Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2012;41:140–150
doi:10.1093/ije/dyr179

The value of twins in epigenetic epidemiology

Jordana T Bell[1,2] and Richard Saffery[3,4]*

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, [2]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK, [3]Cancer, Disease and Developmental Epigenetics, Murdoch Childrens Research Institute, Parkville, Victoria, Australia and [4]Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

*Corresponding author. Murdoch Childrens Research Institute, Flemington Road, Parkville, Victoria, Australia, 3052. E-mail: richard.saffery@mcri.edu.au

Accepted 19 October 2011

During past decades, twin studies have played an important role in genetic epidemiology studies of complex traits. The strength of twin studies lies in the ability to disentangle genetic and environmental factors that contribute to a phenotype, by comparing genetically identical monozygotic twins to dizygotic twins, who share on average 50% of genetic variants. Twin studies now offer the opportunity to study epigenetic variation across the genome with two aims. First, twin studies can improve our understanding of the factors regulating epigenetic variability by assessing the heritability of epigenetic variants. Secondly, the use of twins in epigenetic research is increasingly recognized as an important approach to help unravel the complexities associated with human development and disease. The strategic use of identical twins discordant for complex disease has revealed the importance of linking epigenetic disruption to the disease-associated risk in humans. Lastly, we also discuss the possibility that epigenetic effects on disease may in part explain some of the missing heritability in genome-wide association studies. The study of human epigenetic factors in twins can inform the role of genetics, as well as *in utero* and postnatal environments to the establishment, maintenance and functional consequences of human epigenome variation.

Keywords Epigenetics, twins, DNA methylation, dizygotic, monozygotic, imprinting, phenotypic discordance, heritability

The added value of epigenetic studies in twins

Recent studies have focused on functional genomic approaches in studying complex human disease, at the same time highlighting the importance of characterizing the regulation and dynamic nature of gene expression to help understand the underlying biological mechanisms. Epigenetic mechanisms are crucial regulators of gene expression and, of these, DNA methylation is the most widely studied in mammals and is the focus of this review. The strategic use of twins provides a unique opportunity to circumvent some of the complexities in examining the aetiology of complex disease. Recent DNA methylation studies in twins can improve our understanding of the processes involved in the regulation of epigenetic variation and help disentangle the relative contributions of epigenetics, environment and genetic variation in complex traits.

Twin studies have played an important role in our understanding of individual variation for over a century. The comparison between monozygotic (MZ) and dizygotic (DZ) twins allows an estimation of the heritability of traits, which describes the proportion of the total variance in a trait that is attributable to

genetic factors. Therefore, epigenetic studies in twins can give an insight into the genetic regulation of epigenetic variation. In addition, MZ twin concordance rates can vary across complex diseases. MZ concordance rates near 100% indicate that the trait is likely under genetic control of high penetrance. However, MZ concordance rates for many traits are often much lower than 100%, indicating a role of environmental and/or stochastic epigenetic factors in modulating phenotypes associated with underlying genotypes. Epigenetic mechanisms are widely believed to be the mediators of the influence of environmental factors on the underlying genome, and therefore the use of twins in epigenetic research is becoming increasingly popular.

Interestingly, gene expression studies in twins have accordingly identified not only evidence for heritable gene expression profiles,[1,2] but have also identified genes that show MZ gene expression discordance levels above experimental noise.[1,3] This relationship is even apparent in multiple tissues taken from newborn twin pairs, where MZ pairs can show clear discordance in gene expression profiles at birth, highlighting the importance of the *in utero* environment in determining overall gene expression (and most likely epigenetic) profile.[4]

These findings support the hypothesis that MZ phenotype discordance may be due to divergence in gene expression between MZ twins in response to environmental and stochastic factors. The molecular mechanisms underpinning gene expression and associated phenotypic discordance in MZ twins are likely epigenetic in nature. Therefore, epigenetic studies in disease-discordant MZ twins can identify genes sensitive to the environmental factors that are associated with complex disease. Here, we aim to provide an overview of the use of twin studies in epigenetic research, both for the purpose of improving our understanding of the underlying biology of epigenetic mechanisms through the estimation of epigenetic heritability, and in identifying disease-related epigenetic change.

Twin studies help unravel the regulation of epigenetic profiles

The epigenetic status of a specific genomic region is determined by the sum of genetic and cumulative environmental and stochastic factors. However, the relative contribution of these factors to the epigenetic profile and its stability over time remains poorly understood.

Heritability of epigenetic profiles

The main strength of the classical twin study is the ability to distinguish between the contribution of genetic and environmental effects that contribute to a phenotype of interest, while controlling for age. Twin studies can be applied to epigenetic variation in a similar vein, to dissect the contribution of genes and environment to the establishment of epigenetic profiles. Heritability is an estimate of the proportion of the total phenotypic variance in a population that is attributable to genetic effects. In twins, heritability estimates compare the degree of phenotypic similarities between groups of MZ and DZ twins using concordance rates or intra-class correlation. Twin-based studies can estimate the narrow-sense heritability (h^2), which measures the proportion of trait variance that is due to additive genetic effects. Twin narrow-sense heritability estimates can be calculated as twice the difference between MZ and DZ correlation, $h^2 = 2(r_{MZ} - r_{DZ})$, where r is the correlation coefficient for each twin pair type.[5] In the context of epigenetic heritability, classical twin studies can provide an estimate of the proportion of epigenetic variance at a single locus that can be attributed to genetic variation. In addition, the specific pattern of epigenetic twin-pair correlations at each locus can also inform the proportion of epigenetic variance that can be attributed to shared environment.[6] However, epigenetic heritability estimates will be not only population specific, but also cell, tissue, time and locus specific. These will also depend on the specific epigenetic mechanism and, to a certain extent, on the sensitivity, resolution and coverage of the epigenetic assay.

