STUDY DESIGNS

The continuing value of twin studies in the omics era

Jenny van Dongen[1], P. Eline Slagboom[2], Harmen H. M. Draisma[1], Nicholas G. Martin[3] and Dorret I. Boomsma[1]

Abstract | The classical twin study has been a powerful heuristic in biomedical, psychiatric and behavioural research for decades. Twin registries worldwide have collected biological material and longitudinal phenotypic data on tens of thousands of twins, providing a valuable resource for studying complex phenotypes and their underlying biology. In this Review, we consider the continuing value of twin studies in the current era of molecular genetic studies. We conclude that classical twin methods combined with novel technologies represent a powerful approach towards identifying and understanding the molecular pathways that underlie complex traits.

Classical twin design
The approach used to estimate the importance of genetic and environmental influences on complex trait variation. The estimate of heritability is based on a comparison of resemblence in monozygotic twins (who share all segregating genetic material) and dizygotic twins (who share, on average, half of their segregating genetic material).

[1]Biological Psychology, VU University, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands.
[2]Molecular Epidemiology, Leiden University Medical Center, Netherlands Consortium for Healthy Ageing, Einthovenweg 20, 2333 ZC Leiden, The Netherlands.
[3]Genetic Epidemiology Unit, Queensland Institute of Medical Research, 300 Herston Road, Herston, Brisbane, Queensland 4006, Australia.
Correspondence to D.I.B.
e-mail: di.boomsma@vu.nl
doi:10.1038/nrg3243
Published online 31 July 2012

The classical twin design has been used for decades to estimate the importance of genetic and environmental influences on complex trait variation. Its results have contributed to the awareness that variation in almost every conceivable facet of the human condition is influenced by genetic variation (BOX 1). Traits include intrinsic physical, medical and biochemical characteristics, life-outcome variables, such as income, divorce and mortality, and behavioural traits, including apparently trivial ones such as television watching and Internet use. In fact, for many human phenotypes, heritability estimates derived from twin studies initially encouraged the search for the responsible genetic variation. Through their collaboration in genome-wide association study (GWAS) consortia, large twin registries (TABLE 1; Supplementary information S1 (table)) are nowadays also making an important contribution towards identifying the genetic variation that underlies complex traits and disorders.

Twins offer unique opportunities to genetic research that extend beyond the analysis of phenotypic heritability (BOX 2). Twin designs can provide insight into the genetic aetiology of disease development over time and can aid in the detection of biomarker profiles for medical conditions. For heritable traits, the comparison of discordant monozygotic twins (discordant MZ twins) represents a powerful improvement over the traditional case–control study to search for disease-associated biological marks. The power of this design is demonstrated in a recent study that compared the DNA methylation patterns of MZ twins who were discordant for systemic lupus erythematosus (SLE), and this study identified several genomic regions in which DNA methylation changes were associated with the disease[1]. Novel applications of the classical twin design can provide fundamental insights into the biological mechanisms underlying complex traits. For example, gene expression studies in MZ and dizygotic (DZ) twins have highlighted that variation in genome-wide expression between individuals is due to both genetic and environmental influences and that the importance of these influences may vary across genes and tissues[2,3].

This Review addresses the continuing value of twin studies. We describe various twin study designs with examples of traditional applications, and we describe how twin approaches are now used for tracing disease-causing mutations and for studying various other newly emerging phenotypes (for example, the epigenome, transcriptome, metabolome, proteome and microbiome). We address using discordant MZ twins for the identification of biological mechanisms that are associated with complex traits, for the inference of causality and for the genome-wide analysis of genotype-by-environment (G×E) interaction at variability genes. We also discuss various questions that can be addressed by contrasting data from MZ and DZ twins to establish the heritability of biological marks and to unravel the shared aetiology of associated traits. A range of twin studies is presented, focusing on the initial level of the DNA sequence, down to its expression and intermediate phenotypes, such as metabolites, and ultimately to the clinical endpoints of interest.

Box 1 | **The history of the classical twin study**

The scientific study of twins goes back to 1875, when Francis Galton[104] published his seminal paper 'The history of twins, as a criterion of the relative powers of nature and nurture'. However, Galton was unaware of the distinction between monozygotic (MZ) and dizygotic (DZ) twins. The first studies that investigated the different levels of similarity between MZ and DZ twins were published by Poll (1914)[105] and Siemens (1924)[106], whose interest was pigmented nevi (common moles), a phenotype that is still being studied intensively today because of its importance as a risk factor for melanoma[107]. Not much later, the first twin registries were founded, and power calculations showing that very large sample sizes were needed to obtain reliable estimates of heritability stimulated the foundation of new large registries in the 1980s[108,109]. Consolidation of these registries, new methods for zygosity assessment and improved survey methods coincided with a growing awareness that genetic influences affected a wide range of traits of biomedical and social importance, and an increase in funding to mount large studies. Worldwide, many countries have now set up twin registries[110–112], which have established collections of longitudinal data in twins across age categories from birth[113] to death[34]. Within the past 20 years, very large twin studies have been carried out through mailed, telephone and Internet surveys. Methods linking twin registry data to national databases containing information on cancer and mortality[114], or outcomes of population screens[115], have provided population-based estimates of heritability on samples as large as 44,000 twin pairs.

The continuing importance of twin study designs

Quantitative analysis of genetic and environmental influences. The classical twin design has traditionally been used to study the heritability of disease-related phenotypes and clinical endpoints (TABLE 2). This design has also been widely applied to estimating the extent to which different traits are influenced by the same or different genetic and environmental factors[4]. Multivariate twin models of, for example, symptoms of anxiety and depression provided evidence that comorbidity of these disorders is due to genetic influences that affect the vulnerability to both disorders but that different environments determine whether a vulnerable person develops major depression or generalized anxiety disorder[5,6]. Longitudinal data can be analysed in a similar way: genetic variation in intelligence quotient from ages 1 to 16 is largely attributable to the same genetic influences[7], and the increase in heritability[8] is due to amplification of genetic effects with age. The classical twin design can be extended to model polygenic G×E interactions by testing whether the heritability of a trait varies across different levels of environmental exposures[9]. The heritability of body mass index (BMI) is moderated by physical activity: the higher the level of physical activity, the smaller the genetic influence on BMI[10].

Extending twin models with data from other relatives (such as their parents, siblings, spouses or offspring) enhances statistical power[11] and allows for testing of a much wider range of hypotheses about the causes of human variation, including the role of cultural transmission, social interactions among relatives[12], genetic non-additivity and various mechanisms of assortative mating[13,14]. The 'offspring-of-twins' design is a powerful tool for studying intergenerational associations between environmental variables and outcomes in children[15]. Also, comparing the phenotypic similarity of children of female MZ twins (who socially are cousins but genetically are half-siblings) to the similarity of children of male MZ twins gives insight into the differential importance of paternal and maternal effects; if paternal and maternal effects are equally important, children of male twins and female twins are expected to be equally similar. For birth weight, larger correlations have been observed in children of female twins compared with children of male twins, highlighting the importance of maternal effects for this trait[16].

Classical twin methods continue to be a valuable addition to genetic association studies to establish, for example, the proportion of the heritability that can be explained by newly identified SNPs from GWASs[17]. The current discussion about 'missing heritability' largely stems from the (often great) disparity between estimates of total heritability from twin studies and the proportion of variance accounted for by SNPs from GWASs[18–20], for which many explanations have been proposed[21], including implications that heritability estimates from twin studies may be too high. In our later section on testing classical assumptions, we discuss the relevance of recent molecular findings in twins in the light of the current discussion on 'missing heritability'.

The value of discordant twins. Data from MZ and DZ twins allow for the examination of causal relations in the comorbidity of traits. In this case, information from discordant twins is used in a design that is referred to as the co-twin control method. This method was first used to study the association between smoking and lung cancer[22] and has since been applied to investigate a wide variety of medical hypotheses: for example, to provide evidence against the efficacy of vitamin C in preventing the common cold[23]. The value of the co-twin control design for distinguishing between associations that reflect causality and associations owing to confounding effects of genes or environmental factors (that is, if two traits are affected by the same genetic or environmental influences rather than one trait causing the other) is further exemplified by several recent studies on complex traits, as described below.

Experimental studies in which depressive patients are exposed to various types of exercise regimes suggest that regular exercise causes a reduction in anxious and depressive symptoms. To examine whether this causal relationship is present at the population level, twins who are discordant for exercise behaviour were studied[24]. MZ twins who exercised more than their co-twin did not have fewer symptoms of anxiety and depression. The relationship between exercise behaviour and depression was explained by shared genetic influences rather than by a cause–effect relationship. In another twin study, a reciprocal causal relationship between depression and migraine was revealed[25]. In MZ pairs who were discordant for depression, only the depressed twin had an increased risk of migraine, and in MZ pairs who were discordant for migraine, only the twin with migraine had an increased risk of depression. Furthermore, a co-twin control study of anthropometric traits and cancer found a positive correlation between height and risk of breast and ovarian cancers and indicated correlations between BMI and several types of cancer in some population subgroups[26].

Heritability
The proportion of variation in a trait that is due to heritable differences between individuals in a population: that is, the proportion of variation due to additive genetic effects (that is, narrow-sense heritability) or the proportion of variation due to all genetic effects (that is, broad-sense heritability).

Discordant monozygotic twins
(Discordant MZ twins). Twins who derive from a single fertilized egg cell but who are dissimilar for a certain characteristic or disease. By contrast, concordant MZ twins are phenotypically similar.

Case–control study
The comparison of individuals with a trait or disease of interest (cases) to controls to identify genes or other aspects associated with the trait. Cases and controls can be unrelated or can be relatives (within-family case–control design).

Epigenome
The entire collection of epigenetic marks, including DNA methylation and histone modifications, that regulate the expression of the genome. In contrast to the genome, the epigenome is specific to each cell.

