Twins in Genetic Research

A. Busjahn and D. Boomsma

37

INTRODUCTION
HISTORY OF TWIN RESEARCH
TYPES OF TWIN STUDY DESIGNS
DECIPHERING THE HUMAN GENOME
USE OF TWINS IN MOLECULAR GENETIC STUDIES
TWIN REGISTRIES WORLDWIDE

INTRODUCTION

This chapter introduces the field of genetic research, in which twin studies are a powerful approach to exploring the genetic basis of complex traits. Twin studies take advantage of the existence of mono- and dizygotic (MZ and DZ) twins and use the comparison of resemblances of MZ and DZ twins as the basis for analysis of variations in human traits and diseases. We describe the basic methodology for twin studies, including recent extensions exploiting advances in molecular genetics. We illustrate the value of twin studies by giving a few selected examples from various fields of medicine and psychology, and provide an overview of twin registries which collect data on twins and their relatives worldwide. These registries have been established to obtain an insight into the genetic epidemiology of complex traits and diseases, to study the interaction of genetic factors with sex, age and lifestyle factors and to study the causes of comorbidity between different traits and diseases. Obesity, diabetes, hypertension and psychiatric disorders are examples of common diseases that are a result of various genetic susceptibility factors interacting with environmental risks. Twin registries have been instrumental in establishing the genetic component in susceptibility to these conditions. So far, however, it has been difficult to identify the responsible genes. Because of the design and the often (very) large sample sizes of twin registers, they offer unique opportunities for selected sampling for gene hunting, employing both linkage and association studies.

Genetic epidemiology aims to disentangle and quantify the contributions of genes, shared environment, individual-specific environment and their interactions to variations in human traits. Research questions address the etiology of individual differences (what statisticians call variation) in health and disease, or in continuously varying traits such as height, weight or blood pressure. Variation for a trait in a particular population may be caused by genetic differences between individuals and/or by differences in their environment. The effects of genes and environment may be additive, or they may interact with each other. To explore the etiology of individual differences, or the etiology of clustering of these differences within individuals, it is necessary to collect data from non-random samples. The study sample can include relatives who are genetically related but who grew up in unrelated environments (the so-called adoption design), or relatives who grew up in similar environments but who are of different genetic relatedness (for example the twin design). If the exposure to environmental risk factors can be assessed, these designs also allow the quantification of gene–environment interaction in shaping a particular trait.

Recent advances in statistical modeling allow the simultaneous analysis of many variables in genetic studies. Such advances make new types of analyses possible, such as the multivariate analysis of causes of comorbidity between disorders; analysis of the development of psychopathology over time; the inclusion of covariates in linkage analyses; and the estimation of heritability and linkage conditional to exposure to environmental risk factors. These improvements in

Figure 37.1 Augustine of Hippo

data analysis have in turn led to the establishment of large registries of twins that no longer focus on the assessment of a single phenotype, but collect a wide range of traits and environmental risk factors in twins, as well as in their family members.

HISTORY OF TWIN RESEARCH

Twins have captured the curiosity of researchers for centuries, and proposals to use them as a natural experiment in empirical studies stem from as early as 415 by Augustine of Hippo[1] (Figure 37.1). Galton's classic article on twins[2], published in the 19th century, is often cited as the first iteration of the classical twin method, although it is uncertain whether Galton knew of the distinction between MZ and DZ twins. The systematic analysis of similarity of MZ and DZ twins was introduced by Hermann Werner Siemens, a dermatologist, who formulated the twin rule of pathology: any heritable disease will be more concordant in identical twins than in non-identical twins, and concordance will be even lower in non-siblings[3]. When studying moles, Siemens came up with the clever idea of combining correlation analysis and twin data. He correlated mole counts in one twin with mole counts in the co-twin and contrasted this correlation between MZ and DZ twin pairs. The correlation for mole count in MZ twins, who share all, or nearly all, of their genetic material was 0.4. In DZ twins, who are genetically 50% identical on average, the correlation was only 0.2. The results indicated the importance of genetic factors to the variation in mole count. The larger genetic resemblance in MZ twins is associated with their larger resemblance for the phenotype under study.

TYPES OF TWIN STUDY DESIGNS

Classical twin studies

The classical twin study compares phenotypic resemblances of MZ and DZ twins. MZ twins derive from a single fertilized egg and therefore inherit identical genetic material. DZ twins are genetically as similar as other siblings and share, on average, 50% of their segregating genes. Comparing the resemblance of MZ twins for a trait or disease with resemblance of DZ twins therefore offers an initial estimate of the extent to which genetic variation determines phenotypic variation of that trait. If MZ twins resemble each other more than DZ twins, then the heritability (h^2) of the phenotype can be estimated from twice the difference between MZ and DZ correlations. For example, typical MZ and DZ correlations for depression are around 0.4 and 0.2[4], and therefore heritability is estimated at ~40%. A different pattern of correlations is usually observed for lifestyle factors, indicating the importance of the shared family environment. For smoking initiation during adolescence, typical MZ and DZ correlations are 0.9 and 0.7, leading to a heritability estimate of 40%, but also pointing out the importance of shared environment[5]. The proportion of the variance that is due to shared environment is the difference between the total twin correlation and the part that is explained by heritability (i.e. $r_{MZ} - h^2$ in MZ or $r_{DZ} - h^2/2$ in DZ twins). For smoking initiation, this estimate is around 50% (0.9 – 0.4 based on MZ data or 0.7 – 0.2 based on DZ twin data).