Previous studies have estimated heritability of DNA methylation profiles both at specific genes[7,8] and across the genome.[9] The results indicate that certain regions of the genome show strong evidence of being under genetic control whereas others do not.[10] These findings are further supported by evidence of meiotic transmission of other epigenetic marks such as chromatin states[11] and transcription-factor binding sites,[12] familial clustering of DNA methylation patterns at certain regions,[13] allele specificity in DNA methylation[14,15] and histone modifications,[16] and the identification of Single Nucleotide Polymorphisms (SNPs) that associate with DNA methylation at nearby genes (in *cis*).[17–19]

Epigenetic twin-based heritability estimates provide a measure of the locus-specific variance in epigenetic profiles that is attributed to genetic variants. However, the interpretation of the genetic component requires consideration for two reasons. First, there are no genome-wide studies to date that assess the extent to which twin-based epigenetic heritabilities correspond to the transmission of epigenetic marks across generations.[5,20] In animals (including humans), there are examples of locus-specific *trans*-generational epigenetic inheritance,[21,22] but there is also evidence that the fidelity of transmission of epigenetic changes is reduced with each generation.[23] Therefore, both twin-based and *trans*-generational family-based estimates of epigenetic heritabilities are necessary to understand the mechanisms underlying the genetic component of epigenetic variation at each locus.

Secondly, it is also feasible that DZ twins as a group exhibit more epigenetic differences than MZ twins because the former originate from different zygotes, potentially already epigenetically divergent. In contrast, MZ twins derive from a single zygotic epigenome.

In addition to informing the role of genetic components in epigenetic regulation, twin-based studies can also give an insight into the environmental factors underlying epigenetic variation. However, epigenetic twin studies require careful consideration of the factors that are attributed to common and unique environment, specifically with respect to intra-uterine environment. MZ twins share maternal, obstetric and genetic factors, but differences in epigenetic profile can accumulate in response to differences in environmental exposures and stochastic factors, accumulated both *in utero* and post-natally. In contrast, DZ pairs share many maternal and obstetric factors, but the likelihood of sharing specific genetic variations is the same as that of non-twin siblings. In terms of estimating common environmental effects in epigenetic heritability studies, a distinction could be made to compare DZ twins with specifically di-chorionic MZ twins, because MZ chorionicity has been associated with DNA methylation changes.[8,9] DNA methylation comparisons in MZ mono-chorionic (MZMC) and MZ di-chorionic (MZDC) twins have been conducted across multiple tissues.[8,9] Initial findings surprisingly indicated that MZMC had more dissimilar methylation profiles than MZDC in buccal tissue from 20 MZ twin pairs, using the Human 12K CpG island microarray,[24] which assays approximately 12 000 regions genome wide. However, a subsequent study in a larger sample of 56 MZ twins in multiple tissues[8] observed a mixture of effects of chorionicity on methylation and no consistent direction of effect across tissues. Specifically, MZMCs had more methylation differences than MZDCs in buccal tissue, granulocytes and human umbilical vein endothelial cells (HUVEC), but the opposite pattern was observed for cord blood mononuclear cells (CBMC) and placenta. These observations suggest that placenta sharing does not necessarily result in more similar methylation profiles across tissues. Another recent finding that may give an insight into the environmental factors influencing early-life epigenetic profiles examined the effect of assisted reproductive technology on the DNA methylation levels at three imprinted regions.[25] Although there were no great differences in methylation patterns between naturally conceived twins and twins who were conceived by in vitro fertilization (IVF), a slight increase in methylation variability was observed in the IVF group, suggesting that environmental perturbations may introduce variability in methylation at specific loci.

Several extensions to the classical twin model have been proposed for complex traits[6,26] and these can also be applied to epigenomic data. Multivariate analyses of twin data, in which multiple phenotypes are modelled simultaneously, have been applied to complex phenotypes to search for interactions between phenotypic variables, primarily to gain insight into phenotype causality, and to identify genetic or environmental factors that influence a combination of traits (pleiotropy). All of these approaches can now be applied to epigenetic data; for example, multivariate analyses of different types of epigenetic data, such as DNA methylation, histone modification and nucleosome occupancy sites, can allow inferences about the mechanisms that trigger the initial chromatin change as opposed to those that occur as a consequence of the primary change. Extended twin analyses can also assess the evidence of genotype interactions, for example epistasis, genotype by environment and genotype by gender interactions, which will be important in understanding the regulatory processing underlying epigenomic data.

Finally, age differences between groups of MZ and DZ twin pairs may be a potential confounding factor specific to epigenetic analyses[27,28] and should be considered in any analysis. Age-related effects on epigenetic processes can also be directly examined in twin-based heritability studies. Heritabilities can change over time;[6,29] for example, estimates of the heritability of telomere length vary according to the age of the sample.[30,31] Age-specific heritabilities of epigenetic processes may occur in response to several factors, including genetic effects that affect epigenetic profiles only at a certain developmental stage, changes in the environment that affect the variance contributed by environmental factors and potentially also affect genotype by environment interactions. Therefore, it is of great interest for future studies to specifically examine age-dependent epigenetic heritabilities.