Table 1 | **A selection of twin registries worldwide**

Twin registry name	Registry characteristics	Age	Website	Number of twins or subjects (approximate)*	Number of twins or subjects with DNA available (approximate)*	Biospecimens (available for at least subset of the sample)
Africa						
Bandim Health Project twin registry (Guinea-Bissau)	Population-based with ongoing longitudinal data collection	0–30	http://www.bandim.org	2,500 (twins and singleton controls)	200 twin pairs	Whole blood, plasma
Asia and Australia						
Australian Twin Registry	Population-based with ongoing longitudinal data collection	0–90	http://www.twins.org.au	66,000	12,000 (twins and other family members)	Serum, plasma, buccal cells
Chinese National Twin Registry (CNTR)	Population-based with ongoing longitudinal data collection	All	http://cntr.bjmu.edu.cn	35,000 twin pairs	3,200	Serum, DNA
South Korean Twin Registry (SKTR)	Volunteer preschoolers, cohort of school children, volunteer young adults	0–30	http://www.ktrc.org	10,000 twin pairs	800 twin pairs	Hair, saliva
Keio Twin Registry (Japan)	Adult and adolescent twins from the general population in the Tokyo area	14–30	http://totcop.keio.ac.jp; http://kts.keio.ac.jp; http://kotrec.keio.ac.jp	4,000 twin pairs (plus other family members)	600 twin pairs	Buccal cells, blood
Sri Lankan Twin Registry	Voluntary twin registry component and a population-based database with ongoing data collection	6–94	http://www.ird.lk/Twin%20Registry.php	35,000	Plans to collect DNA from 4,000	Buccal cells
Europe						
The Danish Twin Registry (DTR)	Population-based with ongoing longitudinal data collection	0–107	http://www.sdu.dk/dtr	170,000	20,000	Serum, plasma, buffy coat, saliva, buccal cells, urine
Finnish Twin Cohort study	Population-based with ongoing longitudinal data collection	11–100+	http://www.twinstudy.helsinki.fi	45,000 (plus family members)	14,600 (twins and family members)	Whole blood, serum, plasma, saliva, urine, fat and muscle by biopsy
Netherlands Twin Register (NTR)	Population-based with ongoing longitudinal data collection	0–100	http://www.tweelingenregister.org/en	87,500 (plus family members)	18,000	DNA, RNA, cell lines, serum, plasma, buccal cells, urine, stool
Norwegian Twin Registry (NTR)	Population-based with ongoing longitudinal data collection	18+	www.fhi.no/twins	40,000	4,800	Whole blood, buccal cells, plasma
Swedish Twin Register (STR)	Population-based with ongoing longitudinal data collection	5–100+	http://ki.se/ki/jsp/polopoly.jsp;jsessionid=acR0ziTHzWEclO_cNC?l=en&d=9610	194,000	44,600	Whole blood, serum, saliva
TwinsUK registry	Population-based with ongoing longitudinal data collection	18–90	http://www.twinsuk.ac.uk	12,000	7,000	Whole blood, serum, plasma, buffy coat, saliva, buccal cells, urine, skin, fat, muscle

The comparison of discordant MZ twins offers an alternative to the traditional case–control study. Here, the primary interest is not to infer causality but to identify factors associated with a trait of interest that differ between cases and controls who are perfectly matched for age, sex and genetic background, and who are partly matched for early environmental influences.

Molecular phenotypes and the causes of quantitative trait variation. Technological advances allow an assessment of the extent to which twins resemble each other at the level of molecular processes that contribute to their phenotypic similarity[27]. Thereby, the comparison of discordant MZ twins can lead us into novel pathways associated with disease. A unique advantage of the MZ twin design is the ability to study biological discordance

Transcriptome
The total set of RNA transcripts that are produced in a cell or tissue by transcription of DNA.

Table 1 (cont.) | **A selection of twin registries worldwide**

Twin registry name	Registry characteristics	Age	Website	Number of twins or subjects (approximate)*	Number of twins or subjects with DNA available (approximate)*	Biospecimens (available for at least subset of the sample)
North America						
Mid-Atlantic Twin Registry (MATR; United States)	Population-based, ascertained at birth	0–94	http://www.matr.vcu.edu	56,000	1,500	Whole blood, serum, plasma, buffy coat, saliva, buccal cells
NAS–NRC (National Academy of Sciences–National Research Council) twin registry of Second World War male veterans (United States)	Male twins born between 1917–1927, both of whom served in the military, mostly during the Second World War	85–95	http://iom.edu/Activities/Veterans/TwinsStudy.aspx	31,848	700+	Blood and other materials collected for various investigations
Minnesota Twin Family Study (MTFS; United States)	Ongoing population-based longitudinal study	11–47	http://mctfr.psych.umn.edu	5,000 (plus family)	10,000 (twins and family members)	Blood-derived or saliva-derived DNA
Wisconsin Twin Panel (WTP; United States)	Population-based, longitudinal data, extensive phenotypic characterization, follow-up of selected samples	0–23	http://www.waisman.wisc.edu/twinresearch	19,638 twins (plus parents and siblings)	3,489 (twins, parents and siblings)	Saliva, buccal cells
South America						
Cuban Twin Registry	Population-based with ongoing longitudinal data collection	All		55,400 twin pairs	250 twin pairs	Blood-derived DNA

*Numbers refer to individual twins (rather than twin pairs) unless indicated otherwise. This table shows a selection of some of the large twin registries worldwide. For a more comprehensive table, see Supplementary information S1 (table).

against an equivalent genetic background. Divergence of epigenetic profiles in MZ twins depends on the locus and has been documented for both younger and older age groups[28–30]. In fact, differences in DNA methylation and gene expression are already evident in newborn MZ twins[31,32]. Clearly, environmental and stochastic factors start *in utero* and operate throughout life.

In addition to traditional organismal quantitative traits (such as height and BMI), molecular characteristics (such as gene expression levels, the methylation state of CpG sites in the DNA and the concentration of metabolites in blood and urine) may be regarded as quantitative traits. Variation in molecular traits measured in groups of MZ and DZ twins can be analysed using the classical twin method, like any other phenotype. Multivariate twin analyses address questions that are not easily resolved in any other study design, such as to what extent is the epigenetic regulation and expression of genes across genomic regions influenced by shared genetic factors and to what extent is each region influenced by unique factors? And to what degree do common genetic and environmental mechanisms underlie biological variation across different cells and tissues[33]? The availability of genome-wide DNA marker data allows for novel approaches towards studying G×E interactions, in which MZ twins can play a vital part. By studying variation in a phenotypic trait of interest in MZ twins, it is possible to see not only whether some genotypes confer higher levels of risk for that trait but also whether some contribute to its variability; high variability in the expression of a trait from a common genetic background could explain phenotypic differences between MZ co-twins. Of interest, genetic and environmental factors may influence disease through different pathways (BOX 3). Twin studies can be used to identify aspects of disease that are most related to the underlying genetic liability of individuals and can thereby help to establish clinical criteria and phenotypic definitions that will facilitate the success of GWASs. Other approaches, such as the offspring-of-twins design, may provide insight into transgenerational inheritance of epigenetic regulation and the importance of maternal effects and imprinting on epigenetic marks, although such studies have not yet been published.

An important strength of twin registries lies in the extensive longitudinal collection of data on various phenotypes. Twin studies have indicated that approximately 20–30% of the overall variation in adult lifespan is accounted for by genetic factors[34]. Longitudinal twin studies can be used to identify biomarkers that are associated with ageing: a co-twin control analysis showed that telomere length at advanced age is predictive of survival[35]. MZ twins with the shortest telomeres at the baseline had a threefold greater risk of death during a follow-up period of 7 years than their co-twins with the longest telomere measurements (relative risk = 2.8). The discordant MZ twin design and the classical twin design have received much interest in recent years for studying molecular biology. The following sections will provide an overview of findings from such studies.

Metabolome
The total set of small molecules (for example, lipids, amino acids and sugars) that are the reactants, intermediates or end products of cellular metabolism and that are present in a cell, tissue or complete organism.

Proteome
The entire complement of proteins that are present in a cell, tissue or complete organism.

Microbiome
The entire set of genomes of microorganisms (for example, bacteria, fungi and viruses) that are present in a certain environment: for example, in the human gut.

Variability genes
Genes that contribute to the variation in a phenotype. The genotypes are associated with phenotypic variance rather than with the mean level or frequency of the trait.

Zygosity assessment
The assessment whether same-sex twins are monozygotic or dizygotic is often based on the comparison of DNA markers or alternatively on standardized questionnaires.

Multivariate twin models
Models used for the simultaneous analysis of multiple traits measured in monozygotic and dizygotic twins to estimate the importance of genetic and environmental influences shared ('overlapping') between traits in explaining their clustering, comorbidity or covariance.

Tracing the origin of new mutations

Identifying sequence differences between twins. Although MZ twins originate from one zygote, there is some evidence that their somatic cells are not always identical at the DNA sequence level[36]. A study of healthy MZ twins and singletons suggested that copy number variations (CNVs) may accumulate with ageing in a dynamic fashion[37]. By comparing CNVs in longitudinally collected blood samples of MZ pairs, both increases and decreases in CNV content were found after 10 years (between co-twins and within individual twins). This may reflect fluctuations in the proportions of peripheral blood cells carrying aberrant DNA. By comparing copy numbers in buccal cells of twins and their parents, evidence was found for a pre-twinning *de novo* duplication in a healthy twin pair (that was present in both twins but not in their parents) and a post-twinning *de novo* deletion in one twin from a pair of twins who were concordant for attention problems[38]. A comparison of CNVs in the blood of MZ pairs who were discordant for congenital diaphragmatic hernia and oesophageal atresia found no evidence for structural genomic differences between twins[39]. All of these studies used microarrays, which cover a limited portion of the total content of structural variation in the genome[40]. The application of whole-genome-sequencing techniques may unravel many more sequence differences between MZ twins, including single-nucleotide substitutions.

In 2010, the first study was published that applied whole-genome-sequencing technology in discordant MZ twins[41]. The study entailed a combination of techniques — including whole-genome sequencing, RNA sequencing and genome-wide SNP microarrays — to measure multiple molecular marks in $CD4^+$ cells from female twins who are discordant for multiple sclerosis. Only a small fraction of SNPs and structural variants differed within twin pairs, but no differences were replicated across methods. However, this study should be interpreted as exploratory, as only three discordant pairs were studied. Larger studies are needed to establish whether molecular differences may explain discordance for multiple sclerosis and other diseases in MZ twin pairs.