The application of this type of analysis led to substantial changes in the way we think about the determinants of health and disease and the causes of individual differences in normal and abnormal behavior. During the past decade, a shift has taken place from strict environmental explanations to a more balanced view that recognizes the importance of genes, for example in autism and in attention deficit hyperactivity disorder in children, or in the development of dependence on nicotine, alcohol and other drugs in adults.

Improvements in analysis

Quantitative traits assessed in MZ and DZ twins are traditionally analyzed using analysis of variance (ANOVA) and intraclass correlations to summarize twin resemblance. In large samples the intraclass correlation equals the better known Pearson correlation coefficient. Resemblance between twins for qualitative traits, such as the presence or absence of disease, can be summarized with concordance rates or with tetrachoric correlations. Tetrachoric correlations give the resemblance between relatives for the underlying disease liability and may be used to obtain the heritability of disease liability by doubling the difference between MZ and DZ correlations.

Although heritabilities based on (differences between) correlations can give a good first approximation of heritability, they are of limited use when analyzing and comparing data from different groups, or from longitudinal designs. Multigroup analyses can involve comparison of the genetic architecture of traits in males and females, in groups with different exposures to environmental risk factors or in different cohorts. Structural equation modeling (SEM), or covariance structure modeling, is a more general approach to the analysis of familial resemblances. In SEM, genotypic and environmental effects are modeled as the contributions of unmeasured (latent) variables to the possibly multivariate phenotypic differences between individuals[6]. The latent factors represent the effects of many unidentified influences. For a genetic factor, these effects are due to a possibly large but unknown number of polygenes. The contributions of the latent variables are estimated as regression coefficients in the linear regression of the observed variables on the latent variables. A number of widely available software programs, such as LISREL or Mx, allow estimation of parameters by means of normal theory maximum likelihood (ML) and weighted least squares (WLS). A useful estimator in the Mx program is the raw data likelihood estimator, which handles data from selected samples and from studies in which part of the sample might have missing data. This last situation commonly arises in longitudinal studies.

SEM can accommodate the analysis of sex differences in heritability estimates through the simultaneous analysis of data from male and female twins. It is also possible to test whether the same genes are expressed in males and females by including DZ twins of opposite sex. If the resemblance between twins of the opposite sex is less than would be expected on the basis of the heritability in males and females, this suggests that different genes influence the same trait in the two sexes. Similarly, heritability that is conditional on environmental exposure can indicate the presence of genotype × environment (G × E) interaction[7]. G × E interaction can be detected by including environmental measurements on the basis of which the twin sample can be stratified. For example, the heritability for depression in married women is lower than in unmarried women. Evidence for the effect of G × E interaction on personality comes from a study in Dutch adolescent twins. A religious upbringing greatly reduces the influence of genetic factors on disinhibition, a trait that closely resembles novelty seeking and that is associated with substance use and abuse[8].

Beyond the classical designs

Extending the MZ–DZ design to include parents, siblings, spouses and offspring of MZ and DZ twins offers the possibility to assess the presence of cultural transmission, G × E covariance, non-random mating and social interactions within and between generations[9]. A simple version of the extended twin design, i.e. a study of young-adult twins, their middle-aged parents and a second group of middle-aged twins (of the same age as the parents of the first group of twin pairs) makes it possible to assess the effect of age differences on heritability and on differential gene expression as a function of age. Using this design as a short-cut for a true longitudinal study, Snieder and colleagues obtained evidence that partially different genes influence lipid levels in plasma at different ages[10]. This may be important information for gene-finding studies, as there might only be a limited time period during which genes, which vary over the course of an individual's life, can be detected. Other extended twin studies look at the offspring of MZ twins who are genetically half-sibs but socially cousins. The MZ-offspring design also allows for testing of maternal effects and imprinting by comparing the offspring of male and female MZ pairs.

The analysis of comorbidity

The causes of association and comorbidity between traits can be investigated by generalizing the univariate twin study to multivariate designs, in which more than one phenotype per person is analyzed. Multivariate twin studies ask questions such as: 'does variation in exercise behavior cause variation in depression, or do the traits cluster because they are influenced by a common set of genes?' Or to give another example: 'do low birth weight and hypertension cluster because one disorder increases risk for the other, or is there a common genetic vulnerability?' The answers to such questions lie in the cross-twin cross-trait correlations in MZ versus DZ twin pairs, i.e. the correlation of birth weight in one twin with blood pressure in the other twin. If these

cross-correlations are higher for MZ than for DZ twins, this suggests that the association between traits is genetically mediated.

Twin registries can be extremely helpful to geneticists by coming up with good phenotypes for molecular genetic studies. In such studies, comorbidity, that is, the co-occurrence of two or more disorders, is often seen as a problem: does one exclude patients with comorbid disorders, or can they be included in the study? A solution to this problem may lie in analysis of the causes of comorbidity. A multivariate twin design can establish the extent to which the comorbid phenotypes share a common genetic basis. The full multivariate analysis of such phenotypes should offer more power to detect genes through genome-wide linkage analysis. Marlow and colleagues conducted multivariate genome-wide analyses of quantitative-trait loci that influence reading- and language-related traits, and found that the results of these analyses were substantially clearer than those of previous univariate analyses[11].

Multivariate analyses are also needed for simultaneous modeling of phenotypes (such as depression) and endophenotypes or intermediate phenotypes (such as neuroticism or cortisol levels) to determine their common genetic etiology. The power to detect linkage will only be increased through the use of endophenotypes if their association is due to pleiotropic genetic effects[12].

