The Netherlands Twin Register Biobank: A Resource for Genetic Epidemiological Studies

Gonneke Willemsen,[1] Eco J. C. de Geus,[1] Meike Bartels,[1] C. E. M. Toos van Beijsterveldt,[1] Andy I. Brooks[2], G. Frederique Estourgie-van Burk,[1] Douglas A. Fugman,[2] Chantal Hoekstra,[1] Jouke-Jan Hottenga,[1] Kees Kluft,[3,4] Piet Meijer,[3,4] Grant W. Montgomery,[5] Patrizia Rizzu,[6] David Sondervan,[6] August B. Smit,[7] Sabine Spijker,[7] H. Eka D. Suchiman,[8] Jay A. Tischfield,[2] Thomas Lehner,[9] P. Eline Slagboom[8] and Dorret I. Boomsma[1]

[1] *Department of Biological Psychology, VU University Amsterdam, Amsterdam, Netherlands*
[2] *Rutgers University Cell and DNA Repository, Rutgers University, Department of Genetics, Piscataway, New Jersey, United States of America*
[3] *Gaubius Laboratory, TNO-Quality of Life, Biomedical Research, Leiden, Netherlands*
[4] *Good Biomarker Sciences, Leiden, Netherlands*
[5] *Queensland Institute of Medical Research, Herston, Australia*
[6] *Department of Clinical Genetics, VU Medical Center, Amsterdam, Netherlands*
[7] *Department of Molecular and Cellular Neurobiology, VU University Amsterdam, Amsterdam, Netherlands*
[8] *Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, Netherlands*
[9] *Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, United States of America*

In 2004 the Netherlands Twin Register (NTR) started a large scale biological sample collection in twin families to create a resource for genetic studies on health, lifestyle and personality. Between January 2004 and July 2008, adult participants from NTR research projects were invited into the study. During a home visit between 7:00 and 10:00 am, fasting blood and morning urine samples were collected. Fertile women were bled on day 2–4 of the menstrual cycle, or in their pill-free week. Biological samples were collected for DNA isolation, gene expression studies, creation of cell lines and for biomarker assessment. At the time of blood sampling, additional phenotypic information concerning health, medication use, body composition and smoking was collected. Of the participants contacted, 69% participated. Blood and urine samples were collected in 9,530 participants (63% female, average age 44.4 (SD 15.5) years) from 3,477 families. Lipid profile, glucose, insulin, HbA1c, haematology, CRP, fibrinogen, liver enzymes and creatinine have been assessed. Longitudinal survey data on health, personality and lifestyle are currently available for 90% of all participants. Genome-wide SNP data are available for 3,524 participants, with additional genotyping ongoing. The NTR biobank, combined with the extensive phenotypic information available within the NTR, provides a valuable resource for the study of genetic determinants of individual differences in mental and physical health. It offers opportunities for DNA-based and gene expression studies as well as for future metabolomic and proteomic projects.

Keywords: DNA, RNA, GWAS, mental health, physical health

Since the unravelling of the human genome, systematic sampling of genetically informative data has become part of many epidemiological studies. Biobanks provide the necessary infrastructure for gene discovery studies for mental and physical health. The most valuable biobanks provide high quality genotypic information coupled to deep phenotyping. This is a tall order taking into account the costs and personnel involved in obtaining both kinds of information. That this effort is worthwhile is demonstrated by the results of large scale GWA studies on e.g. lipids, BMI and glucose (e.g. Aulchenko et al., 2009; Diabetes Genetics Initiative, 2007; Dupuis et al., 2010; Kettunen et al., 2009; Prokopenko et al., 2009; Scuteri et al., 2007). As we aim to understand more about the steps that are involved between DNA sequence variation and the ultimate phenotypic expression, there is an increasing need to study intermediate steps, such as gene expression, epigenetics and proteomics.

In January 2004 the Netherlands Twin Register (NTR) undertook a pilot study to examine the feasibility of constructing a biobank, which would include a DNA collection, but also biological material suitable for gene expression studies, the creation of immortalized cell lines and future metabolomic and proteomics studies (Hoekstra et al., 2004; Spijker et al., 2004). As the pilot study was successful, a 4-year project was undertaken to visit twins and their family members who participate in NTR research projects. These NTR

Received 23 March, 2010; accepted 12 April, 2010.

Address for correspondence: Gonneke Willemsen, Department of Biological Psychology, VU University Amsterdam, Van der Boechorststraat 1, The Netherlands. E-mail: ahm.willemsen@psy.vu.nl

projects include longitudinal survey studies (Boomsma et al., 2002, 2006), cardiovascular studies (Kupper et al., 2005, 2006; de Geus, 2007), studies on the etiology of DZ twinning (Hoekstra et al., 2008), studies on cognition and brain function (Smit et al., 2010; van 't Ent et al., 2009) and on psychiatric phenotypes (Distel et al., 2009; van Grootheest et al., 2008; Middeldorp et al., 2007). The NTR Biobank thus aims to combine phenotypic information, often obtained on multiple occasions, with extensive genotypic information. One important additional feature of the NTR Biobank is the fact that information is obtained in family members who are biologically related (e.g. parents and children, siblings, twins) or who share a household but are biologically unrelated (e.g. spouses). Also, a large number of monozygotic (MZ) twins are included, who may be particularly valuable for epigenetics projects and RNA expression studies. After fours years, the collection of biological samples finished in 2008. This paper describes the NTR Biobank and the currently available genotypic and phenotypic information. In addition, we report on the distribution of subjective health reports, medication use and smoking for men and women in two age groups (younger than 50 years and 50 years and over). For body composition, lipid profile, glucose metabolism, haematology, CRP, liver enzymes, creatinine and fibrinogen we present descriptive statistics on younger and older unrelated men and women.

Methods

Participants

Individuals of 18 years and over who participated in NTR research and their family members who were registered as active participants with the NTR were selected for the study. Previous participation in NTR research could consist of completion of one or more surveys as part of a longitudinal study on health, lifestyle and personality, which started in 1991. Mailed questionnaires were sent out to twins and their family members in 1991, 1993, 1995, 1997, 2000, 2002/3, and 2004/5. More than 22,000 participants from 5,500 families (some families are linked and come from larger pedigrees) have participated at least once in these surveys. In addition, individuals could have participated in smaller scale studies that focused on ambulatory cardiovascular responses, laboratory stress reactivity, cognitive function or brain function. The majority of families registered with the NTR were recruited when the twins were adolescents or young adults through City Council registration systems in 1990–91 and in 1992–93. Since 1993, adult twins are recruited through a variety of other approaches, such as advertisements. Detailed information on the longitudinal survey study and additional projects has been provided previously (see Boomsma et al., 2002, 2006).

In addition to the participants already registered with the NTR, an active recruitment was initiated for families in which two sisters both were mothers of twins for a study on the genetics of dizygotic twinning (Hoekstra et al., 2008). Women who were recruited into this study received an interview after which they and their family members (that is their parents or, in case no parents were present, their siblings) were also included in the biobank selection, following the same procedure as for other NTR participants.

The biobank further includes data from young twins and siblings who were registered as newborns by their parents and have been reported on by their parents from birth onwards until they became old enough to complete the reports themselves (Bartels et al., 2007). These twins and siblings participated in a longitudinal study on cognition and health (Estourgie-van Burk et al., in press) and as part of this study they underwent the same blood sampling protocol during a visit to the VU University as was used in the large-scale biobank project.

The study protocol was approved by Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NTR 03-180) and informed consent was obtained from all participants. The study in young twins was in addition approved by the Central Committee on Research Involving Human Subjects.

Procedure

As NTR biobank participants were visited at home, the recruitment of NTR participants in the biobank took place in phases, based on residential address. A particular region of the country was targeted for a period of 4 to 6 weeks and all NTR participants within that region were invited. An invitation letter with an information leaflet was followed by a phone call to answer any questions and when the individual was willing to participate, to make an appointment for the blood sampling. Blood sampling and urine collection was obtained during a home visit, or when preferred, at another address or at work (<1%). Participants were visited by two research assistants, one of which was a trained phlebotomist, who took the blood samples. The other assistant obtained interview data and body composition measures. Visits took place between 7 a.m. and 10 a.m. and lasted 10 to 15 minutes.