Timing the occurrence of *de novo* mutations

A unique advantage of studying disease-causing mutations in MZ twins is that the developmental timing of *de novo* mutations[42] may be tracked if DNA from multiple cell lines is available for both twins. The timing of a mutation in the sodium channel α1 subunit gene (*SCN1A*) that causes Dravet's syndrome was determined by sequencing DNA from several embryonic tissue lineages from a pair of discordant MZ twins[43]. As the mutation was present in all analysed cell lines of the affected twin but not in those of the unaffected co-twin, it was concluded that the mutation had probably occurred at the two-cell stage in the pre-morula embryo. For any disease caused by *de novo* mutations, information about the timing of mutagenesis is of major importance for genetic counselling. Mutations that occur in parental gametes are associated with a negligible risk of recurrence in additional offspring. By contrast, parental germline mosaicism for the mutation is associated with a high recurrence risk because many existing parental gametes will carry the mutation.

Phenotypic impact of epigenetic variation

DNA methylation and disease. In addition to *de novo* mutations in the DNA, epigenetic variation may be another important source of phenotypic variation and discordance in MZ twins. The following example demonstrates this point. In 1997, a pair of MZ girls was born; one of them was healthy, but the other had a severe spinal malformation in which the spinal cord was duplicated. This defect resembled a condition in mice with a mutation in the *Axin1* gene, but no mutation was found in this gene in the twins. However, increased methylation of CpG sites at the *AXIN1* promoter was found in the affected twin as compared with the unaffected twin, and this may have suppressed gene expression and caused the malformation[44].

Although epigenetic variation has not yet been investigated in large twin studies, several small studies

Box 2 | The classical twin design

In the classical twin design, the extent to which phenotypic variation in a trait (V_P) is due to genetic (V_G) and environmental (V_E) influences is estimated as $V_P = V_G + V_E$. Genetic variance can be further decomposed into additive genetic variance (V_A) and variance due to non-additive genetic effects (dominance variance (V_D)): $V_G = V_A + V_D$. Most twin studies, unless they are very large, consider the narrow-sense heritability (h^2), which refers to the proportion of variation that is due to additive genetic variance: $h^2 = V_A / V_P$. Environmental influences (V_E) comprise those that are shared by family members ('the common environment' (V_C)) and influences that are unique to each individual ('the unique environment' (V_U)): $V_E = V_C + V_U$.

These unobserved variance components can be estimated from the observed resemblance (that is, the phenotypic covariance) in monozygotic (MZ) and dizygotic (DZ) twin pairs. MZ twins are derived from a single fertilized egg cell and share (nearly) 100% of their segregating genes, whereas DZ twins are derived from two distinct zygotes and share on average 50% of their segregating genes. Twins of both types share 100% of the common environment and 0% of the unique environment. Therefore, the phenotypic covariance of MZ twins is expected to equal $V_A + V_D + V_C$, and the phenotypic covariance of DZ twins is expected to equal $0.5\,V_A + 0.25\,V_D + V_C$. These expectations are the input (that is, the structural equations) for genetic structural equation modelling (GSEM), a technique by which maximum likelihood estimates of variance components are obtained from twin data. GSEM obtains the expected MZ and DZ covariances given the equations above and compares the outcome to the covariances observed in the data. The maximum likelihood estimates of V_A, V_D, V_C and V_E are those estimates that predict covariances that are most consistent with the observed data. With MZ and DZ data, V_C and V_D cannot be estimated simultaneously. V_D is estimated if there is stronger evidence for non-additive effects (if the MZ correlation is more than twice as large as the DZ correlation), and V_C is estimated if there is stronger evidence for common environmental effects (if the MZ correlation is less than twice as large as the DZ correlation). In extended-twin-family designs, the information from additional types of family relations together with the information from twins allows for estimating V_A, V_D, V_C and V_E simultaneously.

In multivariate twin models, extending the set of equations for the expected covariances allows the modelling of the cross-twin–cross-trait covariance — that is, the covariance of trait one in one twin with trait two in the co-twin. To estimate the degree to which the clustering of different traits or comorbidity of disorders is explained by genetic and environmental influences, the same principles apply as for the expected covariances of twins (for example, MZ twins are expected to share 100% of genetic influences that overlap between traits, whereas DZ twins are expected to share 50%, resulting in a larger cross-twin–cross-trait covariance for MZ twins if the association between traits has a genetic basis).

Genetic non-additivity
Refers to genetic effects that contribute to the phenotypic variance in a non-additive manner. These include the effects of interacting alleles at a single locus (dominance) and interactions between different loci (epistasis).

Assortative mating
Refers to the situation whereby a trait is correlated in spouses because it influences partner choice (phenotypic assortment) or because it correlates with certain environments that influence partner choice (social homogamy). It is also called nonrandom mating.

Maternal effects
Effects that are transmitted from mother to offspring, including genetic effects. The phenotype in offspring can be influenced by: the maternal allele, mitochondrial inheritance, the effects of the prenatal environment (for example, nutrient supply *in utero*) or the maternal supply of RNA or proteins to the egg cell.

Co-twin control method
A method of examining the associations between traits using discordant twins. If monozygotic twins who are discordant for trait 1 are also discordant for trait 2, the association between these traits is unlikely to be confounded by underlying shared genetic or early environmental influences.

Transgenerational inheritance
The transmission of a trait across generations (genetic or cultural inheritance). Epigenetic variation may also be transmitted across generations.

Imprinting
The mechanism that can occur at some loci to silence the expression of one of the two alleles, depending on the parent-of-origin of the allele.

Copy number variations
(CNVs). These refer to large DNA segments (> 1 kb) of which the number of copies is variable (for example, between individuals or between cells within an individual) — for example insertions, deletions and duplications.

demonstrate the promise of the discordant twin design for epigenetics, including studies of Alzheimer's disease[45], autism[46], bipolar disorder[47,48], birth weight[49], cancer[50] and SLE[1]. In MZ twins who are discordant for the autoimmune disorders SLE, rheumatoid arthritis or dermatomyositis, a global decrease in DNA methylation (hypomethylation) was identified in SLE-affected twins, as were regional DNA methylation changes at 49 genes that were enriched for immune function[1]. Many of the genes that were hypomethylated in SLE-affected twins also showed increased expression compared with the healthy co-twin[1]. Integrated studies of DNA methylation and gene expression in discordant twins[51] are particularly valuable for identifying loci at which epigenetic regulation may be associated with disease. Importantly, the dynamic nature of epigenetic variation makes results of epigenetic studies more difficult to interpret compared with genetic studies. Alternatively to being the cause of disease discordance, epigenetic differences may also reflect the effects of disease or the effect of an event occurring in one twin that independently triggered both the disease and the epigenetic changes. Some twin registries have collected longitudinal biological samples, and this allows for identifying epigenetic differences between twins that were already present before the onset of discordance for some diseases. Functional studies will ultimately be required to verify the effect of epigenetic variation.

The classical twin design provides information about the importance of genetic influences on epigenetic variation: comparison of the level of DNA methylation at the imprinted *IGF2–H19* locus in MZ and DZ twins showed that variation in DNA methylation at this locus is mainly determined by heritable factors before middle age[52]. High heritability of epigenetic variation has also been observed for some other loci[53,54], although the average heritability across all loci seems to be low[55].

Differential miRNA expression and disease. The role of non-coding RNAs such as microRNAs (miRNAs)[56–58] is fairly unexplored. In a sample of MZ twins and sibling pairs who were discordant for autism, miRNA expression in lymphoblastoid cell lines was measured, and differential regulation of a number of miRNA transcripts was observed[59]. For two differentially expressed brain-specific miRNAs, the putative target genes inhibitor of DNA binding 3 (*ID3*) and polo-like kinase 2 (*PLK2*), which have been implicated in circadian rhythm signalling and modulation of synapses, were validated by experiments involving knockdown or overexpression of these miRNAs. By combining miRNA data and mRNA expression data, dysregulation of miRNA expression was found to contribute to alterations in target gene expression, which in turn may contribute to disease pathology of autism. The expression of miRNAs was measured in MZ twins who were discordant for lupus nephritis, and differential expression of several miRNAs was observed[60]. Primarily among the gene targets of the most important miRNAs were genes that have a role in interferon signalling (IFN signalling). Together, these studies indicate that the discordant MZ twin design will be a valuable approach towards exploring the role of miRNA expression in complex disease.

Gene expression: causes and disease links

There is wide variation in the heritability of transcript expression across the genome[2,61]. To identify expression quantitative trait loci (eQTLs), variation in expression across tissues of healthy female twin pairs was investigated in a 'matched co-twin analysis'[62]. In the initial stage, SNP associations were tested in one twin of each pair. Although this method of eQTL identification does not require twins, the co-twins in this study served to replicate and validate the identified eQTLs, thus providing extra confidence in the findings.

A frequent use of twin studies is to identify gene expression alterations (on a shared genetic background) that are associated with various disease states; such genes may provide mechanistic insight into disease pathogenesis. A study of gene expression in subcutaneous fat of obesity-discordant MZ twins detected differential expression of a range of genes[63]. Differentially expressed genes included those that are involved in inflammatory pathways (which were upregulated in obese twins) and in mitochondrial branched-chain amino acid (BCAA) catabolism (which were downregulated in obese twins). Interestingly, the largest increase in expression in obese twins was reported for the gene encoding the inflammatory cytokine osteopontin (*SPP1*), which has previously been associated with obesity and insulin resistance in mice. Other diseases for which gene expression changes have been identified in discordant MZ twins include rheumatoid arthritis[64], bipolar disorder[65], schizophrenia[66] and type 1 diabetes[67,68]. A comparison of the skeletal muscle transcriptomes in MZ twins who are discordant for postmenopausal oestrogen-based hormone replacement therapy (HRT) highlights the insights that may be obtained from MZ twins who are discordant for drug treatment, regarding the long-term effects of drug therapies[69]. Several pathways were differentially regulated in twins who received hormonal treatment, and expression differences correlated significantly with differences in muscle performance between the twins. Large twin studies estimating the heritability of expression of individual transcripts have not yet been published.

