-CC1 and UKBS-CC2) collection is a national collection of 3,000 DNA samples from the 12 health regions of Great Britain established in 2005-2006 by a partnership between NHS Blood and Transplant (NHSBT) of England, the Scottish National Blood Transfusion Service and the Welsh Blood Service. The Common Controls collection was established for use as the shared controls in the WTCCC Genome-Wide Association Studies (GWAS) [22]. The English samples from both panels were used in this study.

Blood count measurements. Full blood counts (FBCs) were obtained from EDTA anticoagulated samples of blood drawn from the pouches of the donation collection sets. Samples were transferred at room temperature to a single testing centre in Cambridge and FBCs were measured on a Beckman-Coulter GenS. Measurements were performed between 16-24 hours after phlebotomy.

Replication

Amish study

Samples. 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 (reference below). Participants were relatively healthy volunteers from the Old Order Amish community of Lancaster County, Pennsylvania and their family members [23,24]. Examinations were conducted at the Amish Research Clinic in Strasburg, PA. The Institutional Review Board at the University of Maryland approved all protocols and informed consent was obtained, including permission to use their DNA for genetic studies. Study participants were enrolled within the 2000-

2008 time period. Of the total phenotyped participants, a total of 1,578 had FBC measures and genotype information.

Blood count measurements. The clinical protocol used for blood collection and processing has been described in detail previously [24]. Briefly, venous blood samples from all participants were collected for hematological assessment and DNA genotyping. FBC processing was completed within 24h after venesection. For each trait, outlier observations (6 SD above the mean) at each time of assessment were excluded from analysis.

CBR

Sample. The Cambridge BioResource (CBR) is a collection of pseudo-anonymised DNA samples from 8,000 healthy blood donors that has been established in 2008 and 2010 by the NIHR funded Cambridge Biomedical Research Centre in collaboration with NHSBT for use in genotype-phenotype association studies. Four donors each were enrolled during 2007 and 2009.

Blood count measurements. Full blood counts (FBCs) were obtained from EDTA anticoagulated samples of blood drawn from the pouches of the donation collection sets. FBCs for the first 4,000 samples were performed on an ABX Pentra 60 automated hematology analyser (ABX Diagnostics, Montpellier, France). FBCs for the final 4,000 samples have been performed on a Sysmex XE-2100 and for the purpose of calibration measurements on 500 blood samples were performed on both the Coulter and Sysmex instruments. Measurements were performed between 16-24 hours after phlebotomy.

Cleveland Clinic GeneBank

Sample. The Cleveland Clinic GeneBank study is a hospital-based angiographic study of ~10,000 subjects that has been used previously for discovery and replication of novel genes and risk factors for atherosclerotic CVD [25-28]. Briefly, subject recruitment into GeneBank occurred between 2004 and 2008 and provides an ongoing focus for analyzing the association of biochemical and genetic factors with coronary atherosclerosis in a consecutive cohort of patients undergoing elective cardiac evaluation. Enrollment criteria included stable patients undergoing elective coronary angiography without known myocardial infarction at time of enrollment and ability to give informed consent. Extensive clinical, demographic, laboratory and angiographic data were collected from electronic medical record. Ethnicity information was self-reported. All patients provided written informed consent prior to being enrolled in GeneBank and the Institutional Review Board of the Cleveland Clinic approved the study. For this study we used only healthy controls.

Blood count measurements. Fasting blood samples were collected via arterial sheath prior to commencement of angiography. Platelet number and mean platelet volume were determined within 6 hr of blood draw using an ADVIA 2120 hematology analyzer, which is a flow cytometry-based system that provides a complete blood cell count, a white blood cell differential, and a reticulocyte count.

DESIR

Sample. DESIR is a French cohort from the general population (http://ifr69.vjf.inserm.fr/~desir/context_study.htm). 716 individuals were

genotyped, 178 men and 538 women. Written informed consent was obtained from every participant to the study.

Blood count measurements. Blood was anticoagulated with EDTA. Blood count measurements were performed using either a Technicon H3RTX (Bayer Diagnostics), Puteaux, France or a JT2 analyser (Beckman/Coulter), Roissy, France or an Argos from ABX, Montpellier, France.

INGI Cilento

Sample. INGI Cilento is a population-based study of isolated populations located in the area of the National Park of Cilento e Vallo di Diano. The study includes 2,137 individuals, among them 855 having both phenotype and genotype data were included in this analysis. The ethics committee of Azienda Sanitaria Locale Napoli 1 approved the study design. The study was conducted according to the criteria set by the declaration of Helsinki and each subject signed an informed consent before participating to the study.

Blood count measurements. Blood was anticoagulated with EDTA and FBCs were performed using the automated particle counters Max M analyzer (Coulter Electronics, Miami, USA) (on average within 24 hours from venesection).

GHRAS

Sample. The Greek Health Randomized Aging Study (GHRAS) is a cross-sectional health and nutrition study of an elderly (≥60 years) urban Greek population [29]. The main objectives of the GHRAS are: 1) to record the prevalence of cardiovascular disease, diabetes, and cardiovascular disease risk factors such as hypercholesterolemia, obesity, and hypertension in old and very old subjects, 2) to record the dietary habits of the elderly, and 3) to investigate potential interactions among socioeconomic, lifestyle, dietary, psychological, biochemical, and genetic factors influencing the health status of the elderly. To meet these aims we randomly selected Centers of Open Protection for the Elderly in the Athens region and invited their members to participate in the study. Centers of Open Protection for the Elderly are public entities that provide assistance from social workers, first-degree medical care, and recreational activities for the elderly. The Bioethics Committee of Harokopio University of Athens approved the study protocol and all subjects signed a voluntary consent form. Between 2004 and 2009, 900 volunteers were recruited.

Blood count measurements. Venous blood was anticoagulated with EDTA and FBCs were performed on the same day on a Sysmex automatic analyzer.

LifeLines

Sample. The LifeLines Cohort Study is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the LifeLines project comprises a number of cross-sectional sub-studies which investigate specific age-related conditions. These

include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. A written informed consent was obtained from every participant to the study. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and until August 2010 over 30,000 participants have been included. 3,367 individuals of Caucasian origin were genotyped (1,373 men and 1,994 women).
Blood count measurements. Blood was drawn in BD tubes anticoagulated with EDTA. Blood count measurements were performed using a Sysmex XE2100.

NTR2
Sample. The NTR2 sample was drawn randomly from the NTR-Biobank study [30] optimizing the number of unrelated individuals that were genotyped. All subjects participate in studies of the Netherlands Twin Register (www.tweelingenregister.org). Participants gave informed consent and the study was approved by an ethics committee.
Blood count measurements. Blood counts were obtained using the same protocol described above for the NTR cohort.

OGP -Talana
Sample. Talana is one of ten villages of Ogliastra Genetic Park (OGP), a secluded area of Sardinia, where was conducted a population based cohort study. We recruited and phenotyped 1,035 individuals, 457 men and 578 women that represent 93.6% of resident population. Participants gave a blood sample and underwent anthropometric measurements. For each inhabitant we collected genealogical information dating back to the seventeenth century, medical and pharmacology history data and family history of many diseases. Written informed consent was obtained from every participant in the study.
Blood count measurements. Blood count measurements were performed using Coulter LH Hematology analyzer (Beckman-Coulter, Brea, CA).

SANQUIN-CC
Sample. The Sanquin Common Controls (SANQUIN-CC) collection from 1,231 blood donors was established in 2006 as part of the Bloodomics Consortium and subjects have been recruited from the north-west region of the Netherlands. Participating subjects were recruited at routine Sanquin Blood Bank donation sessions.
Blood count measurements. FBCs were measured from an EDTA anticoagulated blood sample taken from the arm prior to 500 ml full blood donation. Samples were transferred to a single testing centre in Amsterdam and FBC measurements were performed within 4 hours post venesection on a Sysmex XT-2000i instrument.

THISEAS
Sample. The Interactions of SNPs and Eating Association Study (THISEAS) is a cross-sectional CAD case-control cohort, including 1,000 cases and 1,000 controls recruited from an urban Greek population. The Bioethics Committee of Harokopio University of Athens approved the study protocol and all subjects signed a voluntary consent form. Only CAD-free subjects were included in the current work.

Blood count measurements. Venous blood was anticoagulated with EDTA and FBCs were performed on the same day on a Sysmex automatic analyzer.

BioBank Japan

Sample. BioBank Japan Project started in 2003 and have so far collected up to 30,000 cases consisting of 47 diseases in the bank (http://biobankjp.org). A total of 14,697 subjects (9,605 males and 5,092 females) were enrolled in the study and available with the data of PLT, RBC, WBC, Hb, MCH, MCHC, and MCV. All participants provided written informed consent as approved by the ethical committees of the Center for Genomic Medicine, RIKEN and the Institute of Medical Science, the University of Tokyo. Clinical information for the samples in BioBank Japan is collected and updated annually by a self-reporting uniform questionnaire (for birth year, height, weight, and smoking and drinking habits) and from medical records (for laboratory data including hematological and biochemical traits). Since the study population consisted of disease cases, blood counts were adjusted for the affection status of the diseases in addition to age and gender in the association analysis.

Blood count measurements. Full blood counts were performed on fresh blood immediately or within 24 hours using a Coulter.

Study design

The 39 studies participating in the primary meta-analysis, *in silico* replication and *de novo* replication are described in **Table S1** and in further detail in the previous section. Most cohorts include healthy individuals from the general population and of European ancestry, apart for four cohorts of South and East Asian ancestry used for inter-ethnic comparisons. Local ethic committees for all studies granted approval to the study, and all study participants gave informed consent.