Participants were instructed to fast as of 10 p.m. the previous night and to refrain as much as possible from medication use and physical exertion on the day of the home visit and, in the case of smokers, to refrain from smoking one hour prior to the home visit. Fertile women not using hormonal birth control were visited on the 2nd to 4th day of the menstrual cycle, whenever possible. They were asked to phone for an appointment in their region at the start of their menstrual period. Women taking hormonal birth control and who had one week every four weeks when they did not take hormones were visited in that week. Participants were asked to collect their morning urine.

Twin individuals were asked whether they would be willing to provide a swab for buccal DNA (Meulenbelt et al., 1995) to provide additional DNA for zygosity determination and for comparisons with DNA obtained from blood, because twins may show chimearism (van Dijk et al., 1996). Materials for collection of buccal DNA were mailed in batches to participants within a few weeks after the home visits.

Blood Sampling, Processing and Storage

Using a safety lock butterfly needle, eight blood tubes were collected in the following order; 2 × 9 ml EDTA, 2 × 9 ml lithium heparin, 1 × 4.5 ml CTAD, 1 × 2 ml EDTA, 1 × 4.5 ml serum.

To prevent clotting, all tubes were inverted gently 8–10 times immediately after collection. When immediate processing of the sample was required (as is described further below for each of the samples), this took part in a van, which had a small laboratory set up in the back. All samples were subsequently transported to the central laboratory in Leiden. Average transport time was 196 minutes (*SD* = 76) and sample processing started as soon as samples arrived at the lab (full sample processing was completed on average 422 minutes (*SD* = 31) after sample collection). The degree of haemolysis was determined by observing the colour of the plasma after centrifugation. Based on observed haemolysis, there was no association between transport time and haemolysis.

The 2 × 9 ml EDTA anticoagulant tubes were collected for DNA isolation and EDTA plasma. During transport these tubes were stored in melting ice. Upon arrival at the laboratory, the EDTA tubes were centrifuged for 20 minutes at 2000x *g* at 4° C. EDTA plasma was harvested and aliquoted (0.5 ml), snap-frozen in dry ice, and stored at –30° C. The EDTA tubes with buffy coat and red cells samples were stored at –20° C.

For RNA isolation, one of the two 9 ml lithium heparin coagulant tubes was used. Prior to transport 2.5 ml blood was transferred to a Pax-tube, which was then gently inverted a number of times. During transport this Pax-tube, meant for basal RNA determination, was stored at room temperature and upon arrival in the laboratory snap-frozen in dry ice and stored at –30° C. Prior to transport, we added 65 µL Lipopolysaccharide (LPS Sigma L-3012 lot Nr: 36H4130 from Escherichia Coli Serotype 0111:B4) to the remainder of blood (~6.5 ml) in the heparin tube, resulting in a 1 µg/mL concentration. After inverting the tube gently a number of times, it was stored for exactly 6 hours after adding LPS in an insulated box with a constant temperature of 37° C. After 6 hours 2.5 ml of blood was taken from the heparin tube and added to a PAX-tube, inverting the tube gently afterwards. After two hours at room temperature the tube was snap-frozen and stored at –30° C.

The second heparin tube was stored in melting ice during transport and processed at the laboratory within 6 hours of transport. After centrifugation of the tube for 15 minutes at 1000x *g* at 4° C, heparin plasma was obtained and divided into 8 subsamples of 0.5 ml, snap-frozen and stored at –30° C. The processing of this second heparin tube took place in a sterile flow cabinet. For subsequent lymphocyte isolation, 14 ml phosphate-buffered saline (PBS) was added to the remaining blood cells and buffycoat. This diluted blood was then added to 10 ml Ficoll and centrifuged for 25 minutes at 800x *g* at room temperature. The middle layer containing the leucocytes was transferred into a new tube and 45 ml PBS was added. The solution was centrifuged 10 minutes at 250x *g* at room temperature. This washing step was repeated after discarding the supernatant. The lymphocyte pellet was then taken up into 1 ml RPMI: 20% Fetal Calf Serum (FCS), 10% and dimethyl sulfoxide (DMSO). Finally, pellets were frozen and stored overnight at –80° C and then transferred to liquid nitrogen for long-term storage.

The 4.5 ml CTAD tube was collected for future coagulation studies. It was stored during transport in melting ice and upon arrival at the laboratory, centrifuged for 20 minutes at 2000x *g* at 4° C, after which citrated plasma was harvested from the buffy coat and red blood cells, aliquoted (0.5 ml), snap-frozen in dry ice, and stored at –30° C.

The 2 ml EDTA tube was stored at room temperature during transport and used to obtain Hba1c and haematological parameters.

The 4.5 ml serum tube was stored during transport in melting ice. As before with the CTAD and 9 ml EDTA tubes, upon arrival at the laboratory, the serum tube was centrifuged for 20 minutes at 2000x *g* at 4° C, after which serum plasma was harvested from the buffy coat and red blood cells, aliquoted (0.5 ml), snap-frozen in dry ice, and stored at –30° C.

An overview of blood collection, transport and blood handling is listed in Table 1.

At sample collection the tubes were double coded; one code represents the individual according to the twin register ID (family ID and individual ID); the other code is sequential and is used to identify the storage position of the sample. The different types of samples are further identified by the tube size, colour of the cap or storage box. The RNA materials were labeled as A (unchallenged) and B (challenged).

Urine Processing and Storage

The urine sample collected by the participant was transferred to two 10 ml vacutainers and stored in melting ice during transport. In the laboratory the urine tubes were centrifuged for 20 min at 2000x *g* at 4° C. The urine without the debris pellet was then decanted in 10 ml plastic tubes, snap-frozen and stored at –30° C.

Buccal DNA Collection

Twins and triplets who agreed to do a mouth swab were mailed a sample collection kit containing a tube with cotton buds, tubes with collection buffer, a sam-

Table 1

Overview of Blood and Urine Collection, Processing and Storage

At home visit	After home visit, before transport	Transport	At laboratory	Storage	Purpose
2 × 9 ml EDTA blood		In melting ice	Within 6 hours after sampling, plasma was collected	14 subsamples of ~500 µl, snap-frozen, stored at –30° C	Biomarkers
			Red cells, buffy coat and rest plasma were stored	Stored at –20° C	DNA isolation
1 × 9 ml heparin blood		In melting ice	Within 6 hours after sampling, leukocyte cells were collected nitrogen	1 sample stored at –80° C overnight, additionally in liquid	Immortalized cell lines
			Plasma was collected	6 subsamples of ~500 µl, snap-frozen, stored at –30° C	Biomarkers
1 × 9 ml heparin blood	Transfer of 2.5 ml to PAX tube	Room temperature	Snap-frozen on arrival	PAX-tube stored at –30° C, later stored at VU University at –20° C	RNA expression
	LPS added to remaining 6.5 ml	37° C	6 hours after LPS, 2.5 ml blood added to PAX tube; after 2 hours room temperature snap-frozen	PAX-tube stored at –30° C, later stored at VU University at –20° C	RNA expression
1 × 4.5 ml CTAD blood		In melting ice	Plasma collection	4 subsamples of ~500 µl, snap-frozen, stored at –30° C	Biomarkers
1 × 2.5 ml EDTA blood		Room temperature	Within 6–8 hrs, measurement of HbA1c, haematology	No further storage	Hba1c, Haematology
1 × 4.5 ml serum blood		In melting ice	Plasma collection	4 subsamples of ~500 µl, snap-frozen, stored at –30° C	Biomarkers
2 × 9 ml urine		In melting ice	Within 6 hrs, centrifuging to decant urine without debris	2 x 9 ml tubes stored at –20° C	Metabolomics

pling protocol, informed consent forms and a prepaid envelope. They were asked to refrain from eating or drinking one hour prior to collection. Participants took mouth swabs themselves following the sampling protocol which instructed on how to gently rub the cotton swab tips along the inside of the mouth. After rubbing, the mouth swabs were placed in a Falcon tube, containing 0.5 ml of STE buffer (100 mM NaCl, 10 mM Tris and 10 mM EDTA) with proteinase K (0.1 mg/ml) and sodium dodecyl sulfate (0.5%) per mouth swab. In most participants, 16 mouth swabs were collected by conducting 4 separate swabs twice per day during two days. The sample tubes were returned by mail. Upon arrival, swabs were stored at room temperature in the laboratory until DNA extraction.

Biomarkers in Blood

Lipid profile. Total cholesterol, HDL-cholesterol and triglycerides were determined in plasma from the heparin tubes using Vitros 250 total cholesterol, Vitros 250 direct HDL cholesterol and Vitros 250 Triglycerides assays (Johnson & Johnson, Rochester, USA) respectively. LDL-cholesterol was calculated using the Friedewald Equation: LDL = Total cholesterol — HDL — (Triglycerides/5); (Friedewald et al., 1972).