Metabolomics

Metabolites may serve as biomarkers of health and disease[70] and can be quantified in body fluids and tissue samples by approaches such as mass spectrometry and ^{1}H NMR spectroscopy. The first metabolomics study based on ^{1}H NMR spectroscopic analysis of urine and blood plasma from MZ and DZ twin pairs showed that familial factors (such as genetic influences and family environment) explain on average 42% of the variation in individual metabolite peak heights in plasma and 30% of the variation in urine[71]. In two GWASs of metabolite profiles, data from twins allowed the proportion of variance in metabolite levels explained by significantly associated SNPs to be compared with the proportion explained by the total genetic or familial variance[72,73]. Heritability estimates of metabolic measures based on data from 221 MZ twin pairs and

Table 2 | **Heritability estimates from twin studies**

Trait	Heritability	Number of twin pairs (or study type for multiple data sets)*	Refs
Anthropometric			
Height	M: 0.87–0.93; F: 0.68–0.90‡	30,111	126
Body mass index	M: 0.65–0.84‡; F: 0.64–0.79‡	37,000	127
Birth weight	0.42	2,009§	128
Metabolic and cardiovascular			
Diabetes, type 1	0.88	22,650	129
Diabetes, type 2	0.64	13,888	130
Coronary heart disease	M: 0.57; F: 0.38	10,483	131
Systolic blood pressure	0.42	1,617‖	132
Diastolic blood pressure	0.40	1,617‖	132
Markers for cardiovascular disease in blood		12,000 twins	133
High-density lipoprotein (HDL) level	0.66		
Low-density lipoprotein (LDL) level	0.53		
Triglyceride level	0.54		
Glucose level	0.53		
C-reactive protein (CRP) level	0.43		
Brain and central nervous system disorders			
Alzheimer's disease	0.48	662	134
Parkinson's disease	0.34	46,436 twins	135
Migraine	0.34–0.57‡	29,717	136
Multiple sclerosis	0.25–0.76‡	Review	137
Attention-deficit hyperactivity disorder	0.76	Review	138
Autism spectrum disorders	0.71	11,535 twins	139
Schizophrenia	0.81	Meta-analysis	140
Major depression	0.37	Meta-analysis	141
Electroencephalography measures of brain activity			
Alpha power	0.79	Meta-analysis	119
P300 amplitude	0.60	Meta-analysis	119
Magnetic resonance imaging measures of brain structure			
Total brain volume	0.66–0.97	Review	118
Frontal lobe volumes	0.90–0.95	Review	118
Hippocampal volumes	0.40–0.69	Review	118
Skeletal features and disorders			
Bone mineral density	0.60–0.80	Review	142
Osteoarthritis	0.40–0.70	Review	143
Rheumatoid arthritis	0.60	13,502	144
Asthma and pulmonary function			
Asthma	0.60¶	21,135	145
Forced expiratory volume in one second	0.61	4,314 twins	146
Forced vital capacity	0.55	4,314 twins	146
Peak expiratory flow	0.43	4,314 twins	146

340 DZ twin pairs ranged between 23% and 55% for amino acids and other small-molecule metabolites[72]. Estimates were higher for lipids (48–62%) and lipoproteins (50–76%). Although for most direct metabolite measures the total variance explained by significantly associated SNPs was 10% at most, higher estimates of explained variance were observed for certain metabolite ratios. The highest explained variance (25%) was observed for the ratio of linoleic acid to other polyunsaturated fatty acids. The twin-based heritability for this ratio was 62%, implying that 40% of the total heritability can be ascribed to SNPs, which is high compared with most other (clinical) phenotypes.

Whereas traditional enzymatic methods usually provide composite measures of metabolites, ^{1}H NMR gives more detailed insight into the behaviour of individual metabolites in pathways. In a direct comparison, similar estimates of heritability were found for most composite lipid measures on the basis of either enzymatic methods or ^{1}H NMR[72]. This supports the notion that high-resolution metabolomics techniques are reliable.

Similarly to differentially expressed genes, differential levels of other molecules can be linked to disease pathogenesis. After detecting differences in serum and fat tissue lipid profiles in MZ twins discordant for obesity[74], a simulation of lipid bilayer dynamics was carried out using lipidomics and gene expression data from the twins, providing novel functional insights into the biological pathways that underlie adipocyte expansion[75]. This study shows how findings from discordant twin studies may encourage and guide further functional or bioinformatics approaches to obtain in-depth mechanistic insights into the pathological mechanisms that underlie complex traits and disease.

To date, there have been few proteomic studies in twins. A twin study of serum protein levels, as measured by antibody arrays, found that only a small proportion of the variation was attributable to familial factors; however, experimental variation in this study was fairly large[76].

Tissue specificity of molecular variation

In concordance with most molecular and genetic epidemiological studies, most twin studies have been based on peripheral blood. But how well does a molecular profile in blood cells reflect epigenetic and gene expression changes that are associated with different phenotypes and diseases in relevant tissues? Epigenetic changes that arise at later stages of development and throughout life are more likely to be limited to specific tissues or even cells. DNA methylation profiles of MZ twins who are discordant for major psychosis suggest that epigenetic changes related to psychosis may be reflected in peripheral blood[77]. In this study, the most significant methylation change in psychosis-affected twins — that is, hypomethylation at the promoter of the *ST6GALNAC1* gene — was also evident in post-mortem brain tissues of some psychosis patients. However, large studies are warranted to establish how well molecular profiles in blood reflect those occurring in

Table 2 (cont.) | **Heritability estimates from twin studies**

Trait	Heritability	Number of twin pairs (or study type for multiple data sets)*	Refs
Cancer			
Prostate cancer	0.42	21,000	114
Breast cancer (in females)	0.27	23,788	114
Colorectal cancer	0.35	44,788	114
Ageing			
Mortality	0.25	Review	34
Telomere length	0.56	175	35
Lifestyle and life events			
Exercise participation	0.48–0.71‡	37,051	89
Dietary patterns	0.41–0.48	3,262‖	90
Smoking initiation	M: 0.37; F: 0.55	Meta-analysis	147
Smoking persistence	M: 0.59; F: 0.46	Meta-analysis	147
Alcohol abuse or dependence	0.50–0.70	Review	148
Stressful life events	0.28	Meta-analysis	92

*Note that numbers refer to twin pairs unless stated otherwise, and most heritability estimates refer to the narrow-sense heritability (h^2; BOX 2). ‡Range of heritabilities from different countries or study samples. §Female twin pairs with child (offspring-of-twin design). ‖Only females. ¶The original paper reports estimates for various age categories from 3–71 years, separately for males and females. F, females; M, males.

tissues that are relevant to disease, because molecular characteristics, particularly epigenetic and gene expression profiles, are known to be largely tissue-specific. Although many of the relevant disease tissues are difficult if not impossible to obtain from large groups of living subjects, several twin registries are collecting biological samples from various sources other than blood, including saliva, buccal cells, hair, skin, fat, muscle, urine and stool (TABLE 1; Supplementary information S1 (table)).

An issue that is of particular relevance to MZ twins, and possibly also to DZ twins[78], is chimerism. Twins can exchange fetal blood through vascular connections between their circulatory systems. As a result, MZ twins can be haematopoietic chimaeras, and a variable fraction of cells derived from the haematopoietic stem cells (for example, peripheral blood cells) in each twin may actually originate from the co-twin. This process can have implications for the detection of genetic or epigenetic events that are related to discordance originating *in utero*, as some cells in unaffected twins may carry the genetic or epimutation of the co-twin. A study of twins who are discordant for transient neonatal diabetes mellitus type 1 (TNDM1) found that buccal cells only displayed hypomethylation of the *TNDM1* locus in the affected twins, whereas the same epigenetic change was evident in blood samples from both twins[79]. The issue may likewise influence the results of DNA sequence analysis of blood samples from MZ twins[80], although a study in healthy twins suggested that MZ twin concordance for SNPs and copy number in blood versus buccal cells is highly similar[81].

Congenital diaphragmatic hernia
A birth defect that is characterized by malformation of the diaphragm, lung hypoplasia and pulmonary hypertension.

Oesophageal atresia
A congenital malformation of the oesophagus in which the oesophagus does not form an open passage to the stomach and may be connected to the trachea.

Maximum likelihood
Maximum-likelihood estimation obtains estimates of population parameters from a data set by computing the probability (likehood) of obtaining the observed data for a range of different parameter values and evaluating for which values the probability of observing the data is highest.

Dravet's syndrome
A childhood-onset epileptic encephalopathy that is also called severe myoclonic epilepsy of infancy.

Host genetic influences on the microbiome

Studies of the human gut microbiome have revealed considerable variation in the composition of microbial communities between individuals. It remains to be established to what degree this variation is controlled by host genetics[82], but greater similarity has been observed in family members compared to unrelated individuals. A few studies have explored the role of host genetics by comparing various measures of the microbiome in small groups of MZ and DZ twins, but findings have so far been inconclusive, with some studies suggesting that the microbiota are slightly more similar in MZ twins compared with DZ twins[83,84] and other studies observing comparable levels of similarity of the faecal microbiome of MZ and DZ twins[85]. An important factor in the comparison of similarity of individuals is the level that is compared: the overlap between relatives may be small at the organismal level but might be larger at relevant functional levels (for example, the degree to which microbial genes and metabolic pathways are shared).

A few studies in twins searched for microbial signatures that are associated with disease. A comparison of the faecal microbial communities in (concordant) obese and lean MZ twins showed that obesity is associated with various changes, including reduced bacterial diversity and differences in the representation of specific bacterial genes and metabolic pathways[85]. In MZ twins who are discordant for inflammatory bowel diseases, certain gastrointestinal bacterial populations differed in abundance among individuals with different clinical phenotypes of Crohn's disease, which is relevant to our understanding of the pathogenesis behind inflammatory bowel diseases[86]. MZ twins who are discordant for ulcerative colitis differed in the composition of the microbiota and in the expression of human RNA transcripts that are related to oxidative and immune responses in the mucosal epithelium[87]. In affected twins, fewer RNA transcripts correlated with bacterial genera than in unaffected twins, suggesting that ulcerative colitis may be associated with a loss of interaction between the mucosal transcriptional profile and the colonic microbiota.

The interplay of genes and environment

Genetic and environmental influences in many cases do not act independently. Gene–environment correlation (rGE) refers to the situation in which exposure to certain environments is under genetic control[88]. For instance, twin and adoption studies have found that lifestyle factors (for example, exercise participation[89] and diet[90]), life events (for example, divorce[91]) and life circumstances (for example, family environment and social support[92]) are moderately heritable. Thus, influences that are usually considered as measures of 'environment' might often be better described as external factors that are partly under genetic control[93]. By contrast, G×E interaction refers to the scenario in which different genotypes have different reactions to the same environmental exposure[94,95]. By comparing differences in serum lipid levels in MZ twins across pairs with different genotypes, it was found that the Kidd (*JK*) blood group locus is associated with variability in the total cholesterol level[96].