For the analysis of platelet counts and volume we used a standard two-stage design:

Stage 1. A discovery set consisting of 48,666 participants from 23 studies with genome-wide data was used for initial screening of genetic loci associated with PLT. A subset of 13 cohorts with 18,600 individuals had also MPV available (**Table S1**). 4,627 of these individuals from KORA, SHIP, UKBS and TwinsUK were already used in a previous GWAS on MPV and PLT [31]. Associations were considered genome-wide significant when the combined p-value across all samples was $5x10^{-8}$ or less, which corresponds to a Bonferroni correction for the estimated one million independent common variant tests [32].

Stage 2. We further applied the following criteria to prioritise genomic regions for replication using existing GWAS (*in silico*) and newly genotyped (*de novo*) cohorts: i) stage 1 meta-analysis $P \leq 10^{-6}$, ii) SNPs were either unlinked, or located at distances ≥ 500 Kb from the nearest association signal. Stage 2 analysis included 18,838 individuals from 12 additional population-based studies. All had measures of PLT available and 11,594 individuals from 8 studies had in addition measures of MPV. Nine cohorts with available GWAS data (10,773 individuals) contributed to *in silico* replication of all signals with meta-analysis $P \leq 10^{-6}$; to prioritize resources, we typed in the three *de novo* cohorts (8,065 individuals) only signals reaching suggestive p-values after the first round of replication ($5x10^{-8} \leq P \leq 5x10^{-6}$). We attempted to replicate 138 SNPs with a stage 1 meta-analysis $P \leq 10^{-6}$ in 10,773 samples from the 9 *in silico* replication cohorts. SNPs that did not reach the cutoff for genome-wide significance after combined discovery and *in silico* replication analysis (N=55) were then genotyped in a further 8,065 participants from the 3 *de-novo* replication cohorts. All SNPs previously described as associated with PLT and MPV[31] were confirmed in the larger sample of this study.

Genotyping and imputation

GWAS studies. All cohorts used commercially available Affymetrix or Illumina DNA arrays. Quality control on chip level was performed independently for each study. Details on genotyping platforms and data QC used for this analysis by individual studies are reported in **Table S1** and in the **Supplementary Information**. To facilitate meta-analysis, each group performed genotype imputation using either IMPUTE [33], BIMBAM [34] or MACH [35] based on the HapMap Phase II European-American CEU reference panel [36] with parameters and pre-imputation filters as specified in **Table**

S1. Before conducting the meta-analysis we excluded all SNPs with a minor allele frequency (MAF) <0.01. Moreover, imputed SNPs were excluded if the cohort-specific imputation quality score as assessed by r^2.hat (MACH) or .info (IMPUTE) metrics was <0.40; directly genotyped SNPs were excluded if call rate < 0.90 or Hardy-Weinberg $P < 1x10^{-6}$. As is standard for GWAS, we excluded all X-linked SNPs owing to the following reasons: i) the X chromosome has to be treated differently from the autosomes; ii) it cannot be predicted which allele is active, iii) testing males separately results in different sample sizes and power. In total, 2.71 million genotyped or imputed SNPs for PLT and 2.69 million SNPs for MPV were analyzed in the meta-analysis. For *in silico* replication within studies with genome-wide data we used only summary statistics of SNPs in the loci of interest derived from the discovery stage. SNPs were successfully genotyped or imputed, and common quality control filters were applied to each study as specified in **Table S1**.

De-novo genotyping. For stage 2 studies without genome-wide data (CBR, GHRAS and Sanquin Common Controls) the genotypes for the loci of interest were obtained using Sequenom iPLEX according to standard protocols. For the *de-novo* genotyping cohorts, similar filters were applied as specified in **Table S1**: minimal criteria were per-sample and per-SNP call rates > 80% and HWE *P*< 0.001 for all SNPs in all cohorts.

Statistical and bioinformatics analyses

Associations with hematological traits

Each study in stage 1 and stage 2 performed association analyses according to a unified analysis plan. In each study, each genotyped or imputed SNP was tested for association using an additive genetic model. Linear regression models on natural log-transformed (MPV, ln fl) or untransformed (PLT, $1x10^9$/l) dependent variables with adjustments for age, sex and other cohort-specific covariates were employed for studies of unrelated individuals. Linear mixed effects models were used to account for family structure in family-based studies (for details see **Table S1**).

Meta-analyses

To combine association results for both PLT and natural logarithm of MPV across the 23 stage 1 studies as well as 12 stage 2 studies into stage-specific meta-analysis statistics we performed meta-analyses. For stage 1 and stage 2 meta-analyses as well as combined stage 1+2 meta-analyses, we used fixed-effects inverse variance meta-analysis method in which for each SNP a weighted z-statistic was calculated, where weights were proportional to the inverse square of the standard errors of the regression coefficients examined in each sample and selected such that the squared weights sum would be 1. These weighted statistics summarize the overall magnitude and direction of effects relative to a pre-selected reference allele. Calculations were implemented using the METAL package. To assess heterogeneity within European populations in effect direction and size, we estimated heterogeneity p-values based on Cochran's Q statistic.

Population stratification and p-value inflation

P-values of each study were corrected by genomic control inflation factor λ before inclusion into the meta-analysis. This accounts for minor to medium p-value inflation due to residual population structure or other unconsidered confounding factors. The estimated genomic control parameters λ were small for all individual GWAS as well as for the discovery meta-analysis (**Table S1**), suggesting little residual confounding due to population stratification. After meta-analysis, the genome-wide λs for PLT and MPV were 1.079 and 1.039 respectively.

Conditional analysis for detection of secondary association signals

To identify additional independent platelet-associated SNPs at each of the genome-wide significant loci of the discovery stage, we performed an additional genome-wide association analysis for PLT and ln(MPV) including all 23 stage 1 cohorts. Each cohort included genotypes or imputed dosages for all sentinel SNPs selected in the discovery stage meta-analysis as additional covariates to the basic model adjusted for sex, age and cohort-specific adjustments. When data for a sentinel SNP were unavailable, high linkage disequilibrium (LD) proxies were included instead. Association results for each study were combined using fixed-effects meta-analysis as described before. One signal reached genome-wide significance in the secondary meta-analysis, i.e. rs3811444 in *TRIM58* (r^2 with rs7550918 at LOC148824 = 0.002).

Calculation of MPV and PLT genotype score, observed and expected explained variance

Explained variances were estimated independently in the four stage 2 cohorts TwinsUK 610K, Lifelines, NTR2 and Amish study. In each study we calculated a single predictor y_hat for PLT and ln(MPV), where y_hat is the weighted sum over the allelic dosages of the 55 genome-wide significant SNPs for PLT and 29 genome-wide significant SNPs for ln(MPV), respectively. The weights are the effect estimates of the SNPs in the stage 1 discovery meta-analyses (**Table S2**). The explained variance is calculated independently in the four replication cohorts as the adjusted R^2 of the regression of y onto y_hat. We defined a genotype score from the same leading SNPs of MPV and PLT as a weighted sum of the allelic dosages (genotyped or imputed), where the sum of the weights was set to be the number of SNPs and the weights were proportional to the estimate of the effect size for each SNP (effect estimates from the stage 1 discovery meta-analyses). In the population-based TwinsUK 610K study, we estimated the increase per allele by regressing MPV and PLT on the respective genotype score. Furthermore, we applied the method of Park et al.[37] to predict the number of additional PLT- and MPV-associated loci with similar effect sizes to those observed in the GWAS, and the relative variance explained by common variants. The method uses summary statistics of the PLT- and MPV meta-analyses to estimate the total number of susceptibility loci. For the estimation procedure we used unbiased effect sizes of all loci with a p-value $P < 0.05$ in stage 2, minor allele frequencies from the HapMap CEU population and the estimated power

to detect genome-wide signals in stage 1. This method has the limitation that it accounts only for the range of effect sizes that are observed in our study, and thus it provides a conservative estimate of the explained variance and number of additional susceptibility loci. The R-package *INPower* for calculations can be downloaded at http://dceg.cancer.gov/about/staff-bios/chatterjee-nilanjan; power calculations have been performed by using Quanto version 1.2.4 (http://hydra.usc.edu/gxe/).

Analysis of platelet SNPs in non-European samples

To investigate the relevance of our findings in non-European populations, sentinel SNPs reported in **Table S3** were analysed in 9,308 South Asians from LOLIPOP and 14,697 East Asians from the BioBank Japan. Association testing was performed for each SNP-trait pair from **Figure S1**, using the same association testing strategy applied to the primary European samples. The pre-specified statistical significance threshold for replication in the non-European groups was $P \leq 7x10^{-4}$ to account for multiple testing (68 loci tested; **Table S3**). To assess heterogeneity between European and Asian populations in effect direction and size, we estimated heterogeneity p-values based on Cochran's Q statistic. Although we note that these statistics might have formally low power in this study. To assess whether the observed concordance between effect directions in the non-European groups and the primary meta-analysis cohorts was due to chance, we tested the overall number of concordant SNPs, regardless of p-value in the group, via a binomial test with a null expectation of P=0.5.