Epigenetic changes over time

The cumulative effects of environmental and stochastic variation on changing epigenetic profile were first illustrated by a widely cited study that examined both genome-wide and locus-specific DNA methylation variation in a small number of young and middle-aged MZ twins.[28] Although 3-year-old MZ twins showed relatively few epigenetic differences within pairs, those aged 50 years showed considerable variability within pairs, and this was greater if the twins had divergent lifestyles. The statistical analyses included first generating locus-specific summary statistics for each individual and each data type (DNA methylation, histone acetylation, gene expression), followed by estimation of within twin-pair similarities of these summary statistics by calculating Euclidean squared distance within twin-pair. Further analyses compared the locus-specific epigenetic variables with the phenotype using mixed-effect models. The phenotype variables were obtained from principle component analysis applied to the original questionnaire data to reduce them to two uncorrelated components—'ageing' (encompassing age, weight and

height) and 'health' (all variables of disease and pharmacological treatments). The general conclusion was that the epigenetic profile is in constant 'drift', although very few young twins were studied, the resolution of the assays was low, and statistical precision was low.[28] In addition, this was a cross-sectional study and did not evaluate methylation in the same individuals over time, potentially obscuring the distinction of methylation differences that occur over time, to those difference that are present among unrelated individual at birth.

Two studies have directly examined epigenetic drift longitudinally in twins.[32,33] In the first,[32] analysis focused on buccal cell DNA methylation in three genes in 46 MZ and 45 DZ twin pairs at 5 and 10 years of age. Longitudinal change within individuals was calculated by assessing the correlation in methylation at each age. To assess the relative contributions of heritable and environmental/stochastic components to methylation levels, correlations within MZ pairs were compared with correlations within DZ pairs. The results indicated that methylation differences were present in MZ and DZ twins in early childhood and were, in general, not stable over time. In the second study,[33] longitudinal effects were examined at eight loci using two longitudinal blood (11–20 years longitudinal range) and buccal (2–8 years longitudinal range) samples from 34 individuals. Altogether, five of the eight loci were stable over time and across tissues. These studies were the first to reveal the extent of epigenetic discordance in MZ twins in early life and to highlight the instability of methylation levels over time. Importantly, different genomic regions were found to show varying levels of epigenetic divergence over time.

A recent examination of DNA methylation levels at several sites in multiple tissues from newborn twins confirmed that epigenetic drift between genetically identical individuals begins *in utero* and in a tissue-specific manner. In this study, intra-class correlation coefficients (ICCs) were higher in MZ than in DZ twins[8] (Figure 1), supporting previous findings of a role for genetic/heritable factors in the establishment of epigenetic profile.

Epigenetic drift over time has also been observed for singletons and related individuals who are not twins, although some of the differences may in part be attributed to different genetic background.[13] For example, many CpG sites across the genome were identified to change consistently with age in two separate studies.[34,35] Longitudinal changes in epigenetic marks are most likely to occur in a subset of genes.[33]

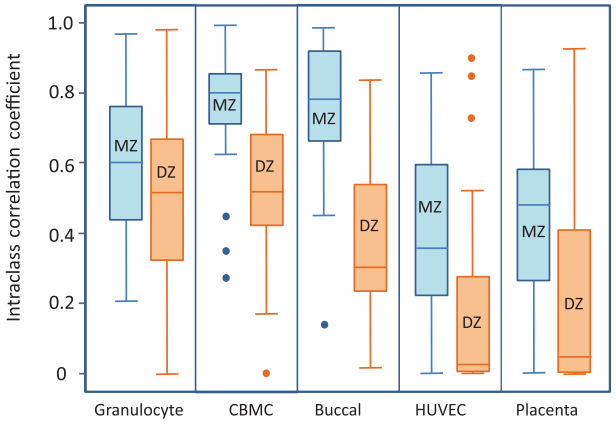

Figure 1 Distribution of intraclass correlations reveals a higher median correlation coefficient for DNA methylation at the IGF2/H19 locus in MZ than DZ twin pairs in five tissues from newborns. The ICC measures the proportion of total variance attributable to within-pair variation within MZ and DZ groups of twins. ICC analysis. Figure adapted from Ref.[8]

Twin studies of epigenetic effects in complex phenotypes

Experiments in genetically identical mice reared in near-identical environmental conditions demonstrate that most phenotypic variation is normally distributed.[36] One possible explanation of this finding is the possibility that epigenetic changes affect complex phenotype independent of genetic or environmental factors. Measurement of epigenetic differences in phenotypically discordant MZ pairs can provide a unique opportunity to identify genes sensitive to the environmental factors that are associated with complex disease.

The Discordant Twin Model

MZ genetic identity

Many studies have reported phenotypic differences between MZ co-twins,[37] and have revealed that non-genetic variation can contribute to disease penetrance and aetiology. The underlying assumption that MZ twins are genetically identical had been directly disproved in several instances: specific point mutations,[38–40] uniparental disomy,[41] triplet repeat expansion,[42,43] chromosomal mosaicism[44] or heteroplasmy for mitochondrial-encoded mutations[45,46] or chromosomal aneuploidies[37] have all been linked to specific phenotypic differences in MZ twins. Importantly, differences in both copy number[47] and telomere length[30,48] have also been described in phenotypically discordant and concordant MZ twin pairs. It is likely that these rare genetic differences, mosaicism, as well as non-genetic (epigenetic) variation within MZ twin pairs underpin the majority of observed phenotypic discordance.

Are epigenetic changes causal or secondary to the trait?

In genetic studies of complex traits, cause and effect can be distinguished because genomic variation is not considered to be variable over an individual's lifespan,

except in relation to 'rare' somatic mutation. However, since epigenetic changes within an individual can vary over time, most epigenetic studies of complex phenotypes would require repeated longitudinal samples from more than one tissue. Unless samples are collected early in life and prior to phenotypic manifestation, unravelling epigenetic cause and effect is problematic. This illustrates some of the caveats associated with inferring causation in any epigenetic association study that links specific epigenetic changes to disease phenotypes.