Case–control studies

Twins are particularly useful in case–control studies, and MZ twins form the ideal case–control study, as they are perfectly matched for genotype and family background. Martin and co-workers studied vitamin C administration in one twin and a placebo in the co-twin. Contrary to popular belief, vitamin C had no effect on the common cold[13]. As alluded to above, studies of the effects of fetal and infant growth on later health (the so-called Barker hypothesis[14]) in twins investigated the etiology of the association between fetal growth and later disease. In the case–control design, this translates into testing whether, within pairs, the twin with the lowest birth weight has the highest risk of developing the disease. An alternative approach looks at differences in birth weight in MZ and DZ twin pairs and at their association with differences in cardiovascular and metabolic parameters. If these associations are due to shared genetic factors, difference scores are uncorrelated in MZ, but not in DZ twins. For blood pressure, the association between low birth weight and high blood pressure in later life seems to be mediated by common genes[15].

In a more general version of the case–control design, MZ discordant twins, only one of whom has a disease, are used to investigate which non-shared environmental influences are related to the disorder[16]. Such discordant MZ pairs might also form the perfect case–control design for gene-expression studies to distinguish between genes that are related to the causes of disease and genes that are expressed as a consequence of disease. Alternatively, such differential expression, congruent with disease discordance, might indicate causal genes that are differentially activated by epigenetic factors[17]. Conversely, detection of somatic mutations of the same gene in the tumors from MZ twins, both of whom have the tumor, might be a powerful way to detect predisposition genes[18].

DECIPHERING THE HUMAN GENOME

Recent advances in genetics, such as completion of the human genome sequence, increased understanding of DNA sequence variants, the availability of low-cost genome-wide tools to monitor them and the development of powerful statistical tools, have all opened new avenues of investigation in human genetics. There is a danger, however, that the usefulness of these tools is limited when applied to human complex traits. Ascertainment bias, problems with phenotypic assessment, lack of follow-up of the phenotypes over time and environmental noise that can arise, for example, from variation during development might all contribute to the fact that the genes that underlie complex traits and diseases in humans have been difficult to identify. During the past decade considerable effort has been devoted to whole genome screens so that quantitative trait loci (QTLs) for complex traits and diseases can be detected. For complex human traits, there is an increasing recognition of the need to understand their population genetics and biometric properties so that phenotypes can be defined in a manner that maximizes the chances of successful gene mapping. In the following section we discuss how twin studies can be of help in this endeavor.

THE USE OF TWINS IN MOLECULAR GENETIC STUDIES

Basic principles

The concordance between MZ twins sets the upper limit on predictions of individual risk that can be made on the basis of the human genome sequence. Discordant MZ schizophrenic twins, for example, illustrate that disease outcome can be very different for two individuals with identical genetic make-up. The MZ twin concordance thus provides important information regarding disease penetrance. If MZ

twins are genotyped at candidate loci, they provide information about locus-specific penetrances.

Twins also offer specific advantages in genome-wide genotyping, such as linkage or association studies, to map QTLs. DZ twins are siblings of the same age. For traits that change with age, the fact that DZ twins are of the same age decreases within-pair variance and therefore increases power for linkage studies. Twins are also matched for a broad range of pre- and postnatal factors, and are more likely than other siblings to have the same father. The value of combining linkage analyses for Mendelian traits in large pedigrees with twin-based QTL linkage was demonstrated in mapping a cholesterol-lowering gene termed CLG[19]. Its mapping to chromosome 13 was initially based on a single Arab pedigree. The subsequent linkage study in German DZ twins not only confirmed the locus, but also added information on the relevance of the as yet unknown gene by showing its influence on lipid levels in the general population.

Risch and Zhang proposed that stringent selection of extreme discordant and concordant (EDAC) pairs might be the only reliable strategy for QTL mapping in humans[20]. Large registries of twins (see below) contain phenotypes of thousands, sometimes tens of thousands, of twins and their family members, and are a suitable source of informative families for linkage studies. Several such QTL mapping projects are currently under way. For example, genes that influence neuroticism and depression are sought in selected samples of twins and their siblings[21]. To ensure that extremely discordant pairs are not selected because of age differences between them, DZ twins of the same age are used. For ordinary sibs, age differences might create large phenotypic differences between them.

Once a trait has been shown to have a significant heritability, the quest for involved genes and their variants is the next logical step. This includes two distinct approaches, linkage analysis (the relation between a chromosomal segment, termed 'locus', and traits) as well as association analysis (the relation between variants of a gene, termed alleles, and a trait). We briefly explain these two approaches as well as some underlying principles of molecular genetics in the following.

Genotyping

DNA as the substrate for molecular genetics is usually extracted from leukocytes. A vial of 10 ml blood is sufficient for most studies. Buccal cells are additional potential sources of DNA, especially in infants. Some researchers prefer DNA from buccal cells over DNA from blood samples for zygosity typing[22]. The technology for genotyping is now a field of its own, with ever-increasing cycles of innovation. One important distinction that is useful to understand better the principles of molecular genetics is the difference between tests for the length of a DNA fragment (e.g. a 'microsatellite' marker) and tests for mutations or polymorphisms (e.g. single nucleotide polymorphisms or SNPs).

Markers for linkage analysis are highly polymorphic fragments of the DNA of known location in the genome. Microsatellite markers are DNA fragments with a number of repetitions of short sequences (CGAT***CACA…CACA***GTGTT). Most of these markers have repetitions of two base pairs and are termed variable number tandem repeats (VNTR). Markers with three or more repetitive bases allow better discrimination between variants. The genotyping starts with an amplification of the locus harboring the marker by polymerase chain reaction (PCR), a stepwise process that doubles selected parts of the chromosome over and over again. The movement of the fragments in a medium such as a gel or a capillary is dependent on its length. The copied fragments are labeled by specific dyes that can be made visible in a DNA sequencer, leading to the band pattern that almost became a trade-mark of genetics. Specific computer programs transform these patterns into numbers that serve as house numbers and can be traced within families.