Mosaicism
The situation in which the tissue of an individual consists of two or more genetically distinct cell lines owing to somatic mutation but originally derived from one (genetically homogeneous) zygote.

Non-coding RNAs
RNA transcripts that are not translated into protein but probably serve a regulatory function.

MicroRNAs
(miRNAs). A type of non-coding RNA with an average length of 22 nucleotides that has been suggested to have an important role in post-transcriptional gene regulation networks.

Lymphoblastoid cell lines
Cell lines derived from lymphocytes that have been immortalized, cultured and stored to provide a renewable source of DNA and RNA.

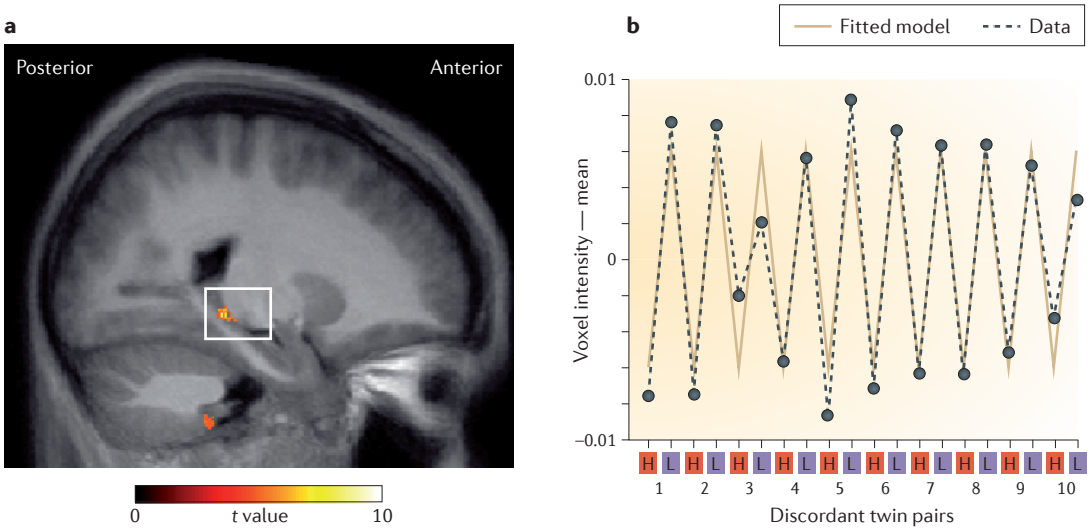
Interferon signalling
(IFN signalling). A signalling system for communication between cells that is involved in the immune response to pathogens and tumours.

Expression quantitative trait loci
(eQTLs). Genomic regions that are associated with the level of expression of an RNA transcript. eQTLs can be tissue-specific.

Mass spectrometry
A technique for determining the mass-to-charge ratio of ions on the basis of their separation in an electromagnetic field. The measured ratios and their relative intensities provide information about both the identity and the abundance of the molecules that gave rise to the ions.

^{1}H NMR spectroscopy
A metabolomics technique that provides information about the structure and quantity of hydrogen-containing molecules. It is based on the absorption and emittance of radiofrequency energy by hydrogen atoms when placed in a strong magnetic field, with wavelengths depending on the atoms' position in the molecule.

Box 3 | **The value of twins in neuroimaging genetics**



Imaging genetics is a form of association analysis in which the phenotype is a measure of brain structure or function (for example, the physiological response of the brain during information processing)[116,117]. Brain imaging studies in twins have contributed substantially to the knowledge that individual differences in brain structure[118] and function[119] are highly heritable. A group of ten male monozygotic (MZ) twin pairs and their non-twin brothers had their brains scanned in a functional magnetic resonance imaging (fMRI) study while they had to memorize a short span of digits (called the digit-memory task)[120]. Before they were asked to recall the digits they memorized, a distraction task was presented in which objects such as fruits, vegetables and tools had to be categorized. When they were distracted by the object categorization task, many men used brain areas that are associated with language for recalling the digits they had memorized. These men took longer to provide the answer than did those who resorted to a visual–spatial memory system to encode the numbers. MZ twins used the same strategy more often than their non-twin brothers, indicating that there are qualitative differences in how individuals think and that these differences have a substantial genetic component.

Another design in imaging genetics compares disease-discordant and disease-concordant MZ twins to assess whether genetic and environmental risk factors for psychiatric disorders act on the same brain regions. Comparisons of discordant MZ twins can highlight brain regions that are susceptible to environmental risk factors. Contrasting MZ twins who both score high on the disease phenotype to those who both score low can be used to identify brain characteristics that are related to genetic risk for disease. An imaging study of bipolar disorder that made use of this design found that white matter pathology in the frontal lobe may be central to the genetic risk of developing bipolar disorder, whereas widespread grey matter abnormalities may be more related to environmental effects and may reflect effects of the illness itself[121]. A study of MZ twins who were discordant or concordant for anxious depression found that environmental risk is highlighted in the left temporal lobe (see the figure)[122]. Most notable were the lower grey matter volumes in the left posterior hippocampus, which contains the main afferent and efferent connections of the hippocampus to the rest of the temporal lobe, in high-risk twins from discordant pairs. The figure illustrates the striking differences in discordant MZ twins, both at the group and individual-pair level. The boxed region in panel **a** shows the left parahippocampal area, where a significant volume reduction was found in the high-risk twin compared with the low-risk co-twin from MZ twin pairs who were discordant for anxious depression. The reduction was not evident in MZ pairs who were concordant for a high risk of depression when compared with MZ twin pairs who were concordant for low risk of depression. The within-pair comparison of discordant MZ pairs is most likely to reveal differences related to environmental exposures, whereas the between-pair comparison of concordant-high pairs and concordant-low pairs is more likely to reveal differences in genetic vulnerability. Therefore, changes in the left parahippocampal area may be specific to an environmentally driven aetiology of anxiety and depression. Colours represent the effect size (*t* value from paired *t* test) of the comparison of grey matter volume between discordant twins. Panel **b** shows the relative responses (individual voxel intensity minus mean voxel intensity in all twins) of ten discordant twin pairs at the most significant voxel in the left parahippocampal area (in this panel, 'H' indicates the twin with high risk of anxious depression and 'L' indicates the low-risk co-twin). Although a substantial overall volume reduction was found in the group of discordant pairs, this figure illustrates that there is a large variation in volume difference across individual discordant pairs. The figure is reproduced, with permission, from REF. 122 © (2007) Elsevier.

A similar approach was used to test whether an interaction between the length polymorphism in the *SLC6A4* serotonin transporter gene and environmental stress is associated with MZ discordance for depression; no evidence was found for this hypothesis[97].

Testing classical assumptions

MZ twins share all of their segregating genes, whereas DZ twins share on average 50%. The assumptions of the classical twin method and the interpretation of results have always been a subject of debate (for a

Table 3 | **MZ and DZ twin concordance for complex disease**

	Probandwise concordance* (%)		Refs
	MZ twins	DZ twins	
Type 1 diabetes	42.9	7.4	129
Type 2 diabetes	34	16	130
Multiple sclerosis	25.3	5.4	149
Crohn's disease	38	2	150
Ulcerative colitis	15	8	150
Alzheimer's disease	32.2	8.7	134
Parkinson's disease	15.5	11.1	151
Schizophrenia	40.8	5.3	152
Major depression	31.1‡ or 47.6§	25.1‡ or 42.6§	153
Attention-deficit hyperactivity disorder	82.4	37.9	154
Autism spectrum disorders	93.7	46.7	155
Colorectal cancer	11	5	114
Breast cancer	13§	9§	114
Prostate cancer	18	3	114

*Defined as $2C / (2C + D)$, where C is the number of concordant affected twin pairs, and D is the number of discordant twin pairs. ‡Concordance in male twin pairs. §Concordance in female twin pairs. DZ, dizygotic; MZ, monozygotic.

detailed discussion of the difficulties related to the concept of heritability, see REF. 98). A first assumption is that MZ twins are genetically identical, for which it has now been proved that there are exceptions to the rule. Still, the difficulty of various whole-genome-sequencing efforts to find any replicable differences between MZ twins[39,41] suggests that DNA sequence differences between MZ twins are not large, although an exact estimation of somatic sequence variation (given the nontrivial error rate in sequencing itself) has not been reported.

The availability of genome-wide marker data also allows us to address the assumption that DZ twins share on average 50% of their segregating genetic material by estimating the true amount of genetic material that DZ twins have inherited from the same parent (that is, identity-by-descent sharing (IBD sharing)). Using genome-wide microsatellite marker data, it was demonstrated that the proportion of IBD sharing in most (95%) DZ twins and siblings lies within the range of 42–58%, with an average close to 50%[99]. Using the empirical IBD measure instead of assumptions about genetic sharing, the heritability of height was estimated at 0.86, which is highly consistent with results from traditional twin studies, providing perhaps the most pertinent evidence to support the estimates of narrow-sense heritability from twin studies.

MZ twins share environmental influence to the same degree as DZ twins. Now that the classical twin design is being used to study epigenetic variation, it is becoming evident that novel attention has to be paid to the assumption that MZ twins share environmental influences to the same degree as DZ twins. Because MZ twins are derived from a single zygote, they may start out with more similar epigenomes than DZ twins, who originate from two zygotes with unique epigenetic profiles. DZ twins may thus start with more epigenetic differences than MZ twins owing to a cause that is not necessarily related to genetic differences. Although this hypothesis remains to be tested, an important observation in this light has been provided by a comparison of small groups of MZ twins that were either monochorionic or dichorionic. The DNA methylation profiles of buccal epithelial cells were more similar in dichorionic MZ twins than in monochorionic MZ twins[55], and this may be related to the timing of splitting of the zygote. Thus, differences in epigenetic resemblance of monochorionic and dichorionic twins may be due to epigenetic divergence of embryonic cells that takes place after the blastomeric stage. Although this issue requires further study in larger samples, it shows that prenatal developmental processes related to twinning may influence the epigenetic resemblance of twins. Importantly, if MZ twins are epigenetically more similar than DZ twins owing to non-genetic causes, the heritability of phenotypes that are epigenetically regulated may be overestimated.