Skewed X-chromosome inactivation in phenotypically discordant MZ twins

The inactivation of the majority of genes on one X-chromosome in females is one of the most widely studied epigenetic phenomena in mammals. This mechanism of dosage compensation equalizes the expression of most X-chromosome genes in males (XY) and females (XX). It typically occurs in a random manner, with ~50% of cells in a female showing inactivation of the maternally derived X and the other 50% showing inactivation of the paternally derived X-chromosome. However, in some individuals a skewed pattern of X-inactivation is apparent, where one parental X in over-represented in a particular tissue or cell type. In some instances, a recessive mutation carried on the active X-chromosome can have profound adverse phenotypic effects. Thus, skewed X-inactivation has been associated with discordance for several disorders including haemophilia,[49] Fragile-X syndrome[43,50] and Duchenne Muscular Dystrophy.[51]

Genome-scale epigenetic investigations in twins and phenotypically discordant MZ twins

Several studies have attempted to identify epigenetic changes associated with complex human disorders by examining relatively low-resolution epigenetic variation at specific gene regions in disease discordant MZ twins (summarized in Table 1). Overall, the established effect sizes are small, with few examples of independent replication of the identified associations. To date, such studies have focused solely on DNA methylation, perhaps due to the availability of genome-wide assays of DNA methylation compared with other epigenetic marks.

Recently, other studies have started to examine the extent of DNA methylation changes across the genome that could potentially explain phenotypic discordance in MZ twins. Perhaps the most successful application of the discordant MZ twin model in epigenetics has been the investigation of DNA methylation changes in three immune related disorders: systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and dermatomyositis (DM).[52] The authors assayed genome-wide DNA methylation profiles at over 1500 CpG sites in five pairs of discordant MZ pairs for each disorder, validated the findings at specific genes in additional discordant twins and matched controls, and also measured overall 5-methylcytosine (5MeC) levels. Differentially methylated regions were calculated using *t*-tests and *P*-values were corrected for multiple testing using false discovery rates.[53] Candidate genes were selected for further validation if the mean methylation difference between affected and non-affected groups surpassed an arbitrary cut-off of 10%. The results highlighted 49 genes, enriched for immune functioning, that showed consistent changes in DNA methylation specifically in SLE, but not in RA or DM. Subsequent validation in matched controls and additional discordant twin pairs confirmed these results. Interestingly, global 5MeC levels showed an overall loss of DNA methylation in the SLE affected individuals, and this was found to be at least in part due to a specific decrease in methylation at repetitive 28S and 18S ribosomal RNA genes.

One the most compelling analysis of factors regulating overall epigenetic profiles was obtained through the high resolution analysis of DNA methylation in three collections of MZ and DZ twin pairs.[9] The samples consisted of white blood cells (WBCs) from 19 MZ twin pairs who were dichorionic and 20 DZ twin pairs matched for age, sex and blood count, and also buccal epithelial cells from 10 monochorionic MZ and 20 DZ age- and sex-matched pairs. Additional WBCs and buccal cells were obtained from each of 10 dichorionic MZ and 10 DZ pairs from another independent twin cohort, and gut biopsies were collected from 18 MZ pairs from another twin population. ICCs for MZ and DZ pairs were calculated for DNA methylation levels at over 6000 genomic regions using a DNA methylation microarray assay.[54] The ICCs clearly showed a higher degree of similarity between the MZ twins as a group, relative to the DZ group ($P = 1.2 \times 10^{-294}$), with some variability between tissues. However, significant epigenetic differences were also found in nearly identical mouse strains, leading the authors to speculate that a differential zygotic epigenetic profile in DZ versus MZ twins, rather than genetic factors, was the primary driver of the ICC differences.[9] Nevertheless, this hypothesis could not be directly tested because of the inability to directly measure the epigenetic profile in single cell zygotes.

The future focus of both genomic and epigenomic analysis of twins will fall on high throughput nucleotide sequencing, which would eventually allow for the detection of epigenetic variation at every site in every gene. Baranzini *et al.*[55] examined genetic, DNA methylation and gene expression profiles by next-generation sequencing in purified $CD4^+$ T cells in a small number (three) of MZ twin pairs discordant for multiple sclerosis (MS). The gene expression data consisted of 50–68 million sequencing reads of messenger RNAs in each twin pair. The analyses of the expression

Table 1 Examples of studies comparing epigenetic profiles in discordant twin pairs

Focus	Number of twin pairs	Biological samples	Epigenetic target/ Methodology	Analysis methodology	Major finding	Ref.
Candidate gene approaches						
Schizophrenia (SCZ)	1 discordant + 1 concordant affected	Lymphocytes	DNA methylation of *DRD2* gene by bisulphite sequencing	Purely observational	Discordant twin with schizophrenia shows more similar *DRD2* methylation profile to affected concordant twins than to unaffected co-twin	54
Birth Weight	12 highly discordant MZ pairs	Buccal cell DNA at age 5	DNA methylation at two CpG sites of *COMT* gene by bisulphite pyrosequencing	Not stated	Highly variable methylation concordance rates between pairs. Site-specific averages of within-pair methylation discordance of 10.3 and 16.1%, with significant correlation in differences between both sites ($r = 0.87$; $P < 0.001$). No correlation between birth weight and methylation level	55
Primary biliary cirrhosis (PBC)	4 discordant MZ, 1 concordant affected	Peripheral blood	DNA methylation by bisulphite sequencing	Not stated	Decreased expression of *CLIC2* and *PIN4* in ¾ PBC affected individuals in discordant pairs. No evidence for DNA methylation changes driving expression	56
Beckwith Wiedemann Syndrome (BWS)	10 discordant MZ, 5 control MZ	Skin fibroblasts and/or peripheral blood	DNA methylation by Southern blotting	Purely observational	BWS affected individuals show loss of imprinting at *KVDMR1* with biallelic expression of *KCNQ1OT1*	57
Silver-Russell Syndrome (SRS)	1 discordant MZ pair	Peripheral blood leukocytes	Combined bisulfite restriction analysis (COBRA) and bisulfite sequencing for two H19 DMRs	Purely observational	Loss of *H19*-DMR in roughly half of cells in the affected twin only with associated decreased expression of *IGF2*	58
Caudal Duplication Anomaly (CDA)	1 discordant MZ pair, 9 unaffected MZ pairs	Peripheral blood mononuclear cells	DNA methylation of *AXIN-1* by bisulphite sequencing	Binomial generalized linear mixed model	Higher methylation at *AXIN-1* promoter in affected vs unaffected co-twin ($P < 0.0001$). Higher methylation in both twins compared with controls ($P = 0.02$). ICC correlation of 0.76 for all 10 MZ pairs.	59
Alzheimer's disease (AD)	1 discordant MZ	Post-mortem temporal neocortex	Various epigenetic markers by immunohistochemistry	Two tailed t-test	Significantly reduced levels of DNA methylation in temporal neocortex neuronal nuclei in AD	60
Body Mass Index (BMI)	16 discordant MZ pairs	Saliva	DNA methylation at nine DMRs implicated in growth, by bisulphite based, amplification and primer extension and high performance liquid chromatography (HPLC). Methylation index calculated as h(C)/h(C) + h(T) where h = peak height on HPLC	Paired t-test and Wilcoxon's signed rank test for methylation–phenotype associations. Linear and rank correlation coefficients for associations of intrapair BMI and methylation differences	Only small intrapair differences in methylation observed. No significant correlations between intrapair BMI differences and intrapair methylation levels	61