Glucose metabolism. Insulin was measured using the Immulite 1000 Insulin Method (Diagnostic Product Corporation, Los Angeles, USA) and glucose was measured using the Vitros 250 Glucose assay (Johnson & Johnson, Rochester, USA). Hba1c was determined in EDTA whole blood directly upon arrival in the laboratory using the Nyocard HbA1c assay (Axis-Shield, Oslo, Norway).

Haematology. Using the Coulter system (Coulter Corporation, Miami, USA) a haematological profile was obtained from EDTA blood directly after the samples arrived in the laboratory. Parameters provided were white blood cell count (WBC), percentages and numbers of neutrophils, lymphocytes, monocytes, eosinophils, basophils as well as red blood cell count, hemoglobulin, hematocrit, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW), platelets and mean platelet volume (MPV).

CRP. From heparin plasma, C-reactive protein (CRP) level was determined using the Immulite 1000 CRP assay (Diagnostic Product Corporation, USA).

Fibrinogen. Fibrinogen levels were determined in citrate plasma, on a STA Compact Analyzer (Diagnostica Stago, France), using STA Fibrinogen (Diagnostica Stago, France).

Liver enzymes. Gamma glutamyl transferase (GGT), alanine aminotransferase (ALT), aspartate aminotransferase (AST) were determined in heparin plasma using the appropriate Vitros assays (Johnson & Johnson, Rochester, USA).

Creatinine. Creatinine level was determined in heparin plasma, using Vitro assay (Johnson & Johnson, Rochester, USA).

DNA Isolation

Blood. From one of the two EDTA tubes with buffy coat and red cells that were stored for each participant, DNA was extracted using the QIAamp DNA Blood Maxi (QIAGEN, Düsseldorf, Germany). DNA concentrations were determined using Nanodrop (NanoDrop Technologies Wilmington, Delaware, USA). All procedures were performed according to the manufacturer's protocols.

Buccal. High molecular genomic DNA from mouth swab samples was extracted using a high salt (KAc) precipitation procedure followed by a standard DNA extraction method as previously described (Beekman et al., 2001; Meulenbelt et al., 1995; Min et al. 2006).

RNA Isolation and cDNA Amplification

After processing and initial storage at -20° C, the PAXgene tubes containing whole blood for total RNA and miRNA extraction were shipped frozen to the Rutgers University Cell and DNA Repository (RUCDR), USA. Upon registration of samples, extractions were performed in a two stage process. In the first stage, for isolation of total RNA, samples were processed 96 at a time using solid phase extraction chemistry on a Qiagen Universal liquid handling system (PAXgene extraction kits as per the manufacturers protocol). Briefly, samples are lysed, neutralized and RNA bound to a plate based solid phase matrix. The bound material is washed and then eluted into a final volume of 120ul which contains purified, nuclease free total RNA from each sample. The range of RNA yield per sample is from 2.5 ug to 5 ug per 2.5 ml of whole blood in each PAXgene tube. During the total RNA extraction process the 'flow through' from the total RNA binding event is collected and bound to a different solid phase extraction matrix (RNAeasy) for collection and enrichment of small RNA species which predominantly includes the miRNA fraction from whole blood samples.

The total RNA and miRNA fractions are stored independently in Micronic 2D tube rack storage systems in multiple aliquots at -80C until accessed for analysis. Total RNA is measured by spectroscopy (NanoDrop) to determine purity and concentration (total yield is calculated from these values) while RNA fidelity is measured by the Agilent Bioanalyzer analysis. All QC analyses are stored in the RUCDR LIMS where each measurement and profile is linked to sample storage information.

Functional testing requires the conversion of total RNA to cDNA in order to make gene expression measurements. To ensure that each sample is a renewable resource, the RUCDR linearly amplifies cDNA from very small starting amounts of total RNA. The chemistry used to create linearly amplified cDNA is NuGEN WT-Pico Ovation chemistry. Briefly, 5 to 20 ng of total RNA from each sample is used to generate a high fidelity cDNA for array hybridization. Using Ribo-SPIA(tm) linear amplification "antisense" cDNA is generated by DNA polymerase replication of a double stranded DNA template prepared from mRNA. Single strand cDNA is prepared from total RNA using a unique DNA/RNA chimeric primer and reverse transcriptase. The primer has a DNA portion that hybridizes randomly to sites of the mRNA and an RNA portion that does not hybridize to the mRNA. The resulting cDNA/mRNA complex has a unique RNA target sequence at the 5′ end of the cDNA. After fragmentation of the mRNA in the cDNA/mRNA complex, DNA polymerase is used to synthesize a second strand including DNA complementary to the 5′ RNA unique sequence incorporated into the first strand. The resulting structure is double stranded DNA with a RNA/DNA heteroduplex of unique sequence at one end. SPIA(tm) is a DNA amplification method developed by NuGEN. SPIA(tm) uses a DNA/RNA chimeric primer, DNA polymerase and RNAse H in a homogeneous, isothermal reaction that provides highly efficient amplification of DNA sequences. In the SPIA(tm) process, RNase H is used to remove the unique RNA sequence in the double stranded cDNA. The sequence now exposed is available for binding a second DNA/RNA chimeric primer. DNA polymerase, together with modified nucleotides in solution, is used to synthesize DNA starting at the 3′ end of the primer and displacing the existing forward strand. RNA at the 5′ end of the newly synthesized strand is again removed with RNase H, exposing part of the priming site for initiation of the next round of DNA synthesis. The process of DNA/RNA primer binding, DNA synthesis and strand displacement and RNA cleavage of the primer is repeated with rapid accumulation of cDNA with sequence complementary (antisense) to the original mRNA. The typical yield from each reaction is approximately 10 ug of cDNA that can be used for any downstream gene expression analysis including and not limited to QPCR, Affymetrix GeneChip, Illumina BeadArray and Agilent Microarrays.

Cell Lines

The cryopreserved leukocytes (CPLs) were shipped in liquid nitrogen dry shippers to the RUCDR, USA for processing and conversion to transformed lymphocyte cell lines (LCLs). To initiate an Epstein-Bar transformation culture, the vials of cryopreserved lymphocytes were thawed in a 37° C water bath, and resuspended in a 25%FBS/RPMI-1640 medium. To remove dimethanolsulfoximide (DMSO), the cell suspension was centrifuged and washed once, after which the mixture was resuspended in 5 ml of transformation medium (25%FBS, RPMI-1640, 10% EBV supernatant, 0.1% phytohemaglutinin). This suspension was placed in a T25 flask on a layer of irradiated MRC-5 human lung fibroblasts (6,000 RADS in a Torrex x-ray cabinet, Faxitron Corp.) and placed in a 37° C humidified incubator with 5% CO2. On day

3–4 of incubation 2 to 3 ml of growth medium (25% FBS/RPMI1640) was added to the culture. Every 3 to 4 days the medium was supplemented by removing 2 to 3 ml of old medium from each flask and replaced with 2 to 3 ml of fresh medium. When the cell density of transformed lymphocytes increases, the cells detach from the adherent cell layer. The detached transformed lymphocyte cell suspension is transferred to a new T-25 flask (devoid of any MRC-5 cells). An additional 2 to 3 ml of fresh medium is added to the flask. This first split is maintained until the transferred lymphocytes are growing robustly, then the LCL culture is split a second time into new T-25 flasks. Once the robustly growing LCL culture has been passaged several times to remove dead cells, the cells are transferred into a T75 flask to provide adequate cell numbers for cryopreservation and DNA extraction.

Phenotype Data Collection During the Home Visit

During the home visit, participants received a brief interview. They were asked what time they got up that morning and whether this was their normal time of waking. Fasting status was checked and any deviation from overnight fasting was recorded. Participants were also asked whether they had a very large or fatty meal the evening before and whether they had physically exerted themselves just prior to the visit. If so, a description of the activity was noted. Participants were instructed in the confirmation letter to have at hand any medication they were using. They were asked about all current medication use and for all medicines the dose, brand and chemical names were recorded directly from the medication packaging. Fertile women not using oral contraceptives were asked to indicate the number of days since the first day of their last menstruation. Women using oral contraceptives reported whether they were in a pill-free week and, if so, the number of days since stopping pill use. In addition, they were asked the brand of the oral contraceptives they were taking. Use of other kinds of hormonal contraceptives was noted.