Twin concordance and disease liability

Relationship between heritability and discordance rates in MZ twins. A high concordance of MZ twins on its own does not imply a high heritability, as demonstrated by concordance for measles. Before immunization was introduced, concordance was close to 100% in both MZ and DZ twins[100]. This indicates that, despite the high concordance in MZ twins, genetic differences between individuals actually contribute little to differences in the vulnerability to this infectious disease. Likewise, a high rate of disease discordance in MZ twins does not rule out the importance of genetic influences. Although MZ twins are usually remarkably similar in appearance, MZ twins who are discordant for disease are often observed (TABLE 3). It is generally assumed that liability to disease is continuous, and disease becomes evident after a threshold has been passed. The probability of observing discordant MZ twins thus depends on the heritability of the underlying liability and on the level of the threshold[101]. Especially for rare disorders (for which the threshold is high), many affected MZ twins are discordant even if the heritability is high (for example, schizophrenia, attention-deficit hyperactivity disorder, autism, multiple sclerosis or type 1 diabetes). From the dimensional view of disease liability, it also follows that despite striking differences in clinical appearance, discordant MZ twins can be quite similar in terms of underlying disease liability (FIGURE 1).

Trait concordance in MZ twins, penetrance and disease risk prediction. The presence of disease-discordant twins indicates that genomes cannot completely predict the disease outcome of individuals, even if most variation in disease outcome between individuals is caused by genetic differences. For example, for schizophrenia, despite the high heritability of 80%, the probandwise concordance between MZ co-twins is only 40–50%.

Lipid bilayer dynamics
The dynamic properties of lipid bilayer membranes, such as thickness, fluidity and permeability, that influence the physiological properties of a cell.

Lipidomics
The comprehensive study of the entire set of lipids in biological systems, such as cells, tissues and organs, using metabolomics techniques.

Chimerism
The situation in which an individual carries some of the genetic material originating from another individual (for example, originating from the co-twin or originating from the mother).

Microbiota
The collection of all microorganisms living in a certain environment (for example, the human gut).

Identity-by-descent sharing
(IBD sharing). Refers to the proportion of alleles in two individuals that are derived identically by descent from a common ancestor.

Monochorionic
Describes twins who share the outer membrane (chorion) surrounding the embryos *in utero*. Monochorionic monozygotic twins result when the zygote splits ≥3 days after fertilization.

Dichorionic
Describes twins who do not share the chorion surrounding the embryos *in utero*. Dizygotic twins are always dichorionic. Dichorionic monozygotic twins result when the zygote splits early after fertilization.

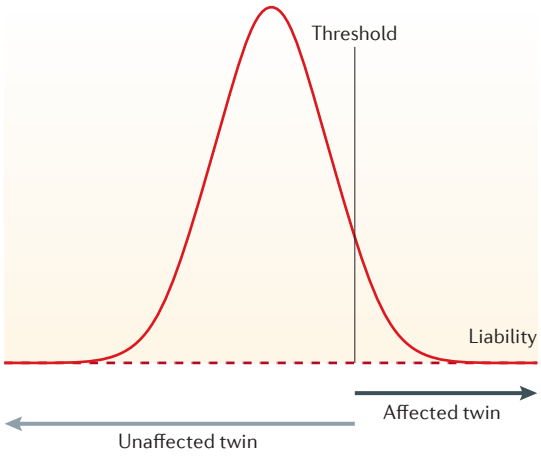


Figure 1 | **Liability threshold model and disease discordance in monozygotic twins.** The liability threshold model assumes that multifactorial diseases result from an underlying continuous character (that is, liability) that is normally distributed in the population[123]. If the combined effects of genetic and environmental influences push an individual's liability across a certain threshold level, the individual is affected. In the population, the proportion of individuals with a liability above the threshold is reflected in the disease prevalence. In discordant monozygotic (MZ) twin pairs, only one twin has a liability above the threshold, although the liability of the unaffected twin may also be high. The black arrow displays the potential range of liabilities of affected twins from discordant MZ twin pairs, and the grey arrow displays the potential range of liabilities of unaffected twins. A comparison of MZ twins who were discordant for congenital diaphragmatic hernia and oesophageal atresia found no differences in genomic structural variation between co-twins[39]. However, structural events in relevant genomic regions that may have contributed to the genetic predisposition of both twins were detected in several pairs; these events were rarely observed in individuals from a healthy control population. A metabolomic study of MZ twins who were discordant for schizophrenia found that, relative to healthy individuals in concordant pairs, the unaffected twins from discordant female pairs showed similar (although smaller) metabolic changes than the affected co-twins[124]. These examples demonstrate that the liability of unaffected twins from discordant pairs may also be elevated. However, this feature does not argue against the value of studying discordant MZ twin pairs to search for the molecular events that caused the affected twin to pass the threshold or events that protected the unaffected twin. Of interest, a study of neurofibromatosis type 1 (NF1) in MZ twins with the same causal mutation in the *NF1* gene but highly variable disease phenotypes revealed considerable variation between twins in DNA methylation at the *NF1* gene[125].

The fact that MZ twin concordance for common disorders is not always high has important implications for genomic risk prediction and the ethical concerns that have been raised in this light. Even if we knew all of the genetic variants that contribute to differences in disease risk between individuals, we would still not be able to predict with certainty the disease risk of all individuals on the basis of their DNA sequence.

Conclusions

Insights that can be obtained from twin studies extend far beyond the classical estimates of heritability. Traditional comparisons of the phenotypic resemblance of twins have been extended to studies of molecular variation across biological samples, providing functional insights into the underlying biology of heritable traits. The study of discordant MZ twins is a powerful method to identify DNA sequence variants, epigenetic variation and metabolites that are associated with disease.

One might feel that there are few aspects of the human condition that have not been investigated in twins; however, new aspects emerge all the time. We have emphasized the value of twin studies in refining phenotypic and clinical definitions and to evaluate biomarkers for disease, but the use of twins can go even further. In recent years, political scientists, sociologists and even economists have become engaged in twin studies. A study of MZ twins who were infected with HIV through blood transfusion at birth but who had strikingly different clinical outcomes used the identical genetic background of twins as a model to study the evolutionary processes and population dynamics that shape viral diversity[102].

In the coming years, longitudinal phenotypic information coupled with biological material collected by worldwide twin registries (TABLE 1; Supplementary information S1 (table)) will be an important resource for large-scale molecular studies. To make optimal use of genetic data collected within twin registries, methods for family-based association analysis are being explored[103]. With the increasing interest in rare genetic variants, there may be renewed interest in linkage studies, in which DZ twins can have an important role. Linkage analysis in DZ twins, contrary to the analysis of non-twin siblings, is not affected by age differences within pairs and is less likely to suffer from non-paternity. Next-generation sequencing across multiple tissues and cell types will facilitate the detection of genome-wide SNPs, CNVs and epigenetic variation in discordant twins at an unprecedented scale, suggesting that twins will continue to provide valuable insights to human genetics.

1. Javierre, B. M. *et al.* Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res.* **20**, 170–179 (2010).
2. McRae, A. F. *et al.* Replicated effects of sex and genotype on gene expression in human lymphoblastoid cell lines. *Hum. Mol. Genet.* **16**, 364–373 (2007).
3. York, T. P. *et al.* Epistatic and environmental control of genome-wide gene expression. *Twin Res. Hum. Genet.* **8**, 5–15 (2005).
4. Martin, N. G. & Eaves, L. J. The genetical analysis of covariance structure. *Heredity* **38**, 79–95 (1977).
5. Kendler, K. S., Heath, A. C., Martin, N. G. & Eaves, L. J. Symptoms of anxiety and symptoms of depression: same genes, different environments? *Arch. Gen. Psychiatry* **44**, 451–457 (1987).
This paper describes one of the first studies that used data from MZ and DZ twins to assess whether the co-occurrence of psychiatric symptoms is explained by a shared genetic or environmental aetiology.
6. Middeldorp, C. M., Cath, D. C., Van Dyck, R. & Boomsma, D. I. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol. Med.* **35**, 611–624 (2005).
7. Brant, A. M. *et al.* The developmental etiology of high IQ. *Behav. Genet.* **39**, 393–405 (2009).
8. Haworth, C. M. *et al.* The heritability of general cognitive ability increases linearly from childhood to