Analysis of pleiotropic effects with erythrocyte traits

For exploration of pleiotropy at the platelet loci, we tested associations of the 85 lead PLT/MPV SNPs with the following three additional erythrocyte traits: Hemoglobin concentration (Hb, g/dl), Mean Erythrocyte Volume (MCV, fL), Erythrocyte Count (RBC, $1x10^{12}$/l). The analysis was done in the HaemGen RBC consortium that is a global meta-analysis of genome-wide association studies investigating genetic factors influencing levels of haemoglobin and other red-blood cell indices. Thirty-two studies with erythrocyte traits including 62,625 individuals contributed to this meta-analysis. The individual studies are ALSPAC (N = 2,526), Amish Study (N = 1,578), INGI Carlantino (N = 520), Cleveland Clinic GeneBank (N = 2,671), INGI Cilento (N = 855), CoLaus (N = 854), DESIR (N = 716), EGCUT(N = 893), EPIC case (N = 844), EPIC cohort (N = 1,847), INGI FVG (N = 1205), INGI Val Borbera (N = 1,662), KORA F3 (N = 1,642), KORA F4 (N = 1,813), LBC1921 (N = 496), LBC1936 (N = 987), Lifelines (N = 8121), LOLIPOP_EW610 (N = 927), LOLIPOP_EW_A (N = 589), LOLIPOP_EW_P (N = 652), MDC (N = 1,699), MICROS/South Tyrol (N = 1,213), NFBC1966 (N = 4,761), PREVEND (N = 3,152), SardiNIA (N = 4,694), SHIP (N = 3,183), SORBS (N = 934), TwinsUK (N = 3,419), UKBS-CC (N = 2,155), NESDA-NTR-QIMR (N= 6,017). Genome-wide association scans were carried out using commercial arrays, with imputation of missing genotypes to HapMap2. Samples with extreme values (+/- 3SD) were excluded, as were SNPs with low minor allele frequency (<1%), call rates (<95%) or imputation quality (r^2.hat (MACH) or .info (IMPUTE) metrics was <0.40). Phenotypic associations were tested in each cohort separately by testing an

additive genetic model. Linear regression models on untransformed traits with adjustments for age, sex and other cohort-specific covariates were used for studies of unrelated individuals. Linear mixed effects models were employed to account for family structure in family-based studies. Study specific statistics were combined by fixed-effects inverse variance meta-analysis implemented in METAL. Principal components were used to adjust for population substructure. Genomic control correction was applied to the results for individual cohorts, and then to the results of meta-analysis. Results of these analyses are provided in **Table S8.**

Genomic Annotation of PLT/MPV association signals

We observed that many of the SNPs with genome-wide significant evidence for association with MPV/PLT levels were in genes (46 out of 68). Therefore, we sought to formally test the hypothesis that SNPs more strongly associated with PLT levels were more likely to overlap genes than those that associated less strongly with PLT levels. To test this hypothesis, we focused on 1,842,709 SNPs that were tested in at least 21 of the 23 cohorts that were included in the PLT meta-analysis described in the main text, including the QIMR cohort. These were then analysed as follows:

1. We grouped SNPs located within 1,000 kb of each other into individual groups (or "clumps") based solely on pair-wise linkage disequilibrium estimated based on genotype data for 700 founders from the QIMR GWAS study. This was performed by modifying the –clump routine implemented in PLINK to use two r2 thresholds (rather than a single threshold): an exclusion (0.1) and an inclusion (0.8) threshold. PLT association P-values were not considered to group SNPs. For example, starting with the first SNP on chromosome 1 (index SNP for the first clump), we (a) identified nearby SNPs with r2≥0.8 with that index SNP and added those to the same LD clump; and (b) identified nearby SNPs with an r2 between 0.1 and 0.8 with that first index SNP and dropped them from subsequent analyses. The remaining nearby SNPs had an r2<0.1 with the index SNP of that first clump. We then moved on to the first of those remaining SNPs, which became the index SNP of the second LD group, and repeated (a) and (b) above. This procedure was repeated until all 1.8 million SNPs had been either grouped (n=471,977) or dropped from subsequent analyses. As a result, we created 71,318 largely independent groups of highly correlated SNPs, with a mean number of SNPs per group of 6.6 (range 1 to 508; median = 3).
2. For each group of SNPs, we determined the lowest and highest genomic coordinate, which corresponded to the physical position of the first and last SNP in the group. For example, a clump on chromosome 1 included 10 SNPs that mapped to a 33 kb interval (or segment) flanked by rs3766180 (1468016 bp) and rs7519837 (1500664 bp). The average segment length across the 71,318 groups of SNPs was 23 kb (range 0 to 1009 kb, median = 7 kb).
3. For each group of SNPs, we determined whether the corresponding chromosomal segment overlapped a known gene, using the first and last exonic base of the longest isoform +/- 50 kb as the gene boundaries (based on the March 2006 UCSC assembly). In this way, each group of SNPs was classified as overlapping a gene or not (GENE indicator). Of the 71,318 groups of SNPs, 43,971 (62%) overlapped at least one gene.

4. We assigned to each group of SNPs a single PLT association P-value, specifically the P-value for the SNP showing the weakest association in the meta-analysis.
5. Each of the 71,318 groups of SNPs was then assigned to a vintile (V) based on the PLT association P-value. For example, the top vintile (V1) had 3,566 SNPs with a PLT association P-value ranging between 10-22 and 0.06, while the bottom vintile (V20) had 3,566 SNPs with a PLT association P-value between 0.9876 and 1.000.
6. Lastly, we used linear regression to test whether the GENE indicator was a significant predictor of V (ie. of the level of association between that group of SNPs and PLT levels). We included in the regression model two potential confounders, namely the minor allele frequency of the SNP (average across all cohorts tested) and imputation confidence metric (average across all cohorts). We also repeated steps (5) and (6) after excluding groups of SNPs with a PLT association P-value <10-5, <0.001, <0.01 and <0.1.

Definition of core genes

To explore biological properties of genes closely associated with the 68 MPV/PLT associated regions, we focused on a set of genes ('core genes') that were intimately associated with the sentinel SNP at each locus. We applied the following criteria for choice of the genes: (i) We first searched all signals for cases where If the sentinel SNP mapped to within a gene, defined as spanning from the transcription start site at 5' to the last nucleotide of the last 3' annotated exon. We identified 41 genes at 41 loci that met these criteria. (ii) We then screened the remaining loci where the sentinel SNP was intergenic to two genes, and selected genes that mapped to less than 10 kb from the sentinel SNP (15 genes at 13 additional loci). A total of 56 genes were selected using these criteria. From the 14 remaining association intervals, no core genes were selected because the sentinel SNP was not within the +/- 10 Kb interval of a gene.

Two genes in the HLA locus (*HLA-B* and *HLA-DOA*) on chromosome 6p21.3 were further removed from the list of core genes (leaving 54 total core genes) to eliminate likely unspecific association signals reflecting i) low biological candidacy; ii) well-documented extended LD beyond the HLA class I and II loci and high gene density makes association patterns in this region difficult to interpret; iii) no expression of *HLA-DOA* in any of the seven blood cell elements of the HaemAtlas other than in B-lymphocytes and no expression in endothelial cells. The core genes are described in **Table 2**.

Canonical pathway analyses

Ingenuity Pathway Analysis (IPA). The core analysis in IPA allows interpretation of the genes in the context of established biological pathways. We ran core analysis of three different sets of genes:

1) A set of genes, defined by the most associated SNP being within ±10Kb of the gene. This set of core genes was used to explore the physical relatedness of

association signals with genes selected purely based on their physical proximity to the association signal.

2) The core genes defined in 1, supplemented by their first order interactors identified as described in the next section.

We limited our analyses to direct relationships annotated within the Ingenuity Knowledge Base reference set (Genes only). The most significant result per set of genes is given in **Table S7**.

Protein-protein interaction (PPI) network

The proteins encoded by the 54 core genes were used as primary baits to develop the PPI network, and the corresponding UNIPROT protein identifier was obtained. To develop a system level network centered on the core proteins we initially searched for first order interactors of the 54 core proteins in public databases. Two different types of resources were used for this initial effort, Reactome and IntAct and related databases. Interactions were initially based on curated literature within the reactome project (www.reactome.org). Reactome comprises a combination of protein protein interactions based on reactions that occur or are mediated by specific proteins and by general reactions that occur or are mediated by classes of proteins (e.g. rho gtpases or tyrosine phosphatases). Secondly, these data were complemented with experimentally identified protein protein interactions obtained from data within the imex consortium (www.imexconsortium.org) - the great majority of data present in the IntAct database. Thus our network is constructed using both general and specific types of reactions. Using the former class of general interactions has the advantage of including proteins with known class but unknown interaction partners, with the assumption that at least one of the proteins in the interacting class is a physical interaction partner of the core protein. This expansion in the number of links improves our ability to identify pathways and propose mechanisms of action with the caveat of loosing information about the real number of interaction partners in the network. Therefore we interpret the network observed in terms of the classes of reactions that our core proteins take part in, and not in terms of the connectivity of such nodes.

Reactome. Reactome (http://www.reactome.org/) is an open source database and website for the exploration and analysis of human biological pathways [38,39]. Reactome contains core datasets for systems biology. Pathways are built from biological reactions that are connected as steps in the pathway. Each reaction describes a biological event, e.g. binding, phosphorylation, transport, or enzymatic event. Reactome content is based on information provided by expert biologists, peer-reviewed to ensure the resulting pathways represent the biological consensus. Every reaction is supported by published experimental data. Pathways are human-centric but may incorporate data from model organisms in the case of conserved functions, although these reactions are clearly differentiated from those that were experimentally determined in humans. Reactome covers many areas of biology such as DNA replication and repair, membrane trafficking, synaptic transmission and receptor-based signaling pathways. Each topic is represented as a hierarchy of pathway diagrams. Pathways relevant to megakaryocyte and platelet biology are

largely within the topic Hemostasis, which at the time of writing (October 2010) contained 43 pathways and 276 reactions. Examples include adhesion to exposed collagen, nitric oxide metabolism, ADP signaling through P2Y purinergic receptors, thrombin activation of proteinase activated receptors, GPVI mediated signaling, aIIbb3 integrin signaling, platelet calcium regulation and platelet degranulation.