SNPs form a very common type of sequence variation. A sequence variation of just a single nucleotide (e.g. CG***A***GAC and CG***T***GAC) may have functional relevance if it changes the resulting protein structure or the regulation of the gene. Even without such consequences, SNPs may be informative when in close proximity to yet undiscovered gene variants. The co-appearance of genetic variants is termed linkage disequilibrium.

Linkage analysis with quantitative traits

Researchers use linkage analysis to try to 'link' DNA markers, whose precise location on one of the 23 pairs of chromosomes is known, to a disease locus or a locus which influences a complex trait. Linkage analysis requires a family-based approach, in which the transmission of genetic markers between generations is tested in relation to the transmission of traits/phenotypes. If a particular DNA marker variant segregates in a pedigree with a disease, then one can assume that the marker locus is close to the disease locus. In linkage analysis of complex traits and continuously distributed phenotypes, the basic concept is identity by descent (IBD). For a marker at a given locus, a child randomly inherits one of the two maternal variants and one of the two paternal alleles. Offspring from the same parents can thus inherit 0, 1 or 2 of the same marker alleles. If similarity for the phenotype corresponds with similarity in marker

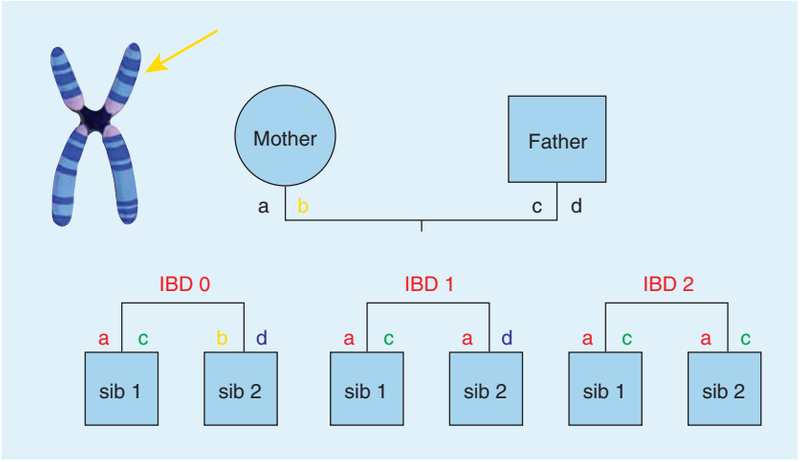

Figure 37.2 Example of a mating between two parents who are both heterozygous for a particular DNA marker. Mother carries alleles 'a' and 'b' and father 'c' and 'd'. Examples of the identity by descent (IBD) status for sibling pairs in the offspring generation are given below the parental generation. In a non-selected population the expectation for the proportion of sibling pairs sharing 0, 1 or 2 alleles IBD is 25%, 50% and 25%, respectively. If the marker is close to a disease locus, the patients will more often be IBD1 or 2

alleles, then again the marker locus is close to the trait locus. When the trait locus influences a quantitative phenotype, it is often referred to as a quantitative trait locus (QTL). The term QTL is also used for complex traits, which have an underlying continuous, genetic liability. In the case of siblings and DZ pairs, three distinct groups can be distinguished. If both sibs inherited the same parental alleles (IBD2), they are identical for alleles at that locus. In contrast, sibs who inherited different parental variants (IBD0) are as different as unrelated subjects for the marker and for genes in that chromosomal area. The remaining group of siblings who share just one of the parental alleles, but not the other allele (IBD1), are as similar for the marker locus as sibs for the total genome (recall that siblings and DZ twins share on average 50% of all segregating genes) (Figure 37.2).

For genetically influenced traits, MZ twins are more alike than DZ twins, and unrelated subjects show the lowest degree of similarity. Likewise, sib pairs who are IBD2 have higher correlations for a particular trait than those who are IBD1 and IBD0, if the marker locus is close to a relevant gene locus. To test the significance of that relation, both regression analysis and variance component analysis methods are widely used and are available in software packages such as Merlin, SOLAR or Genehunter.

MZ twin data do not contribute towards detecting linkage, as MZ twins share all their genetic material identical by descent. However, analyzing MZ phenotypic data simultaneously with linkage data in DZ twins and sibling pairs makes it possible to distinguish between the effects of background genes and shared family environment on the amount of familial variance not accounted for by the QTL.

Association analysis

In association analysis, the interest is usually to look at the influence of allelic variants of a known gene on a particular trait. Linkage analysis compares familial resemblance as a function of IBD status, while association analysis compares the mean levels of the trait (or the absence/presence of disease) in carriers of particular alleles. Linkage thus analyzes variances and covariances, and association compares mean values between groups. This implies that the statistical power in association studies is usually larger than in linkage analysis. Association analysis does not require family structures. In its simplest form, it amounts to counting the numbers of a given allele in patients and healthy controls. Mean differences for a measured quantitative trait can be tested between groups defined by genotype for continuous traits, rather than by comparing genotypes between groups defined by affection status. For a locus with two alleles, A1 and A2, trait levels are compared in subjects with homozygote A1A1, heterozygote A1A2 and homozygote A2A2 genotypes (Figure 37.3).