Participants were asked to rate their physical health as excellent, good, reasonable, moderate, or poor. They also indicated the last time they were ill (less than 1 week ago, less than 1 month ago, or more than 1 month ago) and what this illness was (having a cold (no fever), flu (with fever), infection or other). In the last case, a specification was asked. Participants also indicated whether they smoked or had ever smoked. If so, they were asked to report the number of cigarettes smoked and the number of years they had smoked. Past-smokers were also asked how many years ago they quit smoking. Height was obtained, weight was measured after blood sampling, after removal of shoes, and waist and hip circumference were determined.

Database Organization

Within the NTR data are stored in three different databases; the Administrative database, the Phenotype database and the Genotype database. Personal information such as name, address and relationships between participants are stored in the Administrative database on a separate network that is not physically connected to any other networks and login protected with different authorization levels (see Boomsma et al., 2008 for a more detailed description). The Phenotype database consists of a SPSS-based database directory structure, with separate SPSS files for different projects which may be linked through a unique individual ID. Genotypic information (e.g. microsatellite markers, SNPs) is stored in the Genotypic database, that is Progeny based, separate from the Administrative and Phenotype databases. A subset of the phenotype and genomic databases will also be mirrored in NIH based access controlled data repositories as a resource for the investigator community.

Results

We sent invitation letters out to 14,093 participants (64% of adult NTR participants). In response we received information that 96 individuals had died and 54 lived abroad. Another 2,190 individuals were not reached by phone. Of the 11,753 individuals who were successfully contacted by phone, 8,126 (69%) agreed to participate, 193 (2%) had problems deciding and 3,434 (29%) did not want to participate. For 453 individuals who agreed to participate an appointment was difficult to make, so the final number of participants recruited by this procedure was 7,673. For 30% of non-participants reason of refusal was not stated, 27% no longer wanted to take part in NTR research and 21% indicated they did not feel like taking part. In 8%, problems with blood sampling (e.g. afraid of needles, fainting) were stated as the reason for refusal. A comparison of non-participants with participants on characteristics known from previous questionnaires showed that non-participants were less often female (52% vs. 64%), were slightly younger (average birth year 1963 vs. 1960 in the men, 1964 vs. 1962 in the women) and had less often received higher education (27% vs. 39% in men, 22% vs. 33% in women) than participants.

The final number of participants in the NTR Biobank is 9,530. This number includes the participants who were recruited through the invitation procedure described above (N = 7,673), as well as additional twin mothers and family members who were recruited for the DZ twinning project (N = 1,059) and young adult twins, who were registered as children by their parents, and their siblings (N = 434). An additional 364 participants entered the study because their spouse or family member indicated their willingness to participate during the interview. If these individuals were already registered with the NTR and had also received an invitation for the study, they were scheduled for blood sampling but separate informed consent was still obtained. In case the individual was not yet registered with the NTR (which generally con-

cerned the spouses of participants), they received a registration package at home and information and informed consent for the biobank study.

For 63 of the participants no blood sample could be taken during the visit and a second visit was not possible, thus only urine samples were been obtained. For 37 participants blood sampling failed on the first visit but was successful during a second visit. In 39 cases blood was collected twice, as either the blood collection was not complete at the first occasion or the participant took part in two of the projects that collected blood.

The 9,530 participants in the NTR biobank come from 3,477 nuclear families and some of these families may constitute a larger pedigree. There are 4,293 twins/multiples, 2,070 siblings, 1,209 fathers, 1,587 mothers and 371 spouses of twins/siblings included in the sample. Some persons may have multiple roles, for instance a mother of twins may also be a sister of twins, or a twin may be a spouse of another twin.

There are 1,663 complete twin pairs; 273 monozygotic male, 622 monozygotic female, 144 dizygotic male, 309 dizygotic female, and 299 dizygotic opposite sex complete twin pairs. For 16 complete twin pairs the zygosity is currently unknown but will become available in the near future. In addition, 7 triplet trios participated.

Table 2 provides an overview of family structures available. For example, looking at line 3 there are 1,542 families in which two offspring (this may be multiples or siblings of multiples) participated. In 237 of these families no multiples are present (thus 2 siblings of multiples participated), in 192 families one member of a twin pair and thus one non-twin sibling participated and in 1,113 families a complete twin pair is present (so no additional siblings participated). In 634 of the families in which a complete twin pair participated, these twins were monozygotic. In 818 of the 1,542 families with two offspring no parents participated, in 214 families one parent and in 510 families both parents participated. In 1,385 of the families with two offspring the spouses of twins did not take part, while in 132 cases one spouse and in 25 families two spouses participated.

Sex and age sex distribution. There are 3,554 male and 5,976 female participants in the NTR biobank. Figure 1 presents the age distribution for males and females separately. Based on this distribution we present the data collection for four groups; males younger than 50 yrs (N = 1,996, average age 31.89, SD = 7.66), females younger than 50 yrs (N = 3,741, average age 34.25, SD = 8.09), males aged 50 yrs or older (N = 1,558, average age 61.67, SD = 6.90) and females aged 50 yrs or older (N = 2,235, average age 60.34, SD = 7.48).

Interview data obtained during home visit. Table 3 provides an overview of the data collected at the time of the interview for general health, medication use, smoking behavior, and fasting status for men and women.

Self-reported physical health. The vast majority of the sample reported to be in good or excellent health. In general, the percentage was somewhat lower for women than for men and for older than younger individuals. Men compared to women more often reported excellent health, while the difference in older versus younger individuals was mainly due to fewer old individuals reporting good health.

Medication use. Medication use was markedly higher in the older groups (24% versus 60%) and while women reported more medication use than men, this difference was greatest in the young individuals.

Smoking behavior. The percentage never-smokers was highest among the young individuals with only a small difference between men and women (54.7% vs. 57.5%). In the older individuals fewer men reported never to have smoked than women (23.0% vs. 42.5%). The percentage of current-smokers was highest among the young males (28.1%) and lowest in the older women (19.6%).

Table 2

Family Structure Within the NTR Biobank

N offspring	*N* fam	*N* multiples				*N* MZ pairs	*N* parents			*N* spouses			
		0	1	2	3	0	0	1	2	0	1	2	3
0	333	333	0	0	0	0	1	193	139	332	1	0	0
1	906	200	706	0	0	0	582	123	201	836	70	0	0
2	1542	237	192	1113	0	634	818	214	510	1385	132	25	0
3	529	53	36	434	6	210	264	67	198	457	59	12	1
4	107	24	7	75	1	30	60	22	25	91	13	3	0
5	29	3	3	23	0	13	19	4	6	20	9	0	0
6	17	9	1	7	0	4	13	3	1	14	3	0	0
7	6	1	1	4	0	0	3	1	2	6	0	0	0
8	4	0	0	4	0	1	2	0	2	3	1	0	0
9	3	1	0	2	0	2	2	1	0	3	0	0	0
10	1	0	0	1	0	1	1	0	0	1	0	0	0

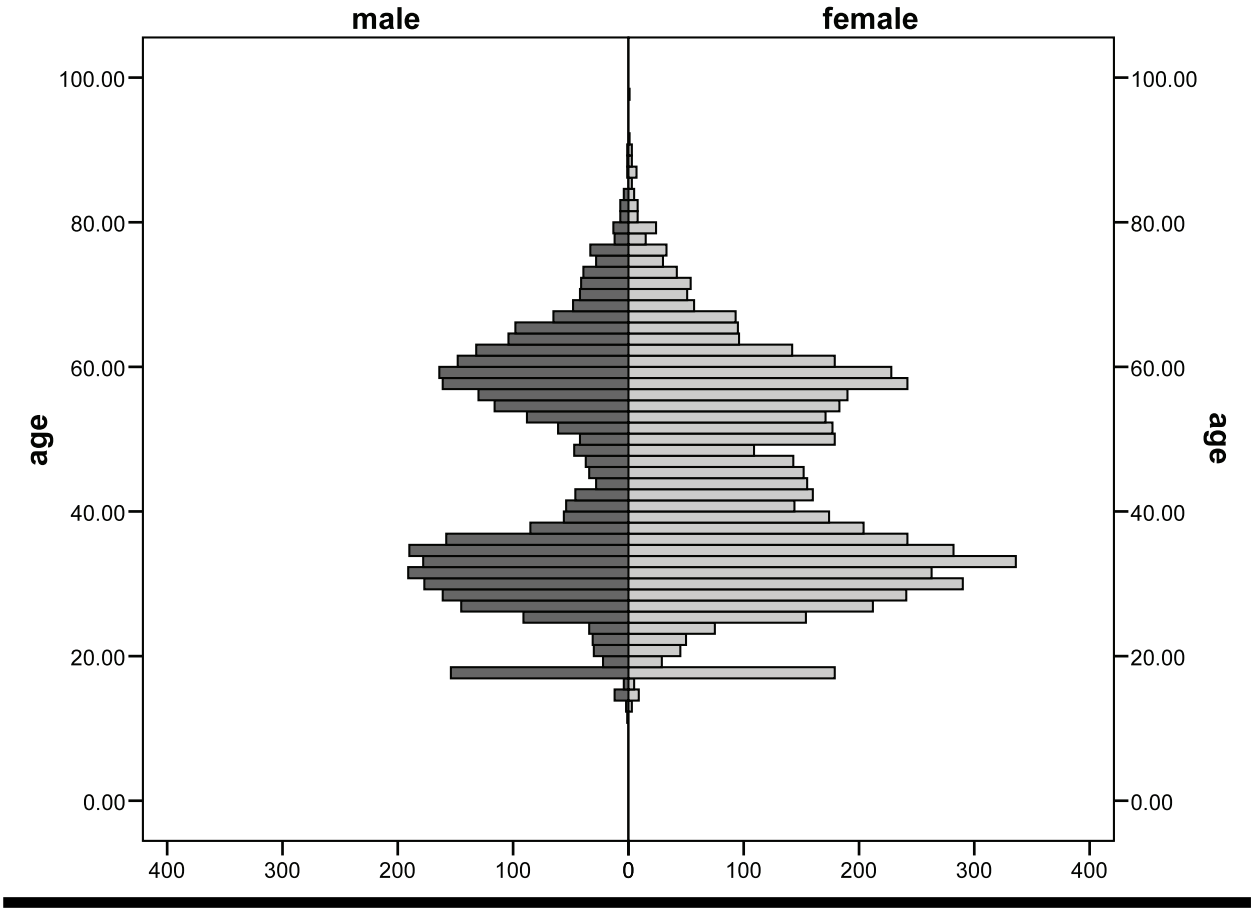

Figure 1

Age distribution in the sample, separate for men and women.

Fasting. More than 94% of the participants kept to the instruction not to eat as of the evening before the blood sampling. This percentage was highest in the older women (96.2%).

Descriptive statistics of continuous variables were estimated in groups of unrelated individuals, so standard deviations are unbiased. This selection results in 1404 males aged < 50 years (average age 32.56, SD = 7.50), 2,307 females aged < 50 years (average age 34.28, SD = 7.89), 1,440 males aged 50 and over (average age 61.76, SD = 6.86) and 1,878 females aged 50 and over (average age 60.33, SD = 7.16). Please note that family members may still be present across groups; for example, father aged 50 years and over and daughter aged < 50 years.

Body Composition

Table 4 presents the descriptive data for body composition for each of the four groups. As expected, men are taller and heavier than women and older individuals are slightly shorter than younger individuals and have a higher weight, BMI, waist and hip circumference and higher waist/hip ratio. Men and women have a similar BMI and hip circumference but overall women have a smaller waist circumference and thus a lower waist/hip ratio.

Parameters Measured in Blood

Table 5, 6 and 7 present the average levels and standard deviations for parameters measured in plasma for the four groups. Only data from fasting individuals were included in the descriptives.

Lipid profile. As can be seen in Table 5, data for total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides show a less favourable lipid profile in older individuals, who have on average higher total cholesterol, LDL cholesterol, and triglyceride levels and lower HDL cholesterol levels than young individuals. In both age groups, the most marked difference between men and women is the higher HDL cholesterol level in women and in the young women this is accompanied by lower LDL cholesterol and triglyceride levels than in the men.

Glucose metabolism. Compared to the younger individuals, glucose and Hba1c are slightly increased but insulin levels are clearly higher in the older individuals. Overall, men and women do not differ much in their average glucose, insulin and Hba1c levels, although the older women have somewhat higher insulin levels than the older men.

Haematology profile. Table 6 presents the descriptives for the haematological parameters. Few marked differ-

Table 3

Descriptives for Subjective Health, Medication Use, Smoking Behavior and Fasting Status

		< 50 years				≥ 50 years			
		Men		Women		Men		Women	
		N	%	*N*	%	*N*	%	*N*	%
Subjective health	Bad	29	1.5	64	1.7	29	1.9	49	2.2
	Poorly	54	2.8	221	6.0	87	5.7	158	7.2
	Reasonable	189	9.7	396	10.8	165	10.7	282	12.9
	Good	1451	74.7	2767	75.3	1074	70.0	1507	68.9
	Excellent	227	11.6	228	6.2	180	11.7	191	8.7
	Missing	46	—	65	—	23	—	48	—
Medication use	No	1583	81.1	2661	72.8	639	41.6	853	38.9
	Yes	370	18.9	995	27.2	898	58.4	1338	61.1
	Missing	43	—	85	—	21	—	44	—
Smoking behavior	Current smoker	551	28.1	796	21.5	364	23.6	433	19.6
	Past smoker	336	17.2	775	21.0	824	53.4	838	37.9
	Never smoker	1071	54.7	2127	57.5	355	23.0	941	42.5
	Missing	38	—	43	—	15	—	23	—
Fasting	Yes	1824	93.7	3474	94.1	1479	95.9	2129	96.2
	No	123	6.3	217	5.9	64	4.1	84	3.8
	Missing	49	—	50	—	15	—	22	—

Table 4

Means and Standard Deviations (SD) for Body Composition Measures in Unrelated[1] Participants

	< 50 years						≥ 50 years					
	Men			Women			Men			Women		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Height (cm)	1404	183.49	7.33	2303	170.19	6.45	1436	178.95	7.03	1869	167.10	6.30
Weight (kg)	1400	82.48	13.16	2298	70.24	13.58	1435	85.74	12.80	1868	74.15	13.38
BMI (kg/m^2)	1400	24.49	3.58	2294	24.31	4.43	1432	26.75	3.50	1862	25.57	4.69
Waist circumference (cm)	1388	102.71	7.16	2279	102.86	9.30	1415	104.35	6.63	1835	105.19	9.41
Hip circumference (cm)	1390	87.61	10.67	2279	79.86	11.17	1420	97.74	10.96	1842	88.98	12.52
Waist/hip ratio (cm/cm)	1388	0.852	0.066	2276	0.775	0.072	1415	0.936	0.075	1835	0.844	0.079

Note: [1] Means and *SD* are based on data from unrelated participants only within each of the four groups (related participants are present across groups; e.g., father over 50 years and a daughter < 50 years).

ences can be seen between the four groups. In the older age groups the difference between men and women in the neutrophil count and percentage becomes stronger, with women having a lower percentage but higher neutrophil count than men. The number and percentage of lymphocytes is larger in men than in women and in both sexes the percentage is higher in older individuals though the count is somewhat lower. The red blood cell count is somewhat higher in young men than in the other three groups and the number of platelets is lower in men than in women in both age groups.

CRP. As can be seen in Table 7, CRP levels are higher in women than in men and higher in the older than in the younger individuals.

Fibrinogen. Fibrinogen levels show a similar pattern as CRP, levels are higher in women and in older individuals.

Liver enzymes. Overall, women have lower levels of GGT, ALT, AST, and creatinine, in both age groups. In older individuals levels of AST and creatinine show a slight increase, but a strong increase can be seen for GGT, in particular for the women, while in older men the ALT level is increased compared to the younger men.

Genotypic Information

DNA. Isolated DNA from blood samples is currently available for 9,461 individuals. The average DNA concentration for these samples is 226.3 (SD 124.98) ng/µl. In addition to DNA extracted from blood samples, buccal DNA was obtained for 2,855 of the participants, including 897 complete twin pairs and 1 set of triplets. As part of the GAIN-NIH Initiative, genome-wide SNP data were obtained using the Perlegen/Affymetrix 5.0 600K platform for 1,875 indi-

Table 5

Means and Standard Deviations (SD) for Lipid Profile and Glucose Metabolism Parameters for Unrelated, Fasting Participants

	< 50 years						≥ 50 years					
	Men			Women			Men			Women		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Total cholesterol (mmol/L)	1290	4.84	1.00	2123	4.77	0.90	1353	5.33	1.04	1768	5.70	1.08
HDL (mmol/L)	1289	1.23	0.31	2122	1.49	0.37	1353	1.26	0.35	1768	1.54	0.43
LDL (mmol/L)	1284	2.95	0.89	2121	2.76	0.82	1348	3.31	0.97	1764	3.50	1.00
Triglycerides (mmol/L)	1290	1.44	0.92	2123	1.13	0.78	1353	1.65	0.93	1768	1.45	0.72
Glucose (mmol/L)	1285	5.46	1.04	2105	5.20	0.65	1342	5.99	1.35	1748	5.73	1.03
Insulin (μIU/ml)	1246	9.02	7.67	2049	8.70	5.51	1317	11.25	11.63	1712	10.78	10.41
Hba1c (%)	1265	5.26	0.60	2081	5.26	0.61	1334	5.54	0.74	1732	5.53	0.69

Table 6

Means and Standard Deviations (*SD*) for Haematological Parameters for Unrelated, Fasting Participants

		< 50 years						≥ 50 years					
		Men			Women			Men			Women		
	Measure	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
WBC	Count	1247	6.43	1.72	2090	6.76	1.90	1331	6.63	2.27	1734	6.30	1.89
Neutrophils	%	1266	51.99	8.24	2079	53.65	9.39	1327	54.94	8.24	1730	51.93	9.16
	Count	1176	3.43	1.29	1978	3.69	1.45	1325	3.68	1.37	1722	3.32	1.34
Lymphocytes	%	1269	35.26	7.48	2084	34.97	8.47	1327	31.72	7.70	1730	35.89	8.59
	Count	1176	2.25	0.65	1978	2.33	0.73	1325	2.09	1.39	1722	2.23	0.85
Monocytes	%	1269	8.96	2.30	2084	7.73	1.98	1327	9.29	2.57	1730	8.29	2.16
	Count	1176	0.58	0.21	1978	0.51	0.16	1325	0.60	0.22	1722	0.51	0.17
Eosinophils	%	1269	3.33	2.02	2084	3.02	2.35	1326	3.56	2.04	1729	3.23	1.91
	Count	1176	0.21	0.14	1978	0.21	0.21	1325	0.23	0.16	1722	0.20	0.14
Basophils	%	1228	0.47	1.24	2044	0.58	1.33	1327	0.47	0.70	1729	0.62	1.93
	Count	1176	0.02	0.13	1978	0.03	0.11	1325	0.02	0.06	1720	0.03	0.15
RBC (10^{12}/L)	Count	1276	5.04	0.36	2093	4.43	0.33	1339	4.84	0.41	1747	4.48	0.36
hemoglobulin (mmol/L)	Count	1276	9.51	0.60	2094	8.34	0.67	1338	9.30	0.76	1747	8.52	0.72
hematocrit	Ratio	1276	0.54	1.92	2094	0.57	2.49	1339	0.60	2.46	1747	0.68	3.22
MCV (fL)	Count	1275	91.26	4.16	2093	91.19	4.92	1339	93.40	4.54	1747	92.31	4.70
MCH (fmol)	Count	1276	2.03	1.99	2093	2.06	2.23	1339	2.08	2.08	1747	2.14	2.61
MCHC (mmol/L)	Count	1276	21.70	17.69	2093	21.28	13.83	1339	21.07	12.17	1746	21.19	13.14
RDW	%	1181	12.22	0.61	1990	12.39	0.98	1329	12.61	0.83	1733	12.55	0.95
platelets	Count	1273	240.03	53.88	2092	267.86	63.77	1339	235.50	61.31	1745	262.18	63.18
MPV (fL)	Count	1180	8.72	0.87	1994	8.85	0.92	1339	8.67	0.89	1745	8.69	0.90

Note: Unless stated differently, all cell counts are in 10^9/L.

viduals (see Boomsma et al., 2008, Sullivan et al., 2009), this same chip was later also used for 25 extra participants. In addition to GWAS data for unrelated individuals, data were also included for trios, that is father, mother and one offspring (Pardo et al., 2009). Furthermore, as part of the GenomEUtwin consortium genome-wide SNP data were obtained in 277 MZF twin individuals using the Illumina 370K platform, and as part of a SPINOZA NWO grant GWA data were obtained for a further 1,352 individuals using the Illumina 660 platform. In total, GWA data are currently available for 3,524 individuals, with 111 of these individuals genotyped on multiple platforms. Genotyping of additional samples, including all complete twin pairs, is currently ongoing.

RNA expression. All materials collected for RNA expression studies, both unstimulated and stimulated samples, were transferred to RUCDR where the materials are prepared for future RNA expression studies. Prior to shipment of all samples, we tested the protocol in 98 samples, which showed good results. Currently totalRNA and miRNA have been extracted and cDNA produced for over 90% of the PAXgene samples.

Cell lines. Materials for creating transformed cell lines were transported in liquid nitrogen dry shippers to the RUCDR. At present transformed cell lines are estab-

Table 7

Means and Standard Deviations (*SD*) for CRP, Fibrinogen and Liver Enzymes for Unrelated, Fasting Participants

	< 50 years						≥ 50 years					
	Men			Women			Men			Women		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
CRP (mg/L)	1265	2.47	8.71	2081	3.82	6.89	1343	3.77	7.69	1738	4.03	8.34
Fibrinogen (g/l)	1174	2.48	0.61	1962	2.69	0.67	1327	2.98	0.74	1712	3.05	0.73
GGT (U/L)	1169	34.96	26.48	1970	23.08	18.21	1330	45.66	45.74	1730	33.83	39.17
ALT (U/L)	1158	12.67	6.20	1868	9.06	5.06	1293	10.92	6.51	1645	9.58	6.05
AST (U/L)	1166	23.33	7.78	1967	19.01	6.60	1324	23.57	8.40	1719	21.52	7.32
Creatinine (mol/L)	1167	94.71	17.16	1967	80.34	30.11	1325	98.14	18.11	1719	81.43	12.85

lished for 4,072 samples with a failure rate of less than 0.1%. These cell lines will provide a renewable resource for ongoing genetic studies.

Additional Phenotype Information From Participation in Longitudinal Surveys and Other Projects

Of the 9,530 participants in the biobank, 90% also participated in longitudinal surveys. Of the biobank participants, 1,654 also took part in studies on cardiovascular functioning, including heart rate and blood pressure monitoring, and 1,202 in studies on cognitive and brain function.

Discussion

Finding genes for complex diseases and traits has proven to be difficult, but as larger studies are performed more polymorphims are found that associate with complex phenotypes. An important requirement is the availability of biobanks with collections of biological material from a large number of participants. In addition, availability of extended phenotypic information, preferably not only biological but also environmental, is needed. The NTR Biobank has such information available on 9,530 participants. The common reason for inclusion in the study was being a twin or multiple or a family member (either biological or nonbiological) of a multiple. Twinning occurs in all layers of society and thus, although females and higher educated individuals are somewhat overrepresented and non-caucasian individuals are underrepresented, the NTR Biobank may be seen as a good representation of the majority of the Dutch population. This is somewhat different from the many biobanks that have been established to study particular diseases.

In addition to the information collected during the home visit, questionnaire information through the longitudinal survey on health, lifestyle and personality is also available for the majority of participants. The influence of genes may thus be examined for a wide variety of phenotypes in the area of physical health, lifestyle, personality and psychiatric disorder. Indeed, genome wide scans within the NTR biobank have been conducted for smoking (Vink et al., 2009), exercise behavior (de Moor et al., 2009), depression (conducted in collaboration with the Netherlands Study of Depression and Anxiety, see Boomsma et al., 2008, Sullivan et al., 2009), lipids (Aulchenko et al., 2009), glucose (Dupuis et al., 2009; Prokopenko et al., 2009) and haematology (Fereirra et al., 2009). A large group of participants completed more than one survey and in the coming years all participants will continue to be asked to participate in the longitudinal survey. Thus, important information on the development of disease and related traits is available and will be added to in the future.

In the surveys individuals were also asked to indicate whether data may be linked to external health registers, such as the Cancer Registration and the Dutch central archive of pathological reports (Palga, Vink et al., submitted). Such linkage will allow us to move outside measured risk factors and questionnaire-based information on disease to clinical diagnosed diseases.

The extent of the biological material collection in the NTR Biobank is such that recent developments in genomic analysis can be easily accommodated. For example, selected blood samples have been used for an epigenetic study (Talens et al., in press) and urine samples for a metabolomic pilot study (unpublished data). The creation of transformed cell lines and the processing of the RNA basal and expression samples add to the possibilities for future studies.