young adulthood. *Mol. Psychiatry* **15**, 1112–1120 (2010).
This paper describes a study based on a large sample of twins from six twin cohorts, showing that the heritability of general cognitive ability increases significantly from childhood to young adulthood.
9. Purcell, S. Variance components models for gene-environment interaction in twin analysis. *Twin Res.* **5**, 554–571 (2002).
This paper describes the implementation of G × E interaction tests in variance component twin analysis, by modelling (unmeasured) genetic effects as a linear function of one or more measures of the environment or moderators.
10. Mustelin, L., Silventoinen, K., Pietiläinen, K., Rissanen, A. & Kaprio, J. Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. *Int. J. Obes.* **33**, 29–36 (2008).
11. Posthuma, D. & Boomsma, D. I. A note on the statistical power in extended twin designs. *Behav. Genet.* **30**, 147–158 (2000).
12. Eaves, L. J. Inferring the causes of human variation. *J. R. Stat. Soc. Ser. A* **140**, 324–355 (1977).
13. Reynolds, C. A., Baker, L. A. & Pedersen, N. L. Models of spouse similarity: applications to fluid ability measured in twins and their spouses. *Behav. Genet.* **26**, 73–88 (1996).
14. van Grootheest, D. S., van den Berg, S. M., Cath, D. C., Willemsen, G. & Boomsma, D. I. Marital resemblance for obsessive-compulsive, anxious and depressive symptoms in a population-based sample. *Psychol. Med.* **38**, 1731–1740 (2008).
15. Magnus, P., Berg, K. & Bjerkedal, T. No significant difference in birth weight for offspring of birth weight discordant monozygotic female twins. *Early Hum. Dev.* **12**, 55–59 (1985).
16. Nance, W. E., Kramer, A. A., Corey, L. A., Winter, P. M. & Eaves, L. J. A causal analysis of birth weight in the offspring of monozygotic twins. *Am. J. Hum. Genet.* **35**, 1211–1223 (1983).
17. Vrieze, S. I. *et al.* An assessment of the individual and collective effects of variants on height using twins and a developmentally informative study design. *PLoS Genet.* **7**, e1002413 (2011).
18. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
19. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
20. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010).
21. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
22. Friberg, L., Cederlof, R., Lundman, T. & Olsson, H. Mortality in smoking discordant monozygotic and dizygotic twins. A study on the Swedish Twin Registry. *Arch. Environ. Health* **21**, 508–513 (1970).
23. Martin, N. G., Carr, A. B., Oakeshott, J. G. & Clark, P. Co-twin control studies: vitamin C and the common cold. *Prog. Clin. Biol. Res.* **103**, 365–373 (1982).
24. de Moor, M. H., Boomsma, D. I., Stubbe, J. H., Willemsen, G. & de Geus, E. J. Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Arch. Gen. Psychiatry* **65**, 897–905 (2008).
25. Ligthart, L., Nyholt, D. R., Penninx, B. W. & Boomsma, D. I. The shared genetics of migraine and anxious depression. *Headache* **50**, 1549–1560 (2010).
26. Lundqvist, E. *et al.* Co-twin control and cohort analyses of body mass index and height in relation to breast, prostate, ovarian, corpus uteri, colon and rectal cancer among Swedish and Finnish twins. *Int. J. Cancer* **121**, 810–818 (2007).
27. Bell, J. T. & Spector, T. D. A twin approach to unraveling epigenetics. *Trends Genet.* **27**, 116–125 (2011).
28. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).
This is the first study indicating that epigenetic profiles of older MZ twins are less similar than those of young MZ twins based on a comparison of global and locus-specific DNA methylation and histone acetylation.
29. Talens, R. P. *et al.* Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* 23 May 2012 (doi:10.1111/j.1474-9726.2012.00835.x).
30. Wong, C. C. *et al.* A longitudinal study of epigenetic variation in twins. *Epigenetics* **5**, 516–526 (2010).
31. Gordon, L. *et al.* Expression discordance of monozygotic twins at birth: effect of intrauterine environment and a possible mechanism for fetal programming. *Epigenetics* **6**, 579–592 (2011).
32. Ollikainen, M. *et al.* DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum. Mol. Genet.* **19**, 4176–4188 (2010).
33. Powell, J. E. *et al.* Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res.* **22**, 456–466 (2012).
34. Hjelmborg, J. B. *et al.* Genetic influence on human lifespan and longevity. *Hum. Genet.* **119**, 312–321 (2006).
This paper describes a study of survival in a large sample of twins, showing that genetic influences on human lifespan are of little importance until the age of 60 but that genes explain an important part of the variation at advanced ages.
35. Bakaysa, S. L. *et al.* Telomere length predicts survival independent of genetic influences. *Aging Cell* **6**, 769–774 (2007).
36. Zwijnenburg, P. J. G., Meijers Heijboer, H. & Boomsma, D. I. Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am. J. Med. Genet. B* **153**, 1134–1149 (2010).
This review provides an overview of studies of MZ twins who are discordant for chromosomal abnormalities, Mendelian disorders and other genetic disorders.
37. Forsberg, L. A. *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* **90**, 217–228 (2012).
38. Ehli, E. A. *et al.* De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on attention problems. *Eur. J. Hum. Genet.* 11 Apr 2012 (doi:10.1038/ejhg.2012.49).
39. Veenma, D. *et al.* Copy number detection in discordant monozygotic twins of congenital diaphragmatic hernia (CDH) and esophageal atresia (EA) cohorts. *Eur. J. Hum. Genet.* **20**, 298–304 (2012).
40. Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nature Rev. Genet.* **12**, 363–376 (2011).
41. Baranzini, S. E. *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351–1356 (2010).
This paper describes a study of female MZ twins who are discordant for multiple sclerosis, and it was the first to report the individual genome sequences of an MZ twin pair based on whole-genome sequencing technology.
42. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nature Rev. Genet.* **13**, 565–575 (2012).
43. Vadlamudi, L. *et al.* Timing of de novo mutagenesis: a twin study of sodium-channel mutations. *N. Engl. J. Med.* **363**, 1335–1340 (2010).
This paper describes a study that sequenced SCN1A in multiple cell lines from MZ twin pairs who are concordant and discordant for Dravet's syndrome to obtain insight into the timing of disease-causing *de novo* mutations.
44. Oates, N. A. *et al.* Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am. J. Hum. Genet.* **79**, 155–162 (2006).
45. Mastroeni, D., McKee, A., Grover, A., Rogers, J. & Coleman, P. D. Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS ONE* **4**, e6617 (2009).
46. Nguyen, A., Rauch, T. A., Pfeifer, G. P. & Hu, V. W. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, *RORA*, whose protein product is reduced in autistic brain. *FASEB J.* **24**, 3036–3051 (2010).
47. Kuratomi, G. *et al.* Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol. Psychiatry* **13**, 429–441 (2008).
48. Rosa, A. *et al.* Differential methylation of the X-chromosome is a possible source of discordance for bipolar disorder female monozygotic twins. *Am. J. Med. Genet. B* **147**, 459–462 (2008).
49. Gao, Y. *et al.* Increased expression and altered methylation of HERVWE1 in the human placentas of smaller fetuses from monozygotic, dichorionic, discordant twins. *PLoS ONE* **7**, e33503 (2012).
50. Galetzka, D. *et al.* Monozygotic twins discordant for constitutive BRCA1 promoter methylation, childhood cancer and secondary cancer. *Epigenetics* **7**, 47–54 (2012).
51. Gervin, K. *et al.* DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet.* **8**, e1002454 (2012).
52. Heijmans, B. T., Kremer, D., Tobi, E. W., Boomsma, D. I. & Slagboom, P. E. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human *IGF2/H19* locus. *Hum. Mol. Genet.* **16**, 547–554 (2007).
53. Coolen, M. W. *et al.* Impact of the genome on the epigenome is manifested in DNA methylation patterns of imprinted regions in monozygotic and dizygotic twins. *PLoS ONE* **6**, e25590 (2011).
54. Gertz, J. *et al.* Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet.* **7**, e1002228 (2011).
55. Kaminsky, Z. A. *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genet.* **41**, 240–245 (2009).
56. Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319**, 1787–1789 (2008).
57. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell. Biol.* **6**, 376–385 (2005).
58. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15** (Suppl. 1), R17–R29 (2006).
59. Sarachana, T., Zhou, R., Chen, G., Manji, H. K. & Hu, V. W. Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. *Genome Med.* **2**, 23 (2010).
60. Te, J. L. *et al.* Identification of unique microRNA signature associated with lupus nephritis. *PLoS ONE* **5**, e10344 (2010).
61. Tan, Q. *et al.* Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Hum. Genet.* **117**, 267–274 (2005).
62. Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
63. Pietiläinen, K. H. *et al.* Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med.* **5**, e51 (2008).
64. Haas, C. S. *et al.* Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins. *Arthritis Rheum.* **54**, 2047–2060 (2006).
65. Matigian, N. *et al.* Expression profiling in monozygotic twins discordant for bipolar disorder reveals dysregulation of the WNT signalling pathway. *Mol. Psychiatry* **12**, 815–825 (2007).
66. Kakiuchi, C. *et al.* Upregulation of ADM and SEPX1 in the lymphoblastoid cells of patients in monozygotic twins discordant for schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **147**, 557–564 (2008).
67. Beyan, H. *et al.* Monocyte gene-expression profiles associated with childhood-onset type 1 diabetes and disease risk: a study of identical twins. *Diabetes* **59**, 1751–1755 (2010).
68. Caramori, M. L. *et al.* Gene expression differences in skin fibroblasts in identical twins discordant for type 1 diabetes. *Diabetes* **61**, 739–744 (2012).
69. Ronkainen, P. H. *et al.* Postmenopausal hormone replacement therapy modifies skeletal muscle composition and function: a study with monozygotic twin pairs. *J. Appl. Physiol.* **107**, 25–33 (2009).
70. Ellis, D. I., Dunn, W. B., Griffin, J. L., Allwood, J. W. & Goodacre, R. Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics* **8**, 1243–1266 (2007).
71. Nicholson, G. *et al.* Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* **7**, 525 (2011).
72. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genet.* **44**, 269–276 (2012).
This was a genome-wide association study of 216 serum metabolites that used twin data to compare the total heritability of each metabolite to the total genetic variance explained by significantly associated SNPs.