Even in the absence of functional relevance of a polymorphism, significant differences between genotype groups can and do arise. These can reflect linkage disequilibrium (LD), meaning that the polymorphism is close to a causal variant. Significant results can also be created by population stratification, a term that refers to the fact that between different groups within a population (e.g. groups with a

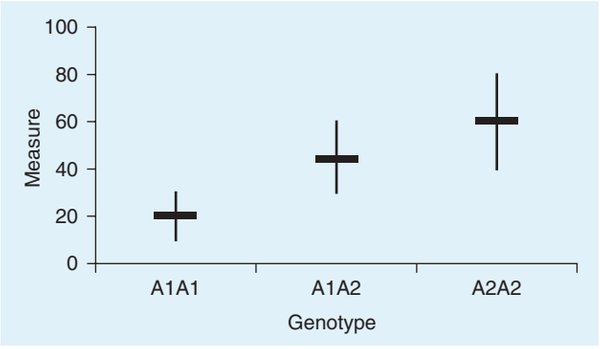

Figure 37.3 Example of association of a quantitative trait ('measure'). Allele A2 has an additive increasing effect as can be seen from comparing subjects without that variant (A1A1) with heterozygous and homozygous carriers of A2

different ethnic background) disease frequencies as well as allele frequencies may vary. The twin design can control for such effects by comparing phenotypes within DZ pairs, taking advantage of the fact that the two twins always belong to the same stratum within the population. If genotypic differences within pairs predict trait differences, this presents strong evidence that the candidate gene is the causal gene (or that it is very close to the causal gene). MZ twins can be used for a special kind of association test. If the differences in trait values within MZ pairs are related to their genotype, this fact constitutes evidence for interaction between genotype and environment, i.e. the genotypes differ in their sensitivity to the environment[23].

TWIN REGISTRIES WORLDWIDE

The twin registers of the world hold an enormous potential for research on the genetics of complex traits. The best of these have existed for decades, and have carefully collected longitudinal data on very large samples of twins and their families. These data include a wide range of important biobehavioral traits and diseases as well as environmental risk factors. Many registries also have large collections of DNA from thousands of subjects. Some registers are population based and, as twins occur pretty much at random in the population, represent some of the best resources for evaluating: first, the importance of genetic variation in liability to disease, through the comparison of similarity in monozygotic and dizygotic twins; second, the significance of genotype by environment (e.g. lifestyle) interaction; and third, the contribution of polymorphisms, in particular typed genes, to the total genetic variance and, by implication, the amount of genetic variance yet to be accounted for. These facts, and the existence of these resources, might not yet be well appreciated in the wider community of genetic and clinical researchers interested in multiple births.

Getting off the ground

Many of the current twin registries are spin-offs from specific research projects, usually in the field of psychology or medicine. The East Flanders Prospective Twin Study, one of the most comprehensive collections of perinatal twin data, was started after the observation of lower degrees of intrauterine hypoxia in second-born twins. The Berlin Twin Register began as a study of the genetics of blood pressure regulation associated with mental stress. Once scientists gain experience with twin research and the databases that support it, they realize the potential of such studies beyond the original question, as well as the necessity to 'nurse' this resource. Resources permitting, a twin collection develops into a register. To define a register, of either twins or patients with a given disease or any other distinguishing characteristic, issues such as sample size and range and sophistication of the data need to be taken into account. For epidemiologic research, existing or emerging population-based registers, including those in Northern Europe, Italy, Korea or Sri Lanka, are of extreme value because of their lack of ascertainment bias. In some countries, sampling of twin pairs is based on computerized population registers, either using direct information on multiple births or applying complex filters that include sharing of date of birth, family name at birth or place of birth or partly sharing identification numbers. Many researchers now realize the value of large data collections for molecular genetic studies and have begun to include siblings, parents and offspring of twins in their research projects.

Data collection

Administrators of most registries maintain contact with the twins and their family members through websites, by sending out newsletters and through mailed surveys. Even DNA samples from buccal swabs have been collected by post[24]. Data collection by mailed questionnaires often results in large data sets such as that from over 45 000 twins and their relatives on neuroticism, a strong risk factor for the development of depression. This effort concluded that familial resemblance for this trait has a simple genetic basis, and rejected alternative models for familial resemblance, such as cultural transmission[25].

Although sample size is clearly important in genetic epidemiology, together with the amount of data that can be collected through questionnaires and large-scale survey studies, benefits also derive from the scope or depth of phenotypes that need to be collected in laboratory settings. To study genetic basis of disease, intermediate phenotypes often must

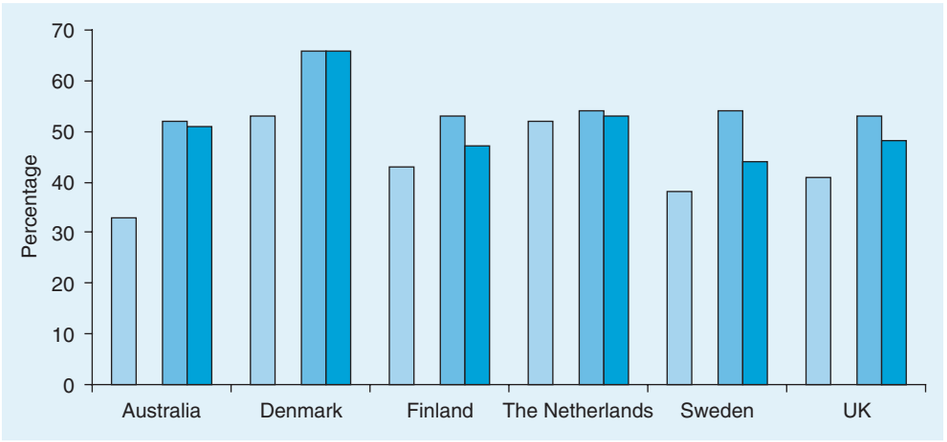

Figure 37.4 Heritability for migraine (light blue), systolic (medium blue) and diastolic (dark blue) blood pressure from six twin registries participating in GenomEUtwin[26]. Articles at: www.ists.qimr.edu.au/journal. html

be determined in more extensive and costly studies. The selection of subsamples can either be random or selective on prior phenotyping from the larger registers (as for example in the Australian or Scandinavian registers) or smaller twin cohorts collected for specific studies. The register of Italian twin athletes and the study of growth before birth and later adult health are good examples of successful in-depth studies on a small scale.

Centralized health databases also greatly facilitate efficient data collection. For example, Finland and other Scandinavian countries have centralized registers for hospital discharge data and for fully reimbursable medications. This information is accessible with a unique personal identifier that is given to each individual at birth and allows record linking for all twins without any self-selection as potential bias. Personal identifiers of twins are used as a query filter for the health databases, and matching records can then be transferred into the twin database.

CONCLUSION AND PROSPECTS

Twin studies have demonstrated a significant genetic contribution to population variation for multifactorial traits such as body height and weight, neuroticism or blood lipid levels and complex diseases such as obesity, depression or cardiovascular disease. Many of these traits and diseases are currently on the increase worldwide, and are influenced by risk factors that include diet, smoking and lack of exercise. Although such 'lifestyle' risk factors that are important for the development of complex diseases are often considered 'environmental', they might themselves be influenced by genes. Twin studies have been useful in assessing the extent to which variation in lifestyle and healthy behavior might be heritable. In fact, recent twin studies have provided considerable evidence that 'lifestyle' risk factors aggregate in families owing to shared genes, in addition to the shared environment. For example, twin studies suggest that differences in eating patterns, smoking initiation and persistence, sports participation and even religious beliefs are all influenced by genetic variation. Heritability for a particular disease thus may reflect the direct influence of disease genes, the influence of genes responsible for variation in lifestyle factors, or the influence of genes which modify the influence of lifestyle on disease risk.

The value of large, unbiased study samples needed to verify the role of genetic variation that underlies common traits is well recognized. We have now moved to an era in which genotyping is relatively cheap and rapid, and the major cost of a study of a complex trait involves family collection and trait phenotyping. This view is reflected by a recent decision by the European Community to fund a large integrated project called GenomEUtwin. The participating twin cohorts, from Scandinavia, The Netherlands, the UK, Italy and Australia, form a collection of more than 0.6 million pairs of twins. Over 30 000 DNA samples, accompanied by informed consent for genetic studies of common diseases, have been collected from these population-based twin cohorts. Combining the data from the major twin registries of Europe will integrate the efforts of the leading genetic and epidemiologic researchers in the field of twin research. In this project, epidemiologic and phenotypic data collection will be integrated. Initial 'proof of principle' genome-wide genotyping efforts will be targeted to 10 000 twins for stature, body mass index (BMI), blood pressure and migraine. These traits all show significant heritability

(Figure 37.4) which, surprisingly, differs very little across countries.

Twins and their family members are often enthusiastic participants in research studies. The increase in the twinning rate in The Netherlands and other countries ensures viability of the application of the classical twin design in genetic epidemiology, in medical, behavioral and psychiatric genetics. Only the close collaboration of geneticists, gynecologists and researchers can ensure optimal use of the twin design.

ACKNOWLEDGMENT

The GenomEUtwin project is supported by the European Union, Contract No. QLG2-CT-2002-01254.

REFERENCES

1. Augustine of Hippo. *De Civitate Dei*. AD 415
2. Galton F. The history of twins as a criterion of the relative powers of nature and nurture. *J R Anthropol Inst G Br Irel* 1875;5:391
3. Siemens HW. *Die Zwillingspathologie: Ihre Bedeutung, ihre Methodik, ihre bisherigen Ergebnisse* [Twin pathology: its importance, its methodology, its previous results]. Berlin: Verlag von Julius Springer, 1924
4. Kendler KS , Neale MC, Kessler RC, *et al*. A population-based twin study of major depression in women. The impact of varying definitions of illness. *Arch Gen Psychiatry* 1992;49: 257–66
5. Boomsma DI, Koopmans JR, Doornen LJP van, Orlebeke JF. Genetic and social influences on starting to smoke: a study of Dutch adolescent twins and their parents. *Addiction* 1994; 89:219–26
6. Neale MC, Cardon LR. *Methodology for Genetic Studies of Twins and Families*, NATO ASI Series D: Behavioural and Social Sciences – Vol. 67. Dordrecht: Kluwer Academic Publishers, 1992
7. Eaves LJ, Last KA, Martin NG, Jinks JL. A progressive approach to non-additivity and genotype–environmental covariance in the analysis of human differences. *Br J Math Stat Psychol* 1997;30:1
8. Boomsma DI, de Geus EJ, van Baal GC, Koopmans JR. A religious upbringing reduces the influence of genetic factors on disinhibition: evidence for interaction between genotype and environment on personality. *Twin Res* 1999;2:115–25
9. Truett KR, Eaves LJ, Walters EE, *et al*. A model system for analysis of family resemblance in extended kinships of twins. *Behav Genet* 1994;24:35–49
10. Snieder H, van Doornen LJP, Boomsma DI. Age-dependency of gene expression for plasma lipids, lipoproteins and apolipoproteins. *Am J Hum Genet* 1997;60:638–50
11. Marlow AJ, Fisher SE, Francks C, *et al*. Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet* 2003;72:561–70
12. Martin N, Boomsma DI, Machin G. A twin-pronged attack on complex traits. *Nat Genet* 1997;17:387–92
13. Martin NG, Carr AB, Oakeshott JG, Clark P. Co-twin control studies: vitamin C and the common cold. *Prog Clin Biol Res* 1982;103:365–73
14. Hales CN, Barker DJ. The thrifty phenotype hypothesis. *Br Med Bull* 2001;60:5–20
15. Ijzerman RG, Stehouwer CD, Boomsma DI. Evidence for genetic factors explaining the birth weight–blood pressure relation. Analysis in twins. *Hypertension* 2000;36:1008–12
16. Asbury K, Dunn JF, Pike A, Plomin R. Nonshared environmental influences on individual differences in early behavioral development: a monozygotic twin differences study. *Child Dev* 2003;74:933–43
17. Petronis A, Gottesman II, Kan P, *et al*. Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr Bull* 2003;29:169–78
18. Forsti A, Jin Q, Sundqvist L, *et al*. Use of monozygotic twins in search for breast cancer susceptibility loci. *Twin Res* 2001;4: 251–9
19. Knoblauch H, Muller-Myhsok B, Busjahn A, *et al*. A cholesterol-lowering gene maps to chromosome 13q. *Am J Hum Genet* 2000;66:157–66
20. Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 1995;268:1584–9
21. Boomsma DI, Beem AL, van den Berg M, *et al*. Netherlands twin family study of anxious depression (NETSAD). *Twin Res* 2000;3:323–34
22. Hall JG. Twinning. *Lancet* 2003;362:735–43
23. Berg K, Kondo I, Drayna D, Lawn R. 'Variability gene' effect of cholesterylester transfer protein (CETP) genes. *Clin Genet* 1989;35:437–45
24. Meulenbelt I, Droog S, Trommelen GJ, *et al*. High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *Am J Hum Genet* 1995;57:1252–4
25. Lake RI, Eaves LJ, Maes HH, *et al*. Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45 850 twins and relatives on two continents. *Behav Genet* 2000;30:223–33
26. Special issue on GenomEUtwin. *Twin Res* 2003;6:issue 5

APPENDIX

Twin studies	*Number of pairs*	*Primary interest*	*Origin of twins*
Australian Twin ADHD Project (ATAP) http://psych.curtin.edu.au/people/hayd.htm	1959	attention deficit hyperactivity disorder and childhood behavioral disorders	Australia
Australian Twin Registry[1] www.twins.org.au	27 582	general resource for medical and scientific research	Australia
Western Australian Twin Register www.ichr.uwa.edu.au	4729	asthma and allergy, attention deficit hyperactivity disorder, early speech and behavior	Australia
East Flanders Prospective Twin Survey (EFPTS)[2] c.derom@pi.de	6050	epidemiology, placentation (chorion type), congenital anomalies, perinatal factors, Barker hypothesis	Belgium
University of British Columbia Twin Project[3] http://www.psychiatry.ubc.ca/	816	personality and the personality disorder	Canada
Chinese National Twin Program (CNTP) cchp@public3.bta.net.cn	4576	establish a population-based national twin registry and study etiologies of common diseases and health-related behavior	China
The Danish Twin Registry[4] askytthe@health.sdu.dk	65 000	aging and age-related health, metabolic and cardiovascular disease, specific diseases	Denmark
The Finnish twin Cohort[5] jaakko.kaprio@helsinki.fi	15 000	health, personality, substance abuse	Finland
Berlin Twin Register[6] (HealthTwiSt) www.healthtwist.de	>900	complex diseases, health-related quantitative traits, pharmacogenetics	Germany
German Observational Study of Adult Twins (GOSAT) and the Bielefeld Longitudinal Study of Adult Twins (BiLSAT)[7] angleitner@uni-bielefeld.de	2509	longitudinal assessment of temperament and personality; generalizability of behavioral genetic findings across methods of personality assessment	Germany
Italian Twin Registry www.gemelli.iss.it	120 000	aging, dementia, cardiovascular diseases, multiple sclerosis, celiac disease, diabetes, asthma, allergies, thyroid diseases, behavioral disorders	Italy
Register of Italian Twin Athletes (RITA) casini@iusm.it	4719	human biology and development, sport and high-level performance	Italy
Twin Register of Rome (TERRY) casini@iusm.it	13 228	lifestyle, development, aging	Italy
Osaka University Aged Twin Registry hayakawa@sahs.med.osaka-u.ac.jp	12 000	aging, dementia, physical diseases, lipids, cognition, lifestyle, life satisfaction, quality of life	Japan
Korean Twin Registry sungjohn@kangwon.ac.kr	154 783	complex human diseases and traits	Korea (South)
Seoul Twin Family Study www.ktrc.org	>4615	cognitive abilities	Korea (South)
Norway Twin Registries[8] mina.bergem@psykiatri.uio.no	>40 000	mental health, obesity, asthma, allergies, health behaviors and perceptions, perinatal influences on health outcomes	Norway

(*Continued*)

Continued

Twin studies	*Number of pairs*	*Primary interest*	*Origin of twins*
The NIPH Twin Panel[9] jennifer.harris@folkehelsa.no	7668	physical health, mental health, asthma, allergies, obesity, health-related behaviors	Norway
National Twin Registry of Sri Lanka www.infolanka.com/org/twin-registry/	20 294	establish nationwide population-based twin register for multidisciplinary research and international collaborations	Sri Lanka
The Swedish Twin Registry[10] www.mep.ki.se	57 405	cancer, cardiovascular diseases, dementia, depression, substance use/abuse, cognition, personality, aging, (common complex diseases)	Sweden
The Swedish Young Male Twins Study Finn.rasmussen@imm.ki.se	1783	risk factors for metabolic syndrome and cardiovascular diseases; overweight and behavioral risk factors	Sweden
Netherlands Twin Register (NTR)[11] www.psy.vu.nl/ntr	30 335	development behavior and emotional problems. Cognition, depression, addiction and cardiovascular risk factors	The Netherlands
St Thomas' UK Adult Twin Registry[12] www.twin-research.ac.uk	10 000	cardiovascular, metabolic, musculoskeletal, dermatological and ophthalmological diseases	UK
Study of growth before birth and adult health g.mcneill@abdn.ac.uk	123	risk factors for coronary heart disease	UK
Twins' Early Development Study (TEDS) a.trouton@iop.kcl.ac.uk	16 810	longitudinal assessment of verbal and non-verbal cognitive development and delay, language development and delay, childhood behavior problems	UK
Northern Region Multiple Pregnancy Register Christopher.Wright@ncl.ac.uk	1216	multiple pregnancy, obstetric and pediatric management and outcomes of pregnancy	UK (North East)
Mid-Atlantic Twin Registry www.matr.vcu.edu	23 000	behavioral and psychiatric	USA
National Academy of Sciences–National Research Council (NAS–NRC) Twin Registry www.iom.edu/twins	15 924	somatic and psychiatric disease, aging, social, psychological and demographic variables	USA
Southern Illinois Twins www.siumed.edu/playlab	126	peer interaction behaviors, preschool cognitive development	USA
Vietnam Era Twin (VET) Registry Birute.Curran@med.va.gov	7500	veterans health, effects of combat, psychiatric disorders, substance abuse	USA
International Twin Study tmack@usc.edu	17 229	etiology of disease, genetic markers	USA and Canada
California Twin Program http://twins.usc.edu/	13 096	etiology of disease, genetic markers	USA (California)

Continued

Twin studies	*Number of pairs*	*Primary interest*	*Origin of twins*
San Diego twin blood pressure study at UCSD http://elcapitan.ucsd.edu/hyper	200	blood pressure, autonomic 'intermediate phenotypes' for high blood pressure	USA (California)
Southern California Twin Register www-rcf.usc.edu/~lbaker	2600	social and moral development, childhood behavior problems, cognitive abilities	USA (California)
Georgia Cardiovascular Twin Study www.mcg.edu/institutes/gpi	534	longitudinal development of bio-behavioral antecedents of cardiovascular disease in youth	USA (Georgia)
Minnesota Twin Family Study (MTFS)[13] www.tc.umn.edu/~mctfr	4723	substance use, related child and adult disorders	USA (Minnesota)
Minnesota Twin Registry www.psych.umn.edu/psylabs/mtfs/	5599	individual differences	USA (Minnesota)
NY Obesity Research Center Child Twin Registry (Project "Grow-2-Gether") http://cpmcnet.columbia.edu/dept/obesectr/NYORC/twins.html	50	food intake, body composition	USA (New York)
Wisconsin Twin Panel www.waisman.wisc.edu/mrddrc	recruitment began with 1989 births	childhood behavioral disorders	USA (Wisconsin)

REFERENCES FOR APPENDIX

1. Hopper JL, Green RM, Nowson CA, *et al*. Genetic, common environment and individual specific components of variance for age- and lean mass-adjusted bone mineral density in 10- to 26-year-old females: a twin study. *Am J Epidemiol* 1998; 147:17–29
2. Derom C, Vlietinck, R, Derom R, Van den Berghe H. Increased monozygotic twinning rate after ovulation induction. *Lancet* 1987;1:1236–8
3. Jang KL, Hu S, Livesley WJ, *et al*. The covariance structure of neuroticism and agreeableness: a twin and molecular genetic analysis of the role of the serotonin transporter gene. *J Personality Soc Psychol* 2001;81:295–304
4. Kjeldsen, MJ, Kyvik KO, Christensen K, Friis ML. Genetic and environmental factors in epilepsy: a population-based study of 11,900 Danish twin pairs. *Epilepsy Res* 2001;44:167–78
5. Cannon TD, Huttunen MO, Lönnqvist J, *et al*. The inheritance of neuropsychological dysfunction in twins discordant for schizophrenia. *Am J Hum Genet* 2000;67:369–82
6. Busjahn A, Knoblauch H, Faulhaber HD, *et al*. QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation* 1999;99:3161–4
7. Borkenau P, Riemann R, Angleitner A, Spinath FM. Genetic and environmental influences on observed personality: evidence from the German Observational Study of Adult Twins. *J Personality Soc Psychol* 2001;80:655–68
8. Bergem ALM, Lannfelt L. Apolipoprotein E type e4 allele, heritability and age at onset in twins with Alzheimer disease and vascular dementia. *Clin Genet* 1997;52:408–13
9. Harris JR, Tambs K, Magnus P. Sex-specific effects for body mass index in the new Norwegian twin panel. *Genet Epidemiol* 1995;12:251–65
10. Lichtenstein P, De Faire U, Floderus B, *et al*. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med* 2002;252:184–205
11. Boomsma DI, de Geus EJC, van Baal GCM, Koopmans JR. Religious upbringing reduces the influence of genetic factors on disinhibition: evidence for interaction between genotype and environment. *Twin Res* 1999;2:115–25
12. Spector TD, Keen RW, Arden NK, *et al*. Influence of vitamin D receptor genotype on bone density in postmenopausal women: a British twin study. *Br Med J* 1995;310:1357–60
13. McGue M, Iacono WG, Legrand L, *et al*. Origins and consequences of age at first drink. I. Associations with substance-use disorders, disinhibitory behavior and psychopathology, and P3 amplitude. *Alcoholism Clin Exp Res* 2001; 25:1156–65