Conclusion

The NTR biobank presents a unique combination of a twin family population-based design, extended biological information and longitudinal survey data collection and offers opportunities for DNA based and RNA expression studies as well as for future metabolomic and proteomic projects. The NTR biobank thus provides a valuable resource for the study of genetic determinants of individual differences in mental and physical health.

Funding

Funding was obtained from the Netherlands Organisation for Scientific Research (NWO): genetic basis of anxiety and depression (904-61-090); Resolving cause and effect in the association between exercise and wellbeing (904-61-193); 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 (40-0056-98-9032), Genotype/phenotype database for behavior and genetic epidemiological studies (40-00506-98-9032), CMSB: Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); the VU University Centre for Neurogenomics and Cognitive Research (CNCR); the European Science Foundation (ESF): Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); the European Science Council (ERC) Genetics of Mental Illness (230374); Rutgers University Cell and DNA Repository cooperative agreement (NIMH U24 MH068457-06), Integration of genomics & transcriptomics in normal twins and major depression (NIMH 1RC2MH089951-01); A 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. This resource is further augmented through a collaboration with the NIMH Center for Genetics Studies (http:\\NIMHGenetics.org) that will make a substantial subset of the NTR biobank available to the research community.

Acknowledgments

We would like to acknowledge Angelique van Bruggen, the technicians of GBS Leiden under supervision of Jacoline Neuteboom, and the VU University IT department for their help during various phases of the NTR Biobank Project.

References

Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., Martin, N.G., Pedersen, N.L., Kyvik, K.O., Kaprio, J., Hofman, A., Freimer, N.B., Jarvelin, M.R., Gyllensten, U., Campbell, H., Rudan, I., Johansson, A., Marroni, F., Hayward, C., Vitart, V., Jonasson, I., Pattaro, C., Wright, A., Hastie, N., Pichler, I., Hicks, A. A., Falchi, M., Willemsen, G., Hottenga, J. J., de Geus, E. J., Montgomery, G. W., Whitfield, J., Magnusson, P., Saharinen, J., Perola, M., Silander, K., Isaacs, A., Sijbrands, E. J., Uitterlinden, A. G., Witteman, J. C., Oostra, B. A., Elliott, P., Ruokonen, A., Sabatti, C., Gieger, C., Meitinger, T., Kronenberg, F., Döring, A., Wichmann, H. E., Smit, J. H., McCarthy, M. I., van Duijn, C. M., Peltonen, L., ENGAGE Consortium. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genetics, 41*, 47–55.

Bartels, M., van Beijsterveldt, C. E., Derks, E. M., Stroet, T. M., Polderman, T. J., Hudziak, J. J., Boomsma, D. I. (2007). Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10,* 3–11.

Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., Posthuma, D., van Beijsterveldt, T. C., Hudziak, J. J., Bartels, M., Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9,* 849–857.

Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C., de Geus, E. J., Beem, A. L., Mulder, E. J., Derks, E. M., Riese, H., Willemsen, G. A., Bartels, M., van den Berg, M., Kupper, N. H., Polderman, T. J., Posthuma, D., Rietveld, M. J., Stubbe, J. H., Knol, L. I., Stroet, T., van Baal, G. C. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research, 5*, 401–406.

Boomsma, D. I., Willemsen, G., Sullivan, P. F., Heutink, P., Meijer, P., Sondervan, D., Kluft, C., Smit, G., Nolen, W. A., Zitman, F. G., Smit, J. H., Hoogendijk, W. J., van Dyck, R., de Geus, E. J., Penninx, B. W. (2008). Genome-wide association of major depression: Description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *European Journal of Human Genetics, 16,* 335–342.

Boomsma, D. I., Willemsen, G., Vink, J. M., Bartels, M., Groot, P., Hottenga, J. J., van Beijsterveldt, C. E., Stroet, T., van Dijk, R., Wertheim, R., Visser, M., van der Kleij, F. (2008). Design and implementation of a twin-family database for behavior genetics and genomics studies. *Twin Research and Human Genetics, 11,* 342–348.

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., Hughes, T. E., Groop, L., Altshuler, D., Almgren, P., Florez, J. C., Meyer, J., Ardlie, K., Bengtsson Boström, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H. N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Råstam, L., Speliotes, E. K., Taskinen, M. R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjögren, M., Sterner, M., Surti, A., Svensson, M., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S. B., Chirn, G. W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., Purcell, S. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science, 316,* 1331–1336.

Distel, M. A., Rebollo-Mesa, I., Willemsen, G., Derom, C. A., Trull, T. J., Martin, N. G., Boomsma, D. I. (2009). Familial resemblance of borderline personality disorder features: Genetic or cultural transmission? *PLoS One, 4,* e5334.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Mägi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R., Egan, J. M., Lajunen, T., Grarup, N., Sparsø, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proença, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S. A., Payne, F., Roccasecca, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Bumpstead, S. J, Charpentier, G, Chen, Y. D., Chines, P., Clarke, R., Coin, L. J., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N., de Geus, E. J., Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R, Franzosi, M. G., Galan, P, Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S, Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A. L., Hassanali, N., Hayward, C., Heath S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hofman, A, Hui J, Hung J, Isomaa, B, Johnson, P. R., Jørgensen, T, Jula A, Kaakinen M, Kaprio J, Kesaniemi, Y. A., Kivimaki, M, Knight, B, Koskinen, S, Kovacs, P, Kyvik K. O., Lathrop G. M., Lawlor, D. A., Le Bacquer, O., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martínez-Larrad, M. T., McAteer, J. B., McCulloch L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra B. A., Orrù, M., Pakyz, R., Palmer, C. N., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F., Pichler, I., Polasek, O., Posthuma, D., Potter S. C., Pouta, A, Province, M. A., Psaty B. M., Rathmann, W., Rayner N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurethsson, G., Sijbrands, E. J., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvänen, A. C., Tanaka, T., Thorand, B., Tichet J, Tönjes, A., Tuomi, T., Uitterlinden A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H, Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C., Yarnell, J. W., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J. H., Zillikens, M. C.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium, Borecki, I. B., Loos, R. J., Meneton, P., Magnusson, P. K., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Ríos, M., Morris, A. D., Lind, L, Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M, Kao W. H., Pankow J. S., Sampson M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll M, Campbell, H, Wilson, J. F.; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators, Bergman, R. N., Buchanan, T. A., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Valle, T. T., Altshuler, D., Rotter, J. I., Siscovick, D. S., Penninx, B. W., Boomsma, D. I., Deloukas, P., Spector, T. D., Frayling, T. M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., van Duijn CM, Aulchenko, Y. S., Cao A, Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., Jarvelin, M. R., Waterworth D. M., Vollenweider P, Peltonen L, Mooser V, Abecasis, G. R., Wareham, N. J., Sladek, R., Froguel, P., Watanabe, R. M., Meigs, J. B., Groop, L., Boehnke, M., McCarthy, M. I., Florez, J. C., Barroso, I. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics, 42,*105–116.

Estourgie-van Burk, G. F. Bartels, M., Boomsma, D. I., Delemarre-van de Waal, H. A. (2009). Body size of twins compared with siblings and the general population: From birth to late adolescence. *Journal of Pediatrics*, *156*, 586–591.

Ferreira, M. A., Hottenga, J. J., Warrington, N. M., Medland, S. E., Willemsen, G., Lawrence, R. W., Gordon, S., de Geus, E. J., Henders, A. K., Smit, J. H., Campbell, M. J., Wallace, L., Evans, D. M., Wright, M. J., Nyholt, D. R., James, A. L., Beilby, J. P., Penninx, B. W., Palmer, L. J., Frazer, I. H., Montgomery, G. W., Martin, N. G., Boomsma, D. I. (2009). Sequence variants in three loci influence monocyte counts and erythrocyte volume. *American Journal of Human Genetics, 85,* 745–749.

Friedewald, W. T., Levy, R. I., Fredrickson, D. S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical Chemistry, 18,* 499–502.

van Grootheest, D. S., Boomsma, D. I., Hettema, J. M., Kendler, K. S. (2008). Heritability of obsessive-compulsive symptom dimensions. *American* Journal *of Medical Genetics Part* B*:* Neuropsychiatric *Genetics, 147B*, 473–478.

Hoekstra, C. (2008). *Dizygotic twinning*. Ipskamp, the Netherlands: PrintPartners.

Hoekstra, C., Meijer, P., Kluft, C., Heutink, P., Smit, G., de Geus, E., Smit, J. H., van Bruggen, A., Montgomery, G. W., Boomsma, D. I. (2004). Genetics of dizygotic twinning: A feasibility study for a biobank. *Twin Research, 7,* 556–563.