PPI in Reactome are classified as: i) *direct complex*: proteins that are part of a complex, and therefore assumed to contact each other (this may not always be the case); ii) *indirect complex*: proteins in different sub-complexes of a complex; iii) *reaction*: proteins that participate in the same reaction. Excludes any protein-protein pair that is in a direct complex, and there is no guarantee that proteins involved in a reaction physically interact; iv) *neighbouring reaction*: connects proteins that participate in consecutive reactions, where the output of a reaction is the input or catalyst for another reaction. Trivial connections due to common small molecules like ATP are prevented by the use of an exclusion list. This interaction type only considers reactions that are connected in a pathway; it excludes cases where the output of a reaction is the input to another, but the reactions are not consecutive pathway steps. 658 edges were identified in Reactome.

IntAct. The same 54 core proteins were also used as bait in a second trawl for interactors by querying the literature-curated interaction databases IntAct [40], MINT [41], InnateDB [42], and other IMEx consortium partners, to develop a system level network. These databases are manually curated and follow the same IMEx level curation standards (http://www.imexconsortium.org/), hence the data from these searches can be pooled. Spoke-expanded complexes and co-localisation studies were excluded from the resulting network, thus removing spurious interactions. Only clustered non-redundant first level interactions between human proteins were used as a conservative approach to minimise the inclusion of false interactions. 132 edges were identified by the IntAct-like searches.

Literature curation. Finally, the PPI network was completed by a manual literature-based curation of the 54 core proteins. Systematic searches were performed using the current gene/protein names and the historic names in the context of the following terms, hematopoiesis, megakaryopoiesis, platelets, interactions with and others. 37 edges were identified by the systematic literature searches.

Transcript profiling of blood cells, blood cell precursors and endothelial cells

Data about transcript levels of all genes in the above cells were extracted from a compendium of expression data sets generated by the Bloodomics Consortium and which in part have been released via Array Express at the European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress). Four main datasets have been used: i) the HaemAtlas which encompasses the results of whole genome expression (WGE) studies performed with RNA samples from the eight main blood cell types; ii) the results of a WGE expression study with RNA samples of platelets from 37 healthy individuals; iii) the WGE results obtained with RNA samples of cultured human

umbilical vein endothelial cells (HUVEC); iv) the WGE results obtained with the RNA samples from CD34+ hematopoietic stem cells (HSCs) and from a study of paired HSC cultures that were differentiated towards megakaryocytes (MKs) and erythroblasts (EBs). Cells were harvested during the 10-day culture at five time points in time (days 3, 5, 7, 9 and 10).

HaemAtlas. The purification of blood cells is described in [43]. In short, RNA was isolated from the six main blood cell types (CD4+ Th (CD4, n=7) and CD8+ Tc lymphocytes (CD8, n=7), CD14+ monocytes (CD14, n=7), CD19+ B lymphocytes (CD19, n=7), CD56+ natural killer (NK) cells (CD56, n=7) and CD66b+ granulocytes (CD66, n=7) obtained from seven healthy blood donors and from MKs and EBs. MKs and EBs were obtained by cultures of cord blood-derived $CD34^+$ HSCs. The former were obtained by cultures for 7 days in a medium supplemented with human recombinant thrombopoietin (THPO) and interleukin-1ß (IL1B) and the latter by cultures for 10 days in the presence of erythropoietin (EPO), interleukin-3 (IL3) and stem cell factor (SCF). To ensure high level purity preparations of both MKs and EBs, cells were flow-sorted using monoclonal antibodies against CD markers: MKs were positive for CD41 and negative for CD34; EBs were negative for CD41 and positive for CD235a. RNA was prepared from two sets of four cell preparations each.

Time-course study of MK and EB cultures. To capture changes in transcript levels over time during the proliferation and differentiation of HSCs towards MKs and EBs paired cultures were generated as described above using CD34+ HSCs from three cord blood donations. Umbilical cord blood of three healthy newborns was collected into cord blood collections bags (MacoPharma, Mouvaux, France) after informed consent. CD34+ HSCs were prepared as in [44] and 92%-98% pure CD34+ HSCs were *in vitro* cultured ($1x10^5$/ml) for 10 days in serum-free media (CellGro-SCGM, Cellgenix, France) supplemented with 50 ng/ml THPO (CellGenix, France) and 10 ng/ml IL1B (Miltenyi Biotech, Surrey, UK) to differentiate into MKs. EBs were in vitro derived from HSCs ($5x10^3$/ml) in the presence of 6 U/ml EPO (R&D Systems, Abingdon, UK), 10 ng/ml IL3 (Miltenyi Biotech, Surrey, UK) and 100 ng/ml SCF (R&D Systems, Abingdon, UK). MKs and EBs were harvested at day 3, 5, 7, 9 and 10 as described in [44] and transferred for RNA isolation into a 5 ml tube, centrifuged at 500 x g for 10 min at RT, resuspended in Trizol (Invitrogen, Paisley, UK) and stored at -80°C. Total RNA was isolated according to manufacturer's instructions. The following murine monoclonal antibodies were used for phenotyping of both lineages at day 10: Fluorescein isothiocyanate (FITC) IgG1 isotype control, Phycoerythrin (PE) IgG1 isotype control, Allophycocyanin (APC) IgG1 isotype control, PE-Cy5 IgG1 isotype control and Pacific Blue (PB) isotype control (BD Bioscience Pharmingen™, Becton Dickinson, Oxford, UK), anti-CD11c V450, anti-CD13 APC, anti-CD14 PB, anti-CD15 V450, anti-CD33 FITC (BD Bioscience Pharmingen™), anti-CD34 PE (Beckman Coulter, High Wycombe, UK), anti-CD36 PE, anti-CD41a APC, anti-CD42a FITC, anti-CD235a FITC and anti-CD66c PE (BD Bioscience Pharmingen™). For phenotyping of cells at all other days anti-CD34 PE and anti-CD41a APC, anti-CD42a FITC and anti-CD235a FITC with matching isotype controls were used. In addition, a ploidy stain of the MKs was performed at each day of harvest as described in [45]. In brief, cells were stained with anti-CD41a APC and with matched isotype control and incubated at 37°C for 30 min

in 500 µl PBE buffer containing 0.1% (v/v) Tween 20, RNAse A (0.1 mg/ml) and propidium iodide (0.05 mg/ml, Sigma-Aldrich, Dorset, UK). Samples were analysed using a 9-colour Cyan-ADP flow cytometer running Summit software version 4.3.02 (Beckman Coulter).

Human Umbilical Cord Vene Endothelial Cells (HUVECs). HUVECs for three cultures were isolated from cord venes and cultured as described by [46] in RPMI 1640 supplemented with 20% human serum, penicillin/streptomycin and fungizone. Cells of the separate donations at the second passage after isolation were trypsinized, pooled and seeded in three separate batches in culture medium on gelatin-coated culture flasks and allowed to grow to confluence over a 5-day period. Endothelial cells were harvested and RNA prepared using Trizol (as above).

Platelets. RNA samples were prepared from leukocyte depleted platelet concentrates obtained by apheresis from 37 individuals selected from the platelet function cohort [47]. Platelets were lysed in Trizol and RNA isolated and RT-PCR was used to measure the level of the pan-leukocyte marker CD45 (*PTPRC*) as to quantify possible contaminating leukocytes. The 37 platelet RNA samples were used for expression studies.

Expression data preparation and release. Altogether 137 RNA samples were applied to Illumina Human Expression BeadChips; platelet RNA samples (n=37) were applied to v1.0, HaemAtlas (n=64) and HUVEC (n=3) RNA samples to v2.0 and the HSC (n=3), MK and EB time-course samples (n=30) to v3.0. The array datasets have been deposited in Array Express under accession numbers E-TABM-633 for the HaemAtlas and E-MTAB-374 for platelets. The HUVEC dataset will be released in Array Express as part of this study. The entire dataset for the time-course study will be deposited as part of a separate publication.

Statistical analyses of gene expression. HaemAtlas microarrays [44] were processed as follows. Raw intensity values were summarised with the “beadarray” [48] package in the R/Bioconductor environment. Signal intensities were transformed with a variance stabilising transform (VST) [49] and arrays were quantile-normalised. The VST transform makes use of the multiplicity of individual probes in Illumina arrays and is asymptotically equivalent to a $\log_2$ transform but performs better at low signal intensities. For ease interpretability of the scale in units of fold change we have chosen to label axes as $\log_2$ transformed rather than the more obscure VST transformed. This does not change the observations. Detection p-values were calculated empirically using the negative control probes present in the Illumina gene-expression arrays. Probes were deemed present if the detection p-value was below 0.01. Probe annotations were taken from [50]. When multiple probes were present in the array, the one with the largest range of expression values across all tissues was chosen. The degree of differential expression between MK and all other cells in the HaemAtlas was quantified using the t-statistic from the Welch's two-sample t-test. To determine whether there was evidence of excess over-expression of the core genes relative to all genes in the array, a one-tailed Wilcoxon rank sum test was carried out between the distributions of t-statistics in these two groups. To

quantify whether gene expression was decreasing, increasing or remaining unchanged, we used the t-statistic associated with the test for the slope being significantly different from zero in a simple linear regression. Positive values reflect increasing expression as a function of times, negative values reflect decreasing expression, and values close to zero reflect no change in expression. We then use the values for the 2.5% and 97.5% quantiles of the t distribution with 4 degrees of freedom to classify gene expression patterns into increasing, decreasing or unchanged.