(continued)

Table 1 Continued

Focus	Number of twin pairs	Biological samples	Epigenetic target/ Methodology	Analysis methodology	Major finding	Ref.
Genome-scale (hypothesis free) approaches						
Systemic Lupus Erythramatousis (SLE)	5 discordant for each disorder −17 SLE sib pairs including 1MZ, 4DZ for validation only	Whole blood: white cell fraction	DNA methylation with Illumina Golden Gate Bead Arrays (more than 1000 measurements)	Student's t-test with False Discovery Rate (FDR) correction for multiple testing	No methylation differences associated with DM and RA	39
Rheumatoid arthritis (RA)			Validation by bisulphite sequencing and pyrosequencing		Consistent methylation changes in SLE twins relative to unaffected co-twins in immune functioning genes	
Dermatomyositis (DM)			Global 5MeC measurement		Global decrease in 5-MeC in SLE and hypomethylation of 28S and 18S *rDNA* genes	
Multiple Sclerosis (MS)	3 discordant MZ pairs	$CD4^+$ T cells	Next Generation sequencing of genomic DNA, mRNA (50–68 million reads) and RRBS (50–90 million reads covering more than 2 million CpG sites)	RRBS sequence aligned using GSNAP.[62] No statistical testing used to determine significance (or otherwise) of observed within pair methylation differences	Between 2 and 176 significant methylation differences (of 2 million CpGs tested) between MZ co-twins (more than 800 methylation differences in unrelated individuals). No evidence for genetic, epigenetic or transcriptome differences underlying MS discordance	42
Bipolar disorder (BPD)	1 discordant MZ pair, 16 unrelated singletons with BPD. Independent set of 14 unrelated BPD singletons	Lymphoblastoid cell lines	MS-RDA with validation by bisulphite DNA sequencing and pyrosequencing	Mann-Whitney *U*-test for analysis of expression and methylation levels. One way Analysis of Variance (ANOVA) with cofactors of age and sex for effect of diagnosis controlled for confounders	10 genomic regions showing differential methylation between co-twins by MS-RDA. Four of these confirmed by bisulphite sequencing of co-twin DNA. Two of these (*PPIEL*, *SMS*) confirmed as significant by pyrosequencing in case-control study design (25 BPD cases and 18 controls) with $P < 0.05$. Methylation status of *PPIEL* inversely correlated with gene expression levels	63
Risk-taking behaviour	1 discordant plus 9 MZ control pairs	Peripheral blood	DNA methylation profiling at more than 12000 CpG island fragments	Multiple technical replicates. FDR correction for multiple testing. 1.15-fold ($\log_2(1.15) = 0.2$) used as experimental/technical variance threshold	No technical comparisons survived FDR correction. Differential methylation of *DLX1*[a] gene (3′UTR) identified specifically in discordant MZ pair ($P < 0.0004$)	43

[a]Transcription factor involved in the formation of GABAergic interneurons.[64]

3′UTR, non-protein coding 3′ untranslated region of RNA transcript; DMR, differentially methylated regions; DRD2, dopaminergic receptor D2; COMT, catechol-*O*-methyltransferase; RRBS, reduced representation bisulphite sequencing; MS-RDA, methylation-sensitive representational difference analysis; GSNAP[62]; PPIEL, peptidylprolyl isomerise E-like; SMS, spermine synthase.

twin-pair data indicated that 57.3% of the variance in gene expression was attributable to between pair differences, 26.3% to unique individual differences and, surprisingly, only 9.4% was attributable to a diagnosis of MS. These data highlight the dynamic nature of gene expression and the challenges in trying to identify disease-related gene expression changes. Despite some evidence for allelic imbalance within discordant twin pairs, no such changes were common across all three MS discordant twin pairs examined. The DNA methylation data consisted of 50–90 million high-quality reads of reduced representation bisulphite sequencing (RRBS) from $CD4^+$ T cells of the three discordant twins, and also showed little evidence of DNA methylation changes that could be specifically associated with the MS phenotype. Although several hundred locus-specific differences were observed between males and females, or between primary and cancerous tissues, the magnitude of epigenetic differences between MZ twins discordant for MS was minimal. However, the sample was relatively small, and consisted of a heterogeneous mix of males, females, Europeans and African Americans. Therefore, although the findings could represent true negative results or disease heterogeneity, it is also possible that the sample size was too small to provide statistical power to detect significant epigenetic differences. Unfortunately, there was no independent validation of the observed epigenetic differences using an alternative methodology, so the robustness of the few observed DNA methylation changes within MZ pairs remains unclear.[55]

Power to detect epigenetic changes in disease-discordant MZ twins

Many factors affect the power of the discordant MZ twin design to detect disease-related differentially methylated regions (DMRs) or differentially methylated probes (DMPs). These include the effect size of the epigenetic variant on disease, the concordance of methylation levels between MZ twins at the locus of interest, the sample size, and the sensitivity, coverage and reproducibility of the methylation assay. Kaminsky *et al*.[56] estimated the power of the discordant MZ twin design for a specific microarray methylation assay.[54] They found that a relatively small number (15–25) of phenotypically discordant twin pairs had sufficient (>80%) power to detect epigenetic changes of 1.2-fold, where an effect size of 1.2-fold change was significantly greater than the null experimental variance threshold for the assay (1.15-fold change). The study also examined the reproducibility and number of technical replicates and as expected, found that increasing the number of technical replicates reduced the technical variance and increased power to detect epigenetic changes. Therefore, limited attempts at power calculations are so far encouraging for the application of genome-scale methylation analysis to large twin-based studies. However, these power estimates do not necessarily apply to other genome-wide methylation assays, which will differ in sensitivity, specificity and coverage. Therefore, the power to detect epigenetic changes in discordant MZ twins needs to be re-evaluated in the context of specific epigenetic assays and in particular for next-generation sequencing technologies.

Epigenetic contribution to phenotype heritability

The missing heritability from genome-wide association studies (GWAS) refers to the observation that most variants identified from GWAS have relatively small effects and explain only a small proportion of familial clustering, leading to the question of how the remaining, 'missing' heritability can be explained.[57,58] Several explanations have been suggested, including rare genetic variants, larger numbers of variants of smaller effects, structural variants, epistasis (gene–gene interaction) and shared environment among relatives. The contribution of epigenetic variation to complex phenotypes (Figure 2) may also in part explain the missing heritability from GWAS.

To examine the contribution of epigenetic effects to phenotypic heritability, a recent study proposed a mathematical model to incorporate genetic, epigenetic and phenotypic variation, and estimate the joint contribution of genetic and epigenetic effects to complex trait susceptibility.[59] The results suggested that while epigenetic variants may add to individual disease risk, they may not necessarily contribute to heritability unless methylation is faithfully transmitted during meiosis. However, more data are needed to test the underlying assumptions in the model, such as the stability of methylation transmission during meiosis, the effect of methylation variants on disease, and the suitability of the multiplicative model of disease risk. To this end, *trans*-generational studies of epigenetic variation in twins can inform the extent of transmission of epigenetic patterns during meiosis, for example, by comparing epigenetic MZ twin concordance profiles with age-controlled epigenetic concordance among relatives of different degree of relatedness across generations.

The possibility that environmentally induced epigenetic changes may be transmitted across generations could theoretically account for a significant proportion of the missing phenotypic heritability by impact on disease penetrance in subsequent generations, and also disease heritability rates. Studies have shown that parental diet, including maternal diet during pregnancy, can impact locus-specific DNA methylation levels and phenotypes in the next generation,[60–64] but as yet no-one has looked at the transmission of environmentally induced epigenetic changes across multiple generations in humans.

A common explanation of the missing heritability from GWAS is low power to detect epistasis, or the interaction between genes. In the context of

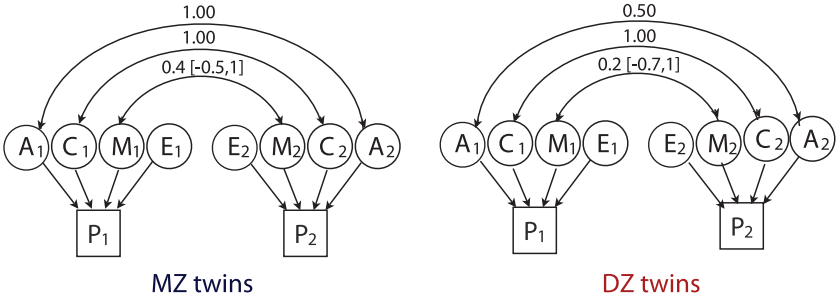

Figure 2 The contribution of DNA methylation to complex phenotypes. Proposed contribution of latent variables, including DNA methylation, to the phenotype of an individual. The figure represents a path model in twins, depicting the contribution of DNA methylation (M) and other factors to the phenotype (P) in twin 1 with correlation estimates in MZ (left) and DZ (right) of latent variables including additive genetic effects. (A_1), common environment (C_1), DNA methylation, (M_1) and unique environment (E_1). Correlation estimates were obtained from previous genetic[65] and epigenetic studies[9] in twins. In siblings, the correlation in M will be lower than that observed in DZ twins due to age differences and associated increased level of cumulative stochastic change. The latent variables depicting DNA methylation status at a genomic region (M), will have resulting phenotypic effects that will be specific to age, sex, population (genetic factors) and the tissue sampled. Figure adapted from Ref.[5]

epigenetics, extensions of epistasis would include the interaction between genetic and epigenetic variants—not only in the context of epigenetic quantitative trait loci (QTLs) but also in terms of genetic–epigenetic effects on disease risk, and interactions between pairs of epigenetic variants. For example, Birney[62] describes two studies that are consistent with such a genetic–epigenetic interaction, one in sheep,[63] and one in the human *FTO* gene in diabetes,[66] and proposes that future analyses should include integrated genetic and epigenome-wide association scans (GWAS/EWAS). Performing studies of genetic–epigenetic interactions in MZ twins would help detect environmentally induced or stochastic epigenetic changes that contribute to the phenotype only via interactions with genetic factors. Such studies may shed light on the interaction between genotype and environment and on the mechanisms underlying incomplete penetrance in complex genetic traits.