Kettunen, J., Perola, M., Martin, N. G., Cornes, B. K., Wilson, S. G., Montgomery, G. W., Benyamin, B., Harris, J. R., Boomsma, D., Willemsen, G., Hottenga, J. J., Slagboom, P. E., Christensen, K., Kyvik, K. O.,

Sørensen, T. I., Pedersen, N. L., Magnusson, P. K., Andrew, T., Spector, T. D., Widen, E., Silventoinen, K., Kaprio, J., Palotie, A., Peltonen, L. (2009). Multicenter dizygotic twin cohort study confirms two linkage susceptibility loci for body mass index at 3q29 and 7q36 and identifies three further potential novel loci. *International Journal of Obesity, 33,* 1235–1242.

Kupper, N., Willemsen, G., Boomsma, D. I., de Geus, E. J. (2006). Heritability of indices for cardiac contractility in ambulatory recordings. *Journal of Cardiovascular Electrophysiology, 17,* 877–883.

Kupper, N., Willemsen, G., Riese, H., Posthuma, D., Boomsma, D. I., de Geus, E. J. (2005). Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertension, 45,* 80–85.

Meulenbelt, I., Droog, S., Trommelen, G. J., Boomsma, D. I., Slagboom, P. E. (1995). High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *American Journal of Human Genetics, 57,* 1252–1254.

Middeldorp, C. M., de Geus, E. J., Beem, A. L., Lakenberg, N., Hottenga, J. J., Slagboom, P. E., Boomsma, D. I. (2007). Family based association analyses between the serotonin transporter gene polymorphism (5-HTTLPR) and neuroticism, anxiety and depression. *Behavior Genetics, 37,* 294–301.

de Moor. M. H,, Liu, Y. J., Boomsma, D. I., Li, J., Hamilton, J. J., Hottenga, J. J., Levy, S., Liu, X. G., Pei, Y. F., Posthuma, D., Recker, R. R., Sullivan, P. F., Wang, L., Willemsen, G., Yan, H., de Geus, E. J., Deng, H. W. (2009). Genome-Wide Association Study of Exercise Behavior in Dutch and American Adults. *Medicine and Science in Sports and Exercise,* epub September 2.

Pardo, L., Bochdanovits, Z., de Geus, E., Hottenga, J. J., Sullivan, P., Posthuma, D., Penninx, B. W., Boomsma, D., Heutink, P. (2009). Global similarity with local differences in linkage disequilibrium between the Dutch and HapMap-CEU populations. *European Journal of Human Genetics, 17,* 802–810.

Prokopenko, I., Langenberg, C., Florez, J. C., Saxena, R., Soranzo, N., Thorleifsson, G., Loos, R. J., Manning, A. K., Jackson, A. U., Aulchenko, Y., Potter, S. C., Erdos, M. R., Sanna, S., Hottenga, J. J., Wheeler, E., Kaakinen M, Lyssenko V, Chen W. M., Ahmadi K, Beckmann J. S., Bergman R. N., Bochud M, Bonnycastle L. L., Buchanan T. A., Cao A, Cervino A, Coin L, Collins F. S., Crisponi L, de Geus E. J., Dehghan A, Deloukas P, Doney A. S., Elliott P, Freimer N, Gateva V, Herder C, Hofman A, Hughes T. E., Hunt S, Illig T, Inouye M, Isomaa B, Johnson T, Kong A, Krestyaninova M, Kuusisto J, Laakso M, Lim N, Lindblad U, Lindgren C. M., McCann O. T., Mohlke K. L., Morris A. D., Naitza S, Orrù M, Palmer C. N., Pouta A, Randall J, Rathmann W, Saramies J, Scheet P, Scott L. J., Scuteri A, Sharp S, Sijbrands E, Smit J. H., Song K, Steinthorsdottir V, Stringham H. M., Tuomi T, Tuomilehto J, Uitterlinden A. G., Voight B. F., Waterworth D, Wichmann H. E., Willemsen G, Witteman J. C., Yuan X, Zhao J. H., Zeggini E, Schlessinger D, Sandhu M, Boomsma D. I., Uda M, Spector T. D., Penninx B. W., Altshuler D, Vollenweider P, Jarvelin M. R., Lakatta E, Waeber G, Fox C. S., Peltonen L, Groop L. C., Mooser, V., Cupples, L. A., Thorsteinsdottir, U., Boehnke, M., Barroso, I., Van Duijn, C., Dupuis, J., Watanabe, R. M., Stefansson, K., McCarthy, M. I., Wareham, N. J., Meigs, J. B., Abecasis, G. R. (2009). Variants in MTNR1B influence fasting glucose levels. *Nature Genetics, 41,* 77–81.

Scuteri, A., Sanna, S., Chen, W. M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orrú, M., Usala, G., Dei, M., Lai, S., Maschio, A., Busonero, F., Mulas, A., Ehret, G. B., Fink, A. A., Weder, A. B., Cooper, R. S., Galan, P., Chakravarti, A., Schlessinger, D., Cao, A., Lakatta, E., Abecasis, G. R. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genetics, 3,* e115.

Smit. D. J, Boersma, M., van Beijsterveldt, C. E., Posthuma, D., Boomsma, D. I., Stam, C. J., de Geus, E. J. (2010). Endophenotypes in a Dynamically Connected Brain. *Behavior Genetics, 40*, 167–177.

Spijker, S., van de Leemput, J. C., Hoekstra, C., Boomsma, D. I., Smit, A. B. (2004). Profiling gene expression in whole blood samples following an in-vitro challenge. *Twin Research, 7,* 564-70.

Sullivan, P. F., de Geus, E. J., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T., Arolt, V., Baune, B. T., Blackwood, D., Cichon, S., Coventry, W. L., Domschke, K., Farmer, A., Fava, M., Gordon, S. D., He, Q., Heath, A. C., Heutink, P., Holsboer, F., Hoogendijk, W. J., Hottenga, J. J., Hu, Y., Kohli, M., Lin, D., Lucae, S., Macintyre, D. J., Maier, W., McGhee, K. A., McGuffin, P., Montgomery, G. W., Muir, W. J., Nolen, W. A., Nöthen, M. M., Perlis, R. H., Pirlo, K., Posthuma, D., Rietschel, M., Rizzu, P., Schosser, A., Smit, A. B., Smoller, J. W., Tzeng, J. Y., van Dyck, R., Verhage, M., Zitman, F. G., Martin, N. G., Wray, N. R., Boomsma, D. I., Penninx, B. W. (2009). Genome-wide association for major depressive disorder: A possible role for the presynaptic protein piccolo. *Molecular Psychiatry, 14,* 359–375.

Talens, R. P., Boomsma, D. I., Tobi, E. W., Kremer, D, Jukema, J. W., Willemsen, G., Putter, H., Slagboom, P. E., Heijmans, B. T. (In press). Variation, patterns and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *The FASEB Journal.*

van Dijk, B. A., Boomsma, D. I., de Man, A. J. (1996). Blood group chimerism in human multiple births is not rare. *American Journal of Medical Genetics, 61*, 264–268.

van 't Ent, D., van Beijsterveldt, C. E., Derks, E. M., Hudziak, J. J., Veltman, D. J., Todd, R. D., Boomsma, D. I., De Geus, E. J. (2009). Neuroimaging of response interference in twins concordant or discor-

dant for inattention and hyperactivity symptoms. *Neuroscience, 164,* 16–29.

Vink, J. M., Smit, A. B., de Geus, E. J., Sullivan, P., Willemsen, G., Hottenga, J. J., Smit, J. H., Hoogendijk, W. J., Zitman, F. G., Peltonen, L., Kaprio, J., Pedersen, N. L., Magnusson, P. K., Spector, T. D., Kyvik, K. O., Morley, K. I., Heath, A. C., Martin, N. G., Westendorp, R. G., Slagboom, P. E., Tiemeier, H., Hofman, A., Uitterlinden, A. G., Aulchenko, Y. S., Amin, N., van Duijn, C., Penninx, B. W., Boomsma, D. I. (2009). Genome-wide association study of smoking initiation and current smoking. *American Journal of Human Genetics, 84,* 367–379.

Vink, J. M., van Kemenade, F. J., Meijer, C. J. L. M., Casparie, M. K., Meijer, G. A., Boomsma, D. I. (Submitted), Cervix smear abnormalities: A study in female twins, their mothers and sisters. *European Journal of Human Genetics.*