Summary of gene expression patterns of core genes. The HaemAtlas and the time-course series of WGE during MK and EB cultures were used to determine which fraction of the 54 core genes is transcribed in MKs. A probe in the array could be assigned to all 54 core genes in the HaemAtlas and time-course experiments (Illumina V2 and V3 arrays respectively). All but six genes were transcribed at some point in the MK lineage, i.e. from CD34+ progenitors to day 10 MK. The genes for which no transcript could be detected by WGE are *GCKR*, *THPO*, *RCOR1*, *CDKN2A*, *WDR66* and *FLJ36031*. *THPO* and *GCKR* genes are highly transcribed in liver tissue and not in HSC-derived cells. The *FLJ63031* was present in the platelet arrays. Specific TaqMan probes for *RCOR1*, *CDKN2A* and *WDR66* were used to more sensitively measure transcript levels in the day 10 MK RNA samples. Positive results were obtained by TaqMan RT-Q-PCR for all three genes, albeit at a very low level.

Lookup of HaemGen platelet sentinel SNPs in eQTL repositories

We used the recombination interval (NCBI36, **Table S5**) containing each of the sentinel SNPs obtained from our meta-analysis as the input region to search public repositories of eQTL data. From each region, we selected only signals where the lead eQTL SNP was either the lead in the HaemGen platelet analysis, or a proxy defined as having high LD ($r^2 \geq 0.8$) with the HaemGen sentinel SNP in the HapMap2 CEU population. eQTL data was accessed through the eQTL browser at the University of Chicago. (**Table S6**)

Model organisms

D. rerio knockdown models

We selected six genes based on the following criteria: i) the genes contained the most associated SNP at each locus; ii) the genes had an unknown function in hematopoiesis at the time of selection, iii) a reliable *D. rerio* ortholog could be identified with over 50% amino acid sequence identity with its human counterpart, iv) the gene showed a detectable gene expression in human MKs and EBs. General maintenance, collection and staging of the wild-type (Tübingen Long Fin) and transgenic *cd41:GFP D. rerio* lines were carried out as previously described [51]. Morpholino antisense oligonucleotides were obtained from GeneTools LLC (Philomath, OR, USA). The oligos (**Figure S6**) were resuspended in sterile water and approximately 1 nl (3-6 ng) was injected in *D. rerio* embryos, at the one- to two-cell stage. Staining of hemoglobin by o-Dianisidine was performed as previously

described in [52]. In brief, unfixed embryos were stained for 15 minutes in the dark, with a solution consisting of o-Dianisidine (0.7 mg/ml), 0.01 M sodium acetate (pH 4.5), 0.65% hydrogen peroxide and 45% (vol/vol) ethanol. Photomicrographs were taken with a Zeiss camera AxioCam HRC attached to a LeicaMZ16 FA dissecting microscope (Leica Microsystems, Wetzlar, Germany).

D. melanogaster knockdown models

In order to analyze genes in an additional suitable in vivo RNAi model system, we obtained fly lines carrying inducible siRNA constructs from the VDRC [53]. To achieve hemocyte specific knockdowns, flies were crossed to the blood specific hml-Gal4 line driving Gal4 expression under control of a hemolectin promoter [54]. Flies were crossed at 29°C and 10-15 larvae per genotype were analyzed at late larval stage L3 7 days post mating. UAS-GFP allowed for microscopic visualization of plasmatocytes and evaluation of cell size and cell number. Subsequently, larvae were incubated at 60°C for 15 minutes, a process that turns the crystal cells black and allows quantification of crystal cells and melanotic tumor formation with a dissecting microscope. Results were confirmed by replication in a minimum of two additional crosses following the same analytic protocol.

M. musculus knockout models

We screened published literature for *M. musculus* knockout models for each of the 54 core genes, gaining functional support for 15 additional loci (*Par1* [55], *Bcl-x(L)* [56], *c-Myb* [57], *Zfpm2* [58,59], *Plec1* [60], *Dock8* [61], *Fads2* [62], *c-Cbl* [63], *Nfe2* [64], *Lnk* [65,66], *Gp1ba* [67], *Dnam1* [68], *Sirpa* [69], *Thpo* [70] and *Itga2b* [71]).

URLs

Public data repositories and pathway resources

Cytoscape	http://www.cytoscape.org/
FlyBase	http://flybase.org/
GRAIL	http://www.broadinstitute.org/mpg/grail/
GWAS studies catalog	http:// www.genome.gov/gwastudies
HapMap	http //hapmap.org
Haplotter	http://haplotter.uchicago.edu/selection/
IntAct	http://www.ebi.ac.uk/intact/
IPA	http://www.ingenuity.com/
Reactome	http //www.reactome.org/
WTCCC	http //www.wtccc.org.uk/
eQTL browser (U. Chicago)	http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/

Statistical analysis software

BIMBAM	http://stephenslab.uchicago.edu/software.html
IMPUTE	http //www.stats.ox.ac.uk/~marchini/software/gwas/impute.html
INPower	http://dceg.cancer.gov/about/staff-bios/chatterjee-nilanjan
MACH	http //www.sph.umich.edu/csg/abecasis/MACH
MACH2QTL	http //www.sph.umich.edu/csg/abecasis/MACH/download
MERLIN	http //www.sph.umich.edu/csg/abecasis/Merlin
METAL	http //www.sph.umich.edu/csg/abecasis/Metal/index.html
PLINK	http //pngu.mgh.harvard.edu/~purcell/plink
ProbABEL	http //mga.bionet.nsc.ru/~yurii/ABEL
Quanto	http://hydra.usc.edu/gxe/
QUICKTEST	http //toby.freeshell.org/software/quicktest.shtml
R	http //www.r-project.org; whole-genome association analysis
SNPTEST	http //www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html
SNPnexus	http://www.snp-nexus.org/

Supplementary references

1. ARIC. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).
2. Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).
3. Cushman, M., Cornell, E.S., Howard, P.R., Bovill, E.G. & Tracy, R.P. Laboratory methods and quality assurance in the Cardiovascular Health Study. *Clin Chem* **41**, 264-70 (1995).
4. Psaty, B.M. *et al.* The risk of myocardial infarction associated with antihypertensive drug therapies. *JAMA* **274**, 620-5 (1995).
5. Smith, N.L. *et al.* Esterified estrogens and conjugated equine estrogens and the risk of venous thrombosis. *JAMA* **292**, 1581-7 (2004).
6. Heckbert, S.R. *et al.* Antihypertensive treatment with ACE inhibitors or beta-blockers and risk of incident atrial fibrillation in a general hypertensive population. *Am J Hypertens* **22**, 538-44 (2009).
7. Holle, R., Happich, M., Lowel, H. & Wichmann, H. KORA--a research platform for population based health research. in *Gesundheitswesen* Vol. 67 Suppl 1 S19-25 (2005).
8. Wichmann, H., Gieger, C. & Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. in *Gesundheitswesen* Vol. 67 Suppl 1 S26-30 (2005).
9. Chambers, J.C. *et al.* Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. in *Nat Genet* Vol. 41 1170-1172 (2009).
10. Deary, I.J., Whiteman, M.C., Starr, J.M., Whalley, L.J. & Fox, H.C. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol* **86**, 130-47 (2004).
11. Deary, I.J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr* **7**, 28 (2007).
12. Pattaro, C. *et al.* The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC Med Genet* **8**, 29 (2007).
13. Krawczak, M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* **9**, 55-61 (2006).
14. Aitken, J.F., Green, A.C., MacLennan, R., Youl, P. & Martin, N.G. The Queensland Familial Melanoma Project: study design and characteristics of participants. *Melanoma Res* **6**, 155-65 (1996).
15. McGregor, B. *et al.* Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet Epidemiol* **16**, 40-53 (1999).
16. Evans, D.M., Frazer, I.H. & Martin, N.G. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res* **2**, 250-7 (1999).
17. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. in *PLoS Genet* Vol. 2 e132 (2006).

18. John, U. *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. in *Soz Praventivmed* Vol. 46 186-94 (2001).
19. Tonjes, A. *et al.* Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Hum Mol Genet* **18**, 4662-8 (2009).
20. Tonjes, A. *et al.* Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany. *Eur J Hum Genet* **18**, 104-10 (2010).
21. Richards, J. *et al.* Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. in *Lancet* Vol. 371 1505-12 (2008).
22. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
23. 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**, 1016-23 (2008).
24. 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**, 823-8 (2008).
25. Bhattacharyya, T. *et al.* Relationship of paraoxonase 1 (PON1) gene polymorphisms and functional activity with systemic oxidative stress and cardiovascular risk. *JAMA* **299**, 1265-76 (2008).
26. Tang, W.H., Wang, Z., Cho, L., Brennan, D.M. & Hazen, S.L. Diminished global arginine bioavailability and increased arginine catabolism as metabolic profile of increased cardiovascular risk. *J Am Coll Cardiol* **53**, 2061-7 (2009).
27. Tang, W.H. *et al.* Subclinical myocardial necrosis and cardiovascular risk in stable patients undergoing elective cardiac evaluation. *Arterioscler Thromb Vasc Biol* **30**, 634-40.
28. Wang, Z., Tang, W.H., Cho, L., Brennan, D.M. & Hazen, S.L. Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. *Arterioscler Thromb Vasc Biol* **29**, 1383-91 (2009).
29. Kanoni, S. & Dedoussis, G.V. Design and descriptive characteristics of the GHRAS: the Greek Health Randomized Aging Study. *Med Sci Monit* **14**, CR204-12 (2008).
30. Willemsen, G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* **13**, 231-45.
31. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. in *Nat Genet* Vol. 41 1182-1190 (2009).
32. Pe'er, I., Yelensky, R., Altshuler, D.A. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. in *Genet. Epidemiol.* Vol. 32 381-385 (2008).
33. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. in *Nat Genet* Vol. 39 906-13 (2007).
34. Servin, B. & Stephens, M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics* **3**, e114 (2007).

35. Li, Y. & Abecasis, G.R. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. in *Am J Hum Genet* Vol. S79 2290 (2006).
36. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
37. Park, J.-H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. in *Nat Genet* (2010).
38. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* **37**, D619-22 (2009).
39. Vastrik, I. *et al.* Reactome: a knowledge base of biologic pathways and processes. *Genome Biol* **8**, R39 (2007).
40. Aranda, B. *et al.* The IntAct molecular interaction database in 2010. *Nucleic Acids Res* **38**, D525-31 (2010).
41. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Lett* **513**, 135-40 (2002).
42. Lynn, D.J. *et al.* InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol* **4**, 218 (2008).
43. Watkins, N.A. *et al.* The HaemAtlas: characterising gene expression in differentiated human blood cells. in *Blood* Vol. 113(19):e1-9. Epub 2009 Feb 19 (2009).
44. Macaulay, I. *et al.* Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. in *Blood* Vol. 109 3260-9 (2007).
45. van den Oudenrijn, S., von dem Borne, A.E. & de Haas, M. Differences in megakaryocyte expansion potential between CD34(+) stem cells derived from cord blood, peripheral blood, and bone marrow from adults and children. *Exp Hematol* **28**, 1054-61 (2000).
46. Jaffe, E.A., Nachman, R.L., Becker, C.G. & Minick, C.R. Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria. *J Clin Invest* **52**, 2745-56 (1973).
47. Goodall, A.H. *et al.* Transcription profiling in human platelets reveals LRRFIP1 as a novel protein regulating platelet function. *Blood* **Epub Sep 10**(2010).
48. Dunning, M.J., Smith, M.L., Ritchie, M.E. & Tavare, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183-4 (2007).
49. Du, P., Kibbe, W.A. & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-8 (2008).
50. Barbosa-Morais, N.L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res* **38**, e17.
51. Westerfield, M. *The Zebrafish Book.*, (Eugene, OR: University of Oregon Press, 1994).
52. Detrich, H.W., 3rd *et al.* Intraembryonic hematopoietic cell migration during vertebrate development. *Proc Natl Acad Sci U S A* **92**, 10713-7 (1995).
53. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* **448**, 151-6 (2007).
54. Cronin, S.J. *et al.* Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. *Science* **325**, 340-3 (2009).

55. Griffin, C.T., Srinivasan, Y., Zheng, Y.W., Huang, W. & Coughlin, S.R. A role for thrombin receptor signaling in endothelial cells during embryonic development. *Science* **293**, 1666-70 (2001).
56. Mason, K. *et al.* Programmed anuclear cell death delimits platelet life span. in *Cell* Vol. 128 1173-86 (2007).
57. Tober, J., McGrath, K.E. & Palis, J. Primitive erythropoiesis and megakaryopoiesis in the yolk sac are independent of c-myb. *Blood* **111**, 2636-9 (2008).
58. Svensson, E.C., Huggins, G.S., Dardik, F.B., Polk, C.E. & Leiden, J.M. A functionally conserved N-terminal domain of the friend of GATA-2 (FOG-2) protein represses GATA4-dependent transcription. *The Journal of biological chemistry* **275**, 20762-9 (2000).
59. Tevosian, S.G. *et al.* FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell* **101**, 729-39 (2000).
60. Andra, K. *et al.* Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. *Genes & development* **11**, 3143-56 (1997).
61. Randall, K.L. *et al.* Dock8 mutations cripple B cell immunological synapses, germinal centers and long-lived antibody production. *Nature immunology* **10**, 1283-91 (2009).
62. Stoffel, W. *et al.* Delta6-Desaturase (FADS2) deficiency unveils the role of omega3- and omega6-polyunsaturated fatty acids. *EMBO J* (2008).
63. Murphy, M.A. *et al.* Tissue hyperplasia and enhanced T-cell signalling via ZAP-70 in c-Cbl-deficient mice. *Mol Cell Biol* **18**, 4872-82 (1998).
64. Shivdasani, R.A. *et al.* Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* **81**, 695-704 (1995).
65. Tong, Z. *et al.* Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. in *Proc Natl Acad Sci USA* Vol. 105 6998-7003 (2008).
66. Takaki, S. *et al.* Control of B cell production by the adaptor protein lnk. Definition Of a conserved family of signal-modulating proteins. *Immunity* **13**, 599-609 (2000).
67. Bergmeier, W. *et al.* The role of platelet adhesion receptor GPIbalpha far exceeds that of its main ligand, von Willebrand factor, in arterial thrombosis. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 16900-5 (2006).
68. Iguchi-Manaka, A. *et al.* Accelerated tumor growth in mice deficient in DNAM-1 receptor. *The Journal of experimental medicine* **205**, 2959-64 (2008).
69. Yamao, T. *et al.* Negative regulation of platelet clearance and of the macrophage phagocytic response by the transmembrane glycoprotein SHPS-1. *J Biol Chem* **277**, 39833-9 (2002).
70. de Sauvage, F.J. *et al.* Physiological regulation of early and late stages of megakaryocytopoiesis by thrombopoietin. *J Exp Med* **183**, 651-6 (1996).

71. Tronik-Le Roux, D. *et al.* Thrombasthenic mice generated by replacement of the integrin alpha(IIb) gene: demonstration that transcriptional activation of this megakaryocytic locus precedes lineage commitment. *Blood* **96**, 1399-408 (2000).

Acknowledgements

ALSPAC. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council (Grant ref: 74882) the Wellcome Trust (Grant ref: 076467) and the University of Bristol provide core support for ALSPAC.

Amish study. We thank our Amish research volunteers for their long-standing partnership in research, and the research staff at the Amish Research Clinic for their hard work and dedication. We are supported by grants and contracts from the NIH including R01 AG18728 (Amish Longevity Study), R01 HL088119 (Amish Calcification Study), U01 GM074518-04 (PAPI Study), U01 HL072515-06 (HAPI Study), U01 HL084756 and NIH K12RR023250 (University of Maryland MCRDP), the University of Maryland General Clinical Research Center, grant M01 RR 16500, P30 DK072488 from the mid-Atlantic Nutrition and Obesity Research Center , the Baltimore Veterans Administration Medical Center Geriatrics Research and Education Clinical.

ARIC. The Atherosclerosis Risk in Communities Study is supported by National Heart, Lung, and Blood Institute (NHLBI) contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022 and grants R01-HL-087641, R01-HL-59367 and R01-HL-086694; National Human Genome Research Institute contract U01-HG-004402; and NIH contract HHSN268200625226C. The infrastructure was partly supported by grant number UL1-RR-025005, a component of the NIH and NIH Roadmap for Medical Research. The authors thank the staff and participants of the ARIC study for their important contributions. Further, ARIC would like to thank the University of Minnesota Supercomputing Institute for use of the blade supercomputers.

Cambridge Bioresource (CBR). The project made use of NHSBT donors from the Cambridge BioResource (http://www.cambridgebioresource.org.uk/). This local resource for genotype-phenotype association studies is supported by a grant from the National Institute for Health Research (NIHR) to the Cambridge Biomedical Research Centre (which supported JS and SM), and by a NIHR programme grant to NHSBT (RP-PG-0310-1002, which supported AA, AR and WHO).

Cardiovascular Health Study (CHS). This CHS research was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133 and NHLBI grants HL080295, HL075366, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center for Research Resources CTSI grant UL 1RR033176 and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Cleveland Clinic Gene Bank. The Cleveland Clinic GeneBank study is supported by NIH grants P01HL076491, P01HL098055, and R01HL103866.

DESIR. The D.E.S.I.R. study has been supported by INSERM contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis; by INSERM (Réseaux en Santé Publique, Interactions entre les déterminants de la santé), Cohortes Santé TGIR, the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Société francophone du diabète , Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche, Topcon. **Members of the DESIR Study Group:** INSERM U1018: B Balkau, P Ducimetière, E Eschwège; INSERM U367: F Alhenc-Gelas; CHU D'Angers: Y Gallois, A Girault; Bichat Hospital: F Fumeron, M Marre; CHU Rennes: F Bonnet; CNRS UMR8090, LILLE: P Froguel; Centres d'Examens de Santé: Alençon, Angers, Caen, Chateauroux, Cholet, Le Mans, Tours; Institute de Recherche Médecine Générale: J. Cogneau; General practitioners of the region; Institut Inter-régional pour la Santé: C Born, E Caces, M Cailleau, O Lantieri, JG Moreau, F Rakotozafy, J Tichet, S Vol. We thank M. Deweider and F. Allegaert for the DNA bank management. We are sincerely indebted to all who participated in this study.

EGCUT. EGCUT received support from EU FP7 grants (201413 ENGAGE, 205419 ECOGENE, 245536 OPENGENE). EGCUT also received targeted financing from Estonian Government SF0180142s08 and from the EU through the ERD Fund for CoExcel. in Genomics. EGCUT authors want to acknowledge M. Hass and V. Soo for the genotyping and technocal help. EGCUT data analyzes were carried out in part in the High Performance Computing Center of University of Tartu.

EPIC - Norfolk. The EPIC Norfolk Study is funded by program grants from the Medical Research Council UK and Cancer Research UK. We thank staff from the MRC Epidemiology Technical Team for carrying out the sample preparation, genotyping and associated quality control work.

GHRAS and THISEAS. We would like to thank all the field investigators for samples and data collection and all volunteers for participation in the studies.

HVH. NHLBI grants R01 HL085251, R01 HL073410, and R01 HL068986.

INGI Friuli Venezia Giulia and INGI Carlantino. This work has been funded by FVG Regional Government L.26-2008 and Fondo Trieste 2008, and Ministry of Health RC16/06.

INGI Cilento. We would like to address a special thank to the Cilento populations. This work was supported by grants from the Italian Ministry of Universities (FIRB - RBIN064YAT, RBNE08NKH7), the Assessorato Ricerca Regione Campania, the Ente Parco Nazionale del Cilento e Vallo di Diano to MC.

INGI Val Borbera. Compagnia di San Paolo, Torino, Italy to DT; Fondazione Cariplo,

Italy to DT; Ministry of Health, Ricerca Finalizzata 2008 to DT

BioBank Japan. We would like to thank all staff of the Laboratory for Statistical Analysis, Medical Informatics, and Genotyping Development at Center for Genomic Medicine, RIKEN and the BioBank Japan Projects for the supports for the study. BioBank Japan Project was supported by Ministry of Education, Culture, Sports, Science and Technology, Japan.

KORA F3 and KORA F4. The KORA research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823). KORA research was also supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. This study was supported through funds from The European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007- 201413.

LifeLines Cohort Study. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists. LifeLines Scientific Protocol Preparation: Rudolf de Boer, Hans Hillege, Melanie van der Klauw, Gerjan Navis, Hans Ormel, Dirkje Postma, Judith Rosmalen, Joris Slaets, Ronald Stolk, Bruce Wolffenbuttel; LifeLines GWAS Working Group: Behrooz Alizadeh, Marike Boezen, Marcel Bruinenberg, Noortje Festen, Lude Franke, Pim van der Harst, Gerjan Navis, Dirkje Postma, Harold Snieder, Cisca Wijmenga, Bruce Wolffenbuttel.

Lolipop. The LOLIPOP Study was supported by British Heart Foundation grant SP/04/002 and the Wellcome Trust. We thank the participants and research teams involved in LOLIPOP. We thank GSK for partly supporting genotyping of the Affymetrix data. We acknowledge on-going support from the Medical Research Council, the National Institute of Health Research, and the Imperial College Healthcare NHS Trust Comprehensive Biomedical Research Centre.

Lothian Birth Cohort 1921 and 1936. We thank the LBC1936 and LBC1921 participants. We thank Sarah Harris, Michelle Luciano, Dave Liewald, Alan Gow, Janie Corley, Caroline Brett, Caroline Cameron, Michelle Taylor and Alison Pattie for data collection, entry and preparation. We thank the study secretary Paula Davies. We thank the nurses and staff at the Wellcome Trust Clinical Research Facility, where subjects were tested and the genotyping was performed. We also thank the staff at the Department of Haematology, Western General Hospital for the hematology measurements. We thank the staff at the Lothian Health Board, and the staff at the SCRE Centre, University of Glasgow. The whole genome association study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC). The LBC1936 research was supported by a programme grant from Research Into Ageing and continues with programme grants from Help the Aged/Research Into Ageing (Disconnected Mind). The LBC1921 data collection was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC) and Medical Research Council (MRC) is gratefully acknowledged.

MICROS / South Tyrol. For the MICROS study, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano, and the South Tyrolean Sparkasse Foundation.

NFBC1966. We acknowledge the support of US National Heart, Lung, and Blood Institute grant HL087679 through the STAMPEED program, grants MH083268, GM053275-14 and U54 RR020278 from the US National Institutes of Health, grant DMS-0239427 from the National Science Foundation, the Medical Research Council of the UK, EURO-BLCS, QLG1-CT-2000-01643 and the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project and grant agreement HEALTH-F4-2007-201413. The authors would like to thank the Center of Excellence in Common Disease Genetics of the Academy of Finland and Nordic Center of Excellence in Disease Genetics, the Sydantautisaatio (Finnish Foundation of Heart Diseases), the Broad Genotyping Center, D. Mirel, H. Hobbs, J. DeYoung, P. Rantakallio, M. Koiranen and M. Isohanni for advice and assistance.

NTR and NTR2. Funding was obtained from the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW): Twin family database for behavior genomics studies (480-04-004); Twin research focusing on behavior (400-05-717); Genotype/phenotype database for behavior genetic and genetic epidemiological studies (911-09-032); Spinozapremie (SPI 56-464-14192); CMSB: Center for Medical Systems Biology (NWO Genomics); NBIC/BioAssist/RK/2008.024); BBMRI –NL: Biobanking and Biomolecular Resources Research Infrastructure; the VU University:

Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA); Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); European Community's Seventh Framework Program (FP7/2007-2013): ENGAGE (HEALTH-F4-2007-201413); the European Science Council (ERC) Genetics of Mental Illness (230374); Collaborative study of the genetics of DZ twinning (NIH R01D0042157-01A); the Genetic Association Information Network, a public–private partnership between the NIH and Pfizer Inc., Affymetrix Inc. and Abbott Laboratories.

OGP –Talana. We thank the Talana population and all the individuals who participated in this study. We are very grateful to the municipal administrations for their collaboration to the project and for economic and logistic support. This work was supported by grants from the Italian Ministry of Education, University and Research (MIUR) no.5571/DSPAR/2002 and (FIRB) D. M. no. 718/Ric/2005.

POPGEN. The German Ministry of Education and Research through the National Genome Research Network supported this study. Infrastructure support was received through the DFG cluster of excellence “Inflammation at Interfaces” and the popgen biobank.

QIMR. We thank the Brisbane twins and their families for their participation; Dixie Statham, Ann Eldridge, Marlene Grace (sample collection); Anjali Henders, Lisa Bowdler, Steven Crooks (sample processing) and David Smyth, Harry Beeby (IT support). Funding: Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496739, 552485, 613627, 552498); Australian Research Council (A7960034, A79906588, A79801419, DP0212016, DP0343921).

Sanquin Common Controls. The Sanquin Common Control collection was established by as part of the European Union FP6-funded Bloodomics project (LSHM-CT-2004-503485) with support of the Sanquin Blood Supply Foundation of the Netherlands.

SHIP. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the ‘Center of Knowledge Interchange’ program of the Siemens AG. This work is also part of the research project Greifswald Approach to Individualized Medicine (GANI_MED). The GANI_MED consortium is funded by the Federal Ministry of Education and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg – West Pomerania (03IS2061A).

SardiNIA. The SardiNIA (ProgeNIA) team was supported by Contract NO1-AG-1-2109 from the NIA, and in part by the Intramural Research Program of the National Institute on Aging (NIA), National Institutes of Health (NIH). The SardiNIA authors are grateful to the many volunteers who generously participate in the study, to the mayors and administrations of the four towns involved and the head of Public Health Unit ASL4 in Ogliastra.

SORBS. Financial support was received from the German Research Council (KFO-152), IZKF (B27) and the German Diabetes Association. We would like to thank Knut Krohn (Microarray Core Facility of the Interdisciplinary Centre for Clinical Research, University of Leipzig) for the genotyping/analytical support and Joachim Thiery (Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig) for clinical chemistry services. We thank Nigel W. Rayner (WTCHG, University of Oxford, UK) for the excellent bioinformatics support. This research of Dr. Inga Prokopenko and Dr. Vasiliki Lagou was part funded through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007-201413.

TwinsUK. The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and (FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413 and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the Dept of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator. The project also received support from a Biotechnology and Biological Sciences Research Council (BBSRC) project grant. (G20234) .The authors acknowledge the funding and support of the National Eye Institute via an NIH/CIDR genotyping project (PI: Terri Young). Genotyping of TwinsUK samples: We thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, Quality Control and Genotyping; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie. Genotyping was also performed by CIDR as part of an NEI/NIH project grant.

UK Blood Services (UKBS-CC1 and UKBS-CC2). The UK Blood Services collection of Common Controls (UKBS-CC collection), funded by the Wellcome Trust grants 076113/C/04/Z and 084183/Z/07/Z and by the Juvenile Diabetes Research Foundation grant WT061858, was established as part of the Wellcome Trust Case-Control Consortium, with the support of the blood services of England, Scotland and Wales. We gratefully acknowledge the generosity of the blood donors in England, Scotland and Wales who have made their DNA samples available to this project as a “free gift”.

Author contributions

Study design group
Christian Gieger, Serena Sanna, Andrew A Hicks, Augusto Rendon, Manuel A Ferreira, Willem H Ouwehand, Nicole Soranzo

Manuscript writing group
Christian Gieger, Aparna Radhakrishnan, Serena Sanna, Andrew A Hicks, Augusto Rendon, Manuel A Ferreira, Willem H Ouwehand, Nicole Soranzo

Data preparation, meta-analysis and secondary analysis group
Aparna Radhakrishnan, Brigitte Kühnel, Weihong Tang, Eleonora Porcu, Giorgio Pistis, Reedik Magi, Manuel A Ferreira, Christian Gieger, Nicole Soranzo

Bioinformatics analyses, Pathway analyses and Protein-protein interaction network group
Stuart Meacham, Jan-Willem N Akkerman, Steve Jupe, Jyoti Khadake, Yasin Memari, Lorenzo Bomba, Augusto Rendon, Willem H Ouwehand

Transcript profiling methods and data group
Katrin Voss, Augusto Rendon, Lorenz Wernisch, Alison H Goodall, Tsun-Po Yang, François Cambien, Jeanette Erdmann, Christian Hengstenberg, Nilesh J Samani, Heribert Schunkert, Panos Deloukas, Willem H Ouwehand.

***M. musculus* models**
Ramiro Ramirez-Solis

***D. rerio* knockdown models**
Ana Cvejic, Jovana Serbanovic-Canic, Derek Stemple, Willem H Ouwehand

***D. melanogaster* knockdown models**
Ulrich Elling, Josef Penninger, Andrew A Hicks

Cohort-specific contributions

Cat.	Cohort	Name	Geno-typing	Pheno-typing	Data Analysis	Cohort PI/Study Design
RBC	ALSPAC	David M Evans			x	
RBC	ALSPAC	Susan M Ring				x
R	Amish	Afshin Parsa		x	x	x
R	Amish	Quince D. Gibson			x	
R	Amish	Braxton Mitchell	x		x	x
R	Amish	Alan R. Shuldiner	x	x		x
D	ARIC	Weihong Tang			x	x
D	ARIC	Aaron R Folsom		x		x
D	ARIC	Saonli Basu			x	
D	ARIC	Wei Feng			x	
D	ARIC	Santhi K Ganesh				x
R	Cleveland Clinic GeneBank	Jaana Hartiala			x	
R	Cleveland Clinic GeneBank	Wilson W.H. Tang	x	x		
R	Cleveland Clinic GeneBank	Hooman Allayee			x	
R	Cleveland Clinic GeneBank	Stanley L. Hazen	x	x		
RBC	CoLaus	François Bastardot			x	
RBC	CoLaus	Micha Hersch				x
D	DESIR	Yann Labrune			x	
D	DESIR	Amélie Bonnefond			x	
D	DESIR	Mario Falchi			x	
D	DESIR	Beverley Balkau		x		
D	DESIR	Philippe Froguel	x	x		
D	EPIC	Claudia Langenberg			x	
D	EPIC	Jing Hua Zhao			x	
D	EPIC	Ruth J. F. Loos				x
D	EPIC	Kay Tee Khaw				x
D	EPIC	Nicholas J. Wareham				x
R	EGCUT	Tõnu Esko			x	
R	EGCUT	Janne Pullat	x	x		x
R	EGCUT	Andres Metspalu		x		x
R	INGI CAR FVG	Nicola Pirastu			x	
R	INGI CARL FVG	Pio D'Adamo			x	
R	INGI CARL FVG	Sheila Ulivi			x	
R	INGI CARL FVG	Paolo Gasparini	x	x		x
D	INGI Cilento	Rossella Sorice	x		x	
D	INGI Cilento	Marina Ciullo				x
D	INGI Cilento	Daniela Ruggiero	x		x	
D	INGI Val Borbera	Giorgio Pistis			x	
D	INGI Val Borbera	Cinzia Sala	x			
D	INGI Val Borbera	Daniela Toniolo	x	x	x	
I	BioBank Japan	Yukinori Okada		x	x	x
I	BioBank Japan	Yusuke Nakamura	x	x		x
I	BioBank Japan	Atsushi Takahashi			x	
I	BioBank Japan	Naoyuki Kamatani	x	x		x
R	LifeLines	Pim van der Harst			x	x
R	LifeLines	Rudolf A de Boer	x	x		
R	LifeLines	Irene Mateo Leach	x	x		x
R	LifeLines	L. Joost van Pelt		x	x	
D+I	LOLIPOP	Weihua Zhang			x	
D+I	LOLIPOP	Paul Elliott	x			x

Cat.	Cohort	Name	Geno-typing	Pheno-typing	Data Analysis	Cohort PI/Study Design
D+I	LOLIPOP	James Scott	x			x
D+I	LOLIPOP	John C Chambers	x	x	x	x
D+I	LOLIPOP	Jaspal S Kooner	x	x	x	x
D	MICROS / South Tyrol	Martin Gögele	x		x	
D	MICROS / South Tyrol	Peter P Pramstaller		x		x
D	MICROS / South Tyrol	Andrew A Hicks				x
RBC	NESDA	Harold Snieder				x
RBC	NESDA	Ilja M Nolte			x	
D	NFBC 1966	Paul F O'Reilly			x	
D	NFBC 1966	Aimo Ruokonen		x		
D	NFBC 1966	Anneli Pouta		x		
D	NFBC 1966	Anna-Liisa Hartikainen	x	x		
D	NFBC 1966	Paavo Zitting	x	x		
D	NFBC 1966	Marjo-Riitta Jarvelin	x	x		x
D+R	NTR+NTR2	Jouke-Jan Hottenga			x	
D+R	NTR+NTR2	Eco JC de Geus		x		
D+R	NTR+NTR2	Gonneke Willemsen		x		
D+R	NTR+NTR2	Dorret I Boomsma				x
R	OGP-Talana	Federico Murgia			x	
R	OGP-Talana	Ginevra Biino				x
R	OGP-Talana	Mario Pirastu				x
D	POPGEN	David Ellinghaus			x	
D	POPGEN	Ute Nöthlings		x		x
D	POPGEN	Stefan Schreiber		x		x
D	POPGEN	Andre Franke			x	
RBC	PREVENT	Gerjan Navis			x	
RBC	PREVENT	Dirk J van Veldhuisen				x
D	QIMR	Manuel A Ferreira			x	x
D	QIMR	Grant W Montgomery	x			
D	QIMR	Ian H Frazer		x		
D	QIMR	Nicholas G Martin	x	x		
R	SANQUIN-CC	Ellen van der Schoot		x		x
R	SANQUIN-CC	Jaap-Jan Zwaginga		x		x
D	SardiNIA	Eleonora Porcu			x	
D	SardiNIA	Andrea Maschio	x			
D	SardiNIA	Francesco Cucca		x		x
D	SardiNIA	David Schlessinger				x
D	SardiNIA	Serena Sanna			x	
D	SardiNIA	Manuela Uda	x	x		x
D	SHIP	Alexander Teumer	x		x	
D	SHIP	Sebastian E Baumeister				x
D	SHIP	Matthias Nauck	x	x		
D	SHIP	Uwe Völker	x			
D	SHIP	Andreas Greinacher		x		x
D	SORBS	Peter Kovacs				x
D	SORBS	Michael Stumvoll				x
D	SORBS	Anke Tönjes	x	x	x	
D	SORBS	Vasiliki Lagou	x		x	
D	SORBS	Inga Prokopenko			x	x
D	UKBS-CC	Aparna Radhakrishnan			x	
D	UKBS-CC	Jennifer Sambrook		x		

Cat.	Cohort	Name	Geno-typing	Pheno-typing	Data Analysis	Cohort PI/Study Design
D	UKBS-CC	Antony Attwood	x	x		x
D	UKBS-CC	Willem H Ouwehand		x		x
R	UKBS-CC	Jonathan Stephens	x			
R	CBR	Jennifer Jolley	x			
R	CBR	Heather Lloyd-Jones	x			
R	CBR	John R Bradley				x
D	CHS-HVH	Joshua C Bis			x	x
D	CHS-HVH	Nicole L Glazer			x	x
D	CHS-HVH	Kerri L Wiggins			x	
D	CHS-HVH	Thomas Lumley			x	x
D	CHS-HVH	Nicholas L Smith		x	x	x
D	CHS-HVH	Kent Taylor	x			
D	CHS-HVH	Susan H Heckbert		x	x	x
D	CHS-HVH	Barbara McKnight			x	x
D	CHS-HVH	Jerome I Rotter	x			
D	CHS-HVH	Bruce M Psaty		x	x	x
R	GHRAS and THISEAS	Stavroula Kanoni			x	
R	GHRAS and THISEAS	Marie-Christine Kyrtsonis		x		
R	GHRAS and THISEAS	Panos Deloukas	x			x
R	GHRAS and THISEAS	George V Dedoussis				x
D	KORA F3 and F4	Christian Gieger			x	x
D	KORA F3 and F4	H-Erich Wichmann		x		x
D	KORA F3 and F4	Angela Döring		x		
D	KORA F3 and F4	Thomas Illig	x			
D	KORA F3 and F4	Brigitte Kühnel			x	
D	KORA F3 and F4	Christa Meisinger		x		x
D	LBC1921 & LBC1936	Lorna M Lopez			x	
D	LBC1921 & LBC1936	Gail Davies			x	
D	LBC1921 & LBC1936	David Porteous		x		x
D	LBC1921 & LBC1936	Albert Tenesa				x
D	LBC1921 & LBC1936	John M Starr				x
D	LBC1921 & LBC1936	Peter M Visscher				x
D	LBC1921 & LBC1936	Ian J Deary				x
D+R	TwinsUK 317k and 610k	So-Youn Shin	x		x	
D+R	TwinsUK 317k and 610k	Massimo Mangino	x			
D+R	TwinsUK 317k and 610k	Chris Hammond	x			x
D+R	TwinsUK 317k and 610k	Swee Lay Thein	x	x		X
D+R	TwinsUK 317k and 610k	Timothy D Spector	x	x		X
D+R	TwinsUK 317k and 610k	Nicole Soranzo	x		x	X

D = Discovery cohort
R = Replication cohort
I = Interethnic replication cohort
RBC = Erythrocyte replication cohort