Conclusions

Epigenetic studies of twins offer the advantage of allowing both a better understanding of the factors regulating epigenetic profiles, and of the role that epigenetic effects have in complex traits. However, there are two main caveats of studying epigenetic variation exclusively in twins. First, there may be epigenetic effects specific to twins that limit their utility as a model to reveal contributors to disease in a broader context. For example, recent data implicate epigenetic mechanisms in the embryo splitting that leads to MZ twins.[67] If epigenetic variants are specific to MZ twins, then MZ vs DZ comparisons aimed at identifying heritable factors may be confounded. These findings may also impact the correlations between epigenetics and MZ phenotypic discordance. Secondly, due to epigenetic profile changes over time, it is difficult to infer causality of any disease-associated epigenetic variants. Furthermore, these time-related changes would result in increasing population variance in epigenetic levels at age-sensitive loci. It is therefore important to use age-matched twin pair where possible for identification of epigenetic biomarkers specifically associated with disease discordance. Despite this potential problem, limited attempts at power calculations are encouraging for large-scale genome-wide methylation analyses in twin studies.

It is highly plausible that over the next decade a full characterization of human genomic, epigenomic and transcriptomic data will be within the reach of most researchers. In the case of epigenomic data, these will be contingent on the provision of multiple tissues from willing donors, ideally sampled at distinct time-points during the donor's lifespan. In many cases, such as with brain tissues, sample collection will need to be done post-mortem. It is likely that the level of complexity and variation revealed within an individual's cells/tissues/organs, as well as that between individuals, will be immense. Such time-dependent epigenetic complexity may hinder the identification of biologically meaningful differences in epigenetic profile that initial studies propose to undertake. Despite this, the capacity to control for genetic variation using MZ twin pairs should prove valuable in helping to unravel the complexity of gene–environment–epigenetic interactions and their contribution to human health and disease.

Funding

National Health and Medical Research Council (Australia) and the Victorian Government's Operational Infrastructure Support Program (to R.E.S.).

Acknowledgements

We would like to thank the anonymous reviewers for their comments. J.T.B. is a Sir Henry Wellcome post-doctoral fellow.

Conflict of interest: None declared.

References

[1] Hasler R, Begun A, Freitag-Wolf S *et al*. Genetic control of global gene expression levels in the intestinal mucosa: a human twin study. *Physiol Genomics* 2009;**38:**73–79.
[2] Tan Q, Christensen K, Christiansen L *et al*. Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Hum Genet* 2005;**117:** 267–74.
[3] Sharma K. Genetic epidemiology of epilepsy: a twin study. *Neurol India* 2005;**53:**93–98.
[4] Gordon L, Joo JH, Andronikos R *et al*. Clues to the effect of intrauterine milieu on gene expression from a study of two tissues from neonatal monozygotic twins. *Epigenetics* 2011;**6:**579–92.
[5] Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet* 2011;**27:**116–25.
[6] Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev* 2002;**3:**872–82.
[7] Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum MCol Genet* 2007;**16:**547–54.
[8] Ollikainen M, Smith KR, Joo EJ *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* 2010;**19:** 4176–88.
[9] Kaminsky ZA, Tang T, Wang SC *et al*. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 2009;**41:**240–45.
[10] Gervin K, Hammero M, Akselsen HE *et al*. Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res* 2011;**21:**1813–21.
[11] McDaniell R, Lee BK, Song L *et al*. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* 2010;**328:**235–39.
[12] Kasowski M, Grubert F, Heffelfinger C *et al*. Variation in transcription factor binding among humans. *Science* 2010; **328:**232–25.
[13] Bjornsson HT, Sigurdsson MI, Fallin MD *et al*. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008;**299:**2877–83.
[14] Kerkel K, Spadola A, Yuan E *et al*. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 2008;**40:**904–08.
[15] Schalkwyk LC, Meaburn EL, Smith R *et al*. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 2010;**86:**196–212.
[16] Kadota M, Yang HH, Hu N *et al*. Allele-specific chromatin immunoprecipitation studies show genetic influence on chromatin state in human genome. *PLoS Genet* 2007; **3:**e81.
[17] Zhang D, Cheng L, Badner JA *et al*. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* 2010;**86:**411–19.
[18] Gibbs JR, van der Brug MP, Hernandez DG *et al*. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010;**6:**e1000952.
[19] Bell JT, Pai AA, Pickrell JK *et al*. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 2011;**12:**R10.
[20] Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 2010;**465:** 721–27.
[21] Rakyan VK, Chong S, Champ ME *et al*. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci USA* 2003;**100:**2538–43.
[22] Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004;**36:**497–501.
[23] Johannes F, Colot V, Jansen RC. Epigenome dynamics: a quantitative genetics perspective. *Nat Rev* 2008;**9:** 883–90.
[24] Heisler LE, Torti D, Boutros PC *et al*. CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. *Nucleic Acids Res* 2005;**33:**2952–61.
[25] Li L, Wang L, Le F *et al*. Evaluation of DNA methylation status at differentially methylated regions in IVF-conceived newborn twins. *Fertil Steril* 2011;**95:** 1975–79.
[26] MacGregor AJ, Snieder H, Schork NJ, Spector TD. Twins. Novel uses to study complex traits and genetic diseases. *Trends Genet* 2000;**16:**131–34.
[27] Fuke C, Shimabukuro M, Petronis A *et al*. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet* 2004;**68**(Pt 3):196–204.
[28] Fraga MF, Ballestar E, Paz MF *et al*. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 2005;**102:**10604–09.
[29] Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. *Nat Rev* 2008;**9:** 255–66.
[30] Slagboom PE, Droog S, Boomsma DI. Genetic determination of telomere size in humans: a twin study of three age groups. *Am J Hum Genet* 1994;**55:**876–82.
[31] Andrew T, Aviv A, Falchi M *et al*. Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *Am J Hum Genet* 2006;**78:**480–86.
[32] Wong CC, Caspi A, Williams B *et al*. A longitudinal study of epigenetic variation in twins. *Epigenetics* 2010;**5:**516–26.
[33] Talens RP, Boomsma DI, Tobi EW *et al*. Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J* 2010;**24:** 3135–44.
[34] Rakyan VK, Down TA, Maslau S *et al*. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 2010;**20:** 434–39.
[35] Teschendorff AE, Menon U, Gentry-Maharaj A *et al*. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 2010;**20:**440–46.

[36] Gartner K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab Anim* 1990;**24:**71–77.
[37] Zwijnenburg PJ, Meijers-Heijboer H, Boomsma DI. Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet* 2010;**153B:**1134–49.
[38] Sakuntabhai A, Ruiz-Perez V, Carter S *et al*. Mutations in ATP2A2, encoding a Ca2+ pump, cause Darier disease. *Nat Genet* 1999;**21:**271–77.
[39] Robertson SP, Thompson S, Morgan T *et al*. Postzygotic mutation and germline mosaicism in the otopalatodigital syndrome spectrum disorders. *Eur J Hum Genet* 2006;**14:** 549–54.
[40] Kondo S, Schutte BC, Richardson RJ *et al*. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nat Genet* 2002;**32:**285–89.
[41] Smith AC, Rubin T, Shuman C *et al*. New chromosome 11p15 epigenotypes identified in male monozygotic twins with Beckwith-Wiedemann syndrome. *Cytogenet Genome Res* 2006;**113:**313–17.
[42] Helderman-van den Enden AT, Maaswinkel-Mooij PD, Hoogendoorn E *et al*. Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities. *J Med Genet* 1999;**36:**253–57.
[43] Kruyer H, Mila M, Glover G, Carbonell P, Ballesta F, Estivill X. Fragile X syndrome and the (CGG)n mutation: two families with discordant MZ twins. *Am J Hum Genet* 1994;**54:**437–42.
[44] Bonilla E, Younger DS, Chang HW *et al*. Partial dystrophin deficiency in monozygous twin carriers of the Duchenne gene discordant for clinical myopathy. *Neurology* 1990;**40:** 1267–70.
[45] Biousse V, Brown MD, Newman NJ *et al*. De novo 14484 mitochondrial DNA mutation in monozygotic twins discordant for Leber's hereditary optic neuropathy. *Neurology* 1997;**49:**1136–38.
[46] Blakely EL, He L, Taylor RW *et al*. Mitochondrial DNA deletion in 'identical' twin brothers. *J Med Genet* 2004; **41:**e19.
[47] Bruder CE, Piotrowski A, Gijsbers AA *et al*. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 2008;**82:**763–71.
[48] Bakaysa SL, Mucci LA, Slagboom PE *et al*. Telomere length predicts survival independent of genetic influences. *Aging Cell* 2007;**6:**769–74.
[49] Bennett CM, Boye E, Neufeld EJ. Female monozygotic twins discordant for hemophilia A due to nonrandom X-chromosome inactivation. *Am J Hematol* 2008;**83:** 778–80.
[50] Willemsen R, Olmer R, De Diego Otero Y, Oostra BA. Twin sisters, monozygotic with the fragile X mutation, but with a different phenotype. *J Med Genet* 2000;**37:**603–04.
[51] Zneimer SM, Schneider NR, Richards CS. In situ hybridization shows direct evidence of skewed X inactivation in one of monozygotic twin females manifesting Duchenne muscular dystrophy. *Am J Med Genet* 1993;**45:**601–05.
[52] Javierre BM, Fernandez AF, Richter J *et al*. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 2010;**20:**170–09.
[53] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc* 1995;**Series B:**289–300.
[54] Schumacher A, Kapranov P, Kaminsky Z *et al*. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res* 2006;**34:**528–42.
[55] Baranzini SE, Mudge J, van Velkinburgh JC *et al*. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 2010;**464:** 1351–56.
[56] Kaminsky Z, Petronis A, Wang SC *et al*. Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res Hum Genet* 2008;**11:**1–11.
[57] Manolio TA, Collins FS, Cox NJ *et al*. Finding the missing heritability of complex diseases. *Nature* 2009;**461:**747–53.
[58] McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008;**17:**R156–65.
[59] Slatkin M. Epigenetic inheritance and the missing heritability problem. *Genetics* 2009;**182:**845–50.
[60] Ng SF, Lin RC, Laybutt DR, Barres R, Owens JA, Morris MJ. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* 2010;**467:**963–66.
[61] Heijmans BT, Tobi EW, Stein AD *et al*. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA* 2008;**105:** 17046–49.
[62] Godfrey KM, Sheppard A, Gluckman PD *et al*. Epigenetic Gene Promoter Methylation at Birth Is Associated With Child's Later Adiposity. *Diabetes* 2011;**60:**1528–34.
[63] Birney E. Chromatin and heritability: how epigenetic studies can complement genetic approaches. *Trends Genet* 2011;**27:**172–76.
[64] Cockett NE, Jackson SP, Shay TL *et al*. Polar overdominance at the ovine callipyge locus. *Science* 1996;**273:**236–8.
[65] Spector TD. Advances in twin and sib-pair analysis. *Greenwich Medical Media* 2000.
[66] Bell CG, Finer S, Lindgren CM *et al*. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PloS One* 2010;**5:**e14040.
[67] Shur N. The genetics of twinning: from splitting eggs to breaking paradigms. *Am J Med Genet C Semin Med Genet* 2009;**151C:**105–09.