73. Nicholson, G. *et al.* A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS Genet.* **7**, e1002270 (2011).
74. Pietiläinen, K. H. *et al.* Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects — a monozygotic twin study. *PLoS ONE* **2**, e218 (2007).
75. Pietiläinen, K. H. *et al.* Association of lipidome remodeling in the adipocyte membrane with acquired obesity in humans. *PLoS Biol.* **9**, e1000623 (2011).
This paper describes a simulation of adipocyte membrane dynamics related to obesity, based on differences in lipid content and differential gene expression detected in discordant monozygotic twins.
76. Kato, B. S. *et al.* Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Sci.* **9**, 73 (2011).
77. Dempster, E. L. *et al.* Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum. Mol. Genet.* **20**, 4786–4796 (2011).
78. van Dijk, B. A., Boomsma, D. I. & de Man, A. J. Blood group chimerism in human multiple births is not rare. *Am. J. Med. Genet.* **61**, 264–268 (1996).
79. Laborie, L. B. *et al.* DNA hypomethylation, transient neonatal diabetes, and prune belly sequence in one of two identical twins. *Eur. J. Pediatr.* **169**, 207–213 (2010).
80. Erlich, Y. Blood ties: chimerism can mask twin discordance in high-throughput sequencing. *Twin Res. Hum. Genet.* **14**, 137–143 (2011).
81. Scheet, P. *et al.* Twins, tissue and time: a comparison of genomic structures. *Twin Res. Hum. Genet.* (in the press).
82. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Rev. Microbiol.* **9**, 279–290 (2011).
83. Stewart, J. A., Chadwick, V. S. & Murray, A. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J. Med. Microbiol.* **54**, 1239–1242 (2005).
84. Zoetendal, E. G., Akkermans, A. D. L., Akkermans-van Vliet, W. M., de Visser, J. A. G. M. & de Vos, W. M. The host genotype affects the bacterial community in the human gastronintestinal tract. *Microb. Ecol. Health Dis.* **13**, 129–134 (2001).
85. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2008).
This paper describes a comparison of faecal microbial communities in monozygotic and dizygotic twins who are concordant for leanness or obesity.
86. Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854 (2010).
87. Lepage, P. *et al.* Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* **141**, 227–236 (2011).
88. Kendler, K. S. & Eaves, L. J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am. J. Psychiatry* **143**, 279–289 (1986).
89. Stubbe, J. H. *et al.* Genetic influences on exercise participation in 37.051 twin pairs from seven countries. *PLoS ONE* **1**, e22 (2006).
90. Teucher, B. *et al.* Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res. Hum. Genet.* **10**, 734–748 (2007).
91. Middeldorp, C. M., Cath, D. C., Vink, J. M. & Boomsma, D. I. Twin and genetic effects on life events. *Twin Res. Hum. Genet.* **8**, 224–231 (2005).
92. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).
93. Vinkhuyzen, A. A. E., Van Der Sluis, S., De Geus, E. J. C., Boomsma, D. I. & Posthuma, D. Genetic influences on 'environmental' factors. *Genes Brain Behav.* **9**, 276–287 (2010).
94. Caspi, A. & Moffitt, T. E. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nature Rev. Neurosci.* **7**, 583–590 (2006).
95. Caspi, A. *et al.* Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. *Proc. Natl Acad. Sci. USA* **104**, 18860–18865 (2007).
96. Berg, K. Variability gene effect on cholesterol at the Kidd blood group locus. *Clin. Genet.* **33**, 102–107 (1988).
97. Wray, N. R. *et al.* Use of monozygotic twins to investigate the relationship between 5HTTLPR genotype, depression and stressful life events: an application of item response theory. *Novartis Found. Symp.* **293**, 48–59 (2008).
98. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era: concepts and misconceptions. *Nature Rev. Genet.* **9**, 255–266 (2008).
99. Visscher, P. M. *et al.* Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Med. Genet.* **81**, 1104–1110 (2007).
100. Jörgensen, G. in *Erbgefüge* (ed. Vogel, F.) 581–665 (Springer, 1974).
101. Smith, C. Heritability of liability and concordance in monozygous twins. *Ann. Hum. Genet.* **34**, 85–91 (1970).
102. Tazi, L. *et al.* HIV-1 infected monozygotic twins: a tale of two outcomes. *BMC Evol. Biol.* **11**, 62 (2011).
103. Ott, J., Kamatani, Y. & Lathrop, M. Family-based designs for genome-wide association studies. *Nature Rev. Genet.* **12**, 465–474 (2011).
104. Galton, F. The history of twins, as a criterion of the relative powers of nature and nurture. *J. Anthropol. Institute Great Britain Ireland* **5**, 391–406 (1876).
105. Mayo, O. Early research on human genetics using the twin method: who really invented the method? *Twin Res. Hum. Genet.* **12**, 237–245 (2009).
106. Siemens, H. W. Die Zwillingspathologie. *Mol. Gen. Genet.* **35**, 311–312 (1924).
107. Zhu, G. *et al.* A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur. J. Hum. Genet.* **15**, 94–102 (2007).
108. Jinks, J. L. & Fulker, D. W. Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of the human behavior. *Psychol. Bull.* **73**, 311–349 (1970).
This is a classical paper that describes the application of the biometrical genetic approach initiated by R. A. Fisher to the analysis of twin and family data.
109. Martin, N. G., Eaves, L. J., Kearsey, M. J. & Davies, P. The power of the classical twin study. *Heredity* **40**, 97–116 (1978).
110. Boomsma, D. I. Twin registers in Europe: An overview. *Twin Res.* **1**, 34–51 (1998).
111. Busjahn, A. & Hur, Y. M. Twin registries: an ongoing success story. *Twin Res. Hum. Genet.* **9**, 705 (2006).
112. Peltonen, L. GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res.* **6**, 354–360 (2003).
113. Llewellyn, C. H., van Jaarsveld, C. H., Johnson, L., Carnell, S. & Wardle, J. Nature and nurture in infant appetite: analysis of the Gemini twin birth cohort. *Am. J. Clin. Nutr.* **91**, 1172–1179 (2010).
114. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
In this study, twin data from Scandinavian twin registries were linked to national health records of cancer diagnosis to obtain heritability estimates for various types of cancer based on large twin samples.
115. Vink, J. M. *et al.* Cervix smear abnormalities: linking pathology data in female twins, their mothers and sisters. *Eur. J. Hum. Genet.* **19**, 108–111 (2011).
116. De Geus, E. J. C. Introducing genetic psychophysiology. *Biol. Psychol.* **61**, 1–10 (2002).
117. Mattay, V. S., Goldberg, T. E., Sambataro, F. & Weinberger, D. R. Neurobiology of cognitive aging: insights from imaging genetics. *Biol. Psychol.* **79**, 9–22 (2008).
118. Peper, J. S., Brouwer, R. M., Boomsma, D. I., Kahn, R. S. & Hulshoff Pol, H. E. Genetic influences on human brain structure: a review of brain imaging studies in twins. *Hum. Brain Mapp.* **28**, 464–473 (2007).
119. Van Beijsterveldt, C. E. M. & Van Baal, G. C. M. Twin and family studies of the human electroencephalogram: a review and a meta-analysis. *Biol. Psychol.* **61**, 111–138 (2002).
120. Koten, J. W. *et al.* Genetic contribution to variation in cognitive function: an fMRI study in twins. *Science* **323**, 1737–1740 (2009).
121. van der Schot, A. C. *et al.* Influence of genes and environment on brain volumes in twin pairs concordant and discordant for bipolar disorder. *Arch. Gen. Psychiatry* **66**, 142–151 (2009).
122. De Geus, E. J. C. *et al.* Intrapair differences in hippocampal volume in monozygotic twins discordant for the risk for anxiety and depression. *Biol. Psychiatry* **61**, 1062–1071 (2007).
123. Falconer, D. S. *Introduction to Quantitative Genetics* (Ronald Press Co., 1960).
124. Tsang, T. M., Huang, J. T., Holmes, E. & Bahn, S. Metabolic profiling of plasma from discordant schizophrenia twins: correlation between lipid signals and global functioning in female schizophrenia patients. *J. Proteome Res.* **5**, 756–760 (2006).
125. Harder, A. *et al.* Monozygotic twins with neurofibromatosis type 1 (NF1) display differences in methylation of *NF1* gene promoter elements, 5'untranslated region, exon and intron 1. *Twin Res. Hum. Genet.* **13**, 582–594 (2010).
126. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).
127. Schousboe, K. *et al.* Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res.* **6**, 409–421 (2003).
128. Clausson, B., Lichtenstein, P. & Cnattingius, S. Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG* **107**, 375–381 (2000).
129. Hyttinen, V., Kaprio, J., Kinnunen, L., Koskenvuo, M. & Tuomilehto, J. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs. *Diabetes* **52**, 1052–1055 (2003).
130. Kaprio, J. *et al.* Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* **35**, 1060–1067 (1992).
131. Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).
132. Zhang, S. *et al.* Genetic and environmental contributions to phenotypic components of metabolic syndrome: a population-based twin study. *Obesity* **17**, 1581–1587 (2009).
133. Rahman, I. *et al.* Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res. Hum. Genet.* **12**, 286–294 (2009).
134. Pedersen, N. L., Gatz, M., Berg, S. & Johansson, B. How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Ann. Neurol.* **55**, 180–185 (2004).
135. Wirdefeldt, K., Gatz, M., Reynolds, C. A., Prescott, C. A. & Pedersen, N. L. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol. Aging* **32**, 1923–1928 (2011).
136. Mulder, E. J. *et al.* Genetic and environmental influences on migraine: a twin study across six countries. *Twin Res.* **6**, 422–431 (2003).
137. Hawkes, C. H. & MacGregor, A. J. Twin studies and the heritability of MS: a conclusion. *Mult. Scler.* **15**, 661–667 (2009).
138. Faraone, S. V. *et al.* Molecular genetics of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **57**, 1313–1323 (2005).
139. Lundstrom, S. *et al.* Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation. *Arch. Gen. Psychiatry* **69**, 46–52 (2012).
140. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
141. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
142. Peacock, M., Turner, C. H., Econs, M. J. & Foroud, T. Genetics of osteoporosis. *Endocr. Rev.* **23**, 303–326 (2002).
143. Spector, T. D. & MacGregor, A. J. Risk factors for osteoarthritis: genetics. *Osteoarthr. Cartil.* **12** (Suppl. 1), 39–44 (2004).
144. MacGregor, A. J. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
145. Thomsen, S. F., Van Der Sluis, S., Kyvik, K. O., Skytthe, A. & Backer, V. Estimates of asthma heritability in a large twin sample. *Clin. Exp. Allergy* **40**, 1054–1061 (2010).
146. Ingebrigtsen, T. S. *et al.* Genetic influences on pulmonary function: a large sample twin study. *Lung* **189**, 323–330 (2011).
147. Li, M. D., Cheng, R., Ma, J. Z. & Swan, G. E. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**, 23–31 (2003).

148. Agrawal, A. & Lynskey, M. T. Are there genetic influences on addiction: evidence from family, adoption and twin studies. *Addiction* **103**, 1069–1081 (2008).
149. Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D. & Ebers, G. C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl Acad. Sci. USA* **100**, 12877–12882 (2003).
150. Halfvarson, J. Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm. Bowel Dis.* **17**, 6–12 (2011).
151. Tanner, C. M. *et al.* Parkinson disease in twins. *JAMA* **281**, 341–346 (1999).
152. Cardno, A. G. *et al.* Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch. Gen. Psychiatry* **56**, 162–168 (1999).
153. Kendler, K. S. & Prescott, C. A. A population-based twin study of lifetime major depression in men and women. *Arch. Gen. Psychiatry* **56**, 39–44 (1999).
154. Levy, F., Hay, D. A., McStephen, M., Wood, C. & Waldman, I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *J. Am. Acad. Child Adolesc. Psychiatry* **36**, 737–744 (1997).
155. Rosenberg, R. E. *et al.* Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163**, 907–914 (2009).

Acknowledgements
This work was supported by the European Research Council (ERC 230374) and the Institute for Health and Care Research (EMGO+).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION

Dorret I. Boomsma's homepage: http://www.tweelingen-register.org
Australian Twin Registry: http://www.twins.org.au
Integrated Research on Developmental Determinants of Ageing and Longevity (IDEAL): http://www.ideal-ageing.eu
Netherlands Consortium for Healthy Ageing: http://www.healthy-ageing.nl
Queensland Twin Registry: http://www.qtwin.org.au

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF