Twin discordance and disease: not just an environmental cause?

Witold Czyz (MD) 1,2, Julia M Morahan (PhD)1,2, George Ebers (MD) 1,2, Sreeram V Ramagopalan (DPhil) 1,2,3,4,5

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2 Nuffield Department of Clinical Neurosciences (Clinical Neurology), University of Oxford, Oxford, United Kingdom; 3 Blizard Institute, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, London, United Kingdom; 4 London School of Hygiene and Tropical Medicine, London, United Kingdom

**Correspondence should be addressed to:**

Dr Sreeram V Ramagopalan, Blizard Institute,
Queen Mary University of London, Barts and The London School of Medicine and Dentistry
4 Newark Street, London, E1 2AT, United Kingdom, Tel: 020 7377 7472, Fax: 020 7377 7033

E-mail: s.ramagopalan@qmul.ac.uk
Abstract

Monozygotic (MZ) twins have been used for decades in genetic-epidemiological studies to tease out the relative contributions of genes and the environment to a trait. Phenotypic discordance in MZ twins has traditionally been ascribed to non-shared environmental factors acting after birth, however recent data shows this explanation as being far too simple. We review here other reasons for discordance, including differences in the \textit{in utero} environment, mosaicism and epigenetics. Epigenetic differences are gaining increasing recognition. While it is clear that in specific cases epigenetic alterations provide a causal factor in disease aetiology, the overall significance of epigenetics in twin discordance as well as its dependence on environmental and genetic factors remains unclear. Epigenomic profiling studies have recently shed more light on the dynamics of temporal methylation change and methylome heritability, yet have not given a definite answer on the relevance to disease due to limitations in establishing causality. This review explores the subject of epigenetics as another component in human phenotypic variability focusing on evidence from MZ twin studies.
**Phenotypic Variability and Discordance**

The extent to which phenotypic traits are heritable has been a subject of scientific interest at least since Galton’s classic twin study design[1]. Twins offer a unique means to study inheritance. Monozygotic (MZ) twins arise from a single zygote and have always been thought to inherit identical genomic sequences[1]. Dizygotic (DZ) twins arise from two different zygotes and, just like siblings, inherit on average 50% of identical genomic sequence. To assess the relative contribution of genes to a trait, comparisons are made between MZ and DZ twin concordance, with a greater MZ than DZ concordance rate implicating a role for genetics in determining the trait[1]. Phenotypic discordance between MZ twins has traditionally been ascribed to non-shared environmental exposures. However, recent research highlights this as being a too simplified explanation. In this review, we focus on potential sources of MZ discordance.

**Environment and Discordance**

There are a range of early environmental factors that need to be considered as potentially explaining MZ twin discordance. Twinning itself is thought to be a rare malformation and a stochastic event, although there exists evidence for familiality [1-3]. MZ twins occur in about 3.5 in 1000 pregnancies or 4 in 1000 live births[4, 5]. Depending on the time of zygote splitting, MZ twins can be divided into four groups[1]. If the zygote splits within 3 days, the twins are dichorionic and diamniotic (DC DA) (18-36% of all MZ births). If splitting occurs after that time but before the 7th day, the twins are monochorionic but diamniotic (60-80% of cases)[1, 6]. If division occurs between days 7 and 14, twins are monochorionic and monoamniotic (MC MA). This type of twining accounts for 2-4% of all MZ twins[6]. Conjoined twins arise when splitting happens after the 13th or 14th day[1, 6]. All multifoetal pregnancies are more prone to complications (like foetal malnutrition and growth restriction, premature birth), with six times higher mortality rate than in singletons and a shorter average duration of twin pregnancy (35 weeks)[2, 7-15]. Intrauterine growth restriction (IUGR) is a common issue in twin pregnancies affecting 12-47% of all twin pairs[16]. It often leads to discordance in birth weight[16-18] and has been linked with discordance for a range of phenotypes like height, head circumference, intelligence, language comprehension and
expression, fine motor performance, balance, coordination, and visual-motor perception [16, 18, 19].

The potential causes of IUGR include genetic predisposition, in-utero crowding, uneven allocation of blastomere and uneven blood supply, as well as placental dysfunction (complications like abruptio placentae, infarcts, stem vessel thrombosis, velamentous insertion of the cord, and single umbilical artery)[16, 17, 19-21]. IUGR is even more pronounced in MC twins where differences in placental sharing and vascularisation lead to occasional unequal blood and nutrient sharing, and, in about 15% MC diamniotic pregnancies, result in twin-to-twin transfusion syndrome (TTTS)[4, 16, 19]. MC twins have a higher incidence of congenital heart disorders and TTTS increases this risk even further[4]. However, even in the absence of twin to twin transfusion, MC twins are seven times more likely to develop congenital heart disease, which usually manifests in one twin only[3-5, 11, 19, 22]. The higher risk nature of multiple pregnancies, their proclivity towards complications and twin-twin competition for maternal resources should increase the probability of a skewed environment affecting the twins in utero[23, 24].

After birth, any non-shared environmental exposure including factors such as diet, smoking, toxin exposure and infection could possibly contribute towards twin discordance [20, 25-31]. Moreover, early phenotypic differences arising in twins could potentially cause shared exposures to have different effects leading to dissimilarity between twins.

**De Novo Mutations and Genetic Mosaicism**

It is always assumed that MZ twins are genetically identical, but a wealth of data is accumulating to show that this is not necessarily the case. Mosaicism for de novo mutations, retro-transpositions, indels, duplications and chromosomal rearrangements may play a role in MZ twin discordance[32-44]. The rate for de novo base substitutions has been estimated at about $10^{-8}$ per base pair per generation making some genetic differences between adult twins likely[45]. Postzygotic mutations have been linked with discordance for diseases such as oral-facial-dygital syndrome type 1 or Joubert syndrome[46-49]. Postzygotic point mutations have been found to be the source of MZ twin discordance in Van der Woude syndrome, Darier’s Disease and neurofibromatosis type 1 while mosaicism for chromosomal
abnormalities has been implicated in MZ discordance for conditions like Turner syndrome, trisomy 21, trisomy 13, skin pigmentation and sex phenotypes[50, 51]. Postzygotic karyotypic mosaicism caused by faulty mitotic division has also been reported in cases of Ullrich-Turner syndrome[20].

Copy number variants (CNVs) accounting for a major portion of the genome, are highly polymorphic and relatively unstable, with mutation rates 100 to 10 000 times higher than single base substitutions[52]. Phenotypic discordance in MZ twins may in part be caused by de novo mutations of CNVs and CNV mosaicism[32, 34, 36, 53]. Indeed, it has been indicated that de novo CNVs may occur at a rate of 10% per twinning event, however so far studies failed to link CNV mosaicism to any specific case of phenotypic discordance in MZ twins[40, 54-56].

Additionally, unequal exchange of cells during gestation might potentially lead to discordant fetomaternal microchimerism[20].

**Developmental Noise and Stochasticity**

Some variation is inevitable as a result of transcriptional/translational stochasticity entailed by the random movements of molecules and the complexity of their interactions [57-63]. It should be expected that such noise can lead to markedly different effects in identical environmental conditions[64]. The effect of developmental stochasticity might amass in a drift-like fashion and thus be more relevant to discordance in complex polygenic traits like height or weight, which develop over long periods of time[58]. Stochastic events like unequal division of the inner mass cells during twining, unequal allocation of the developmental markers or precursor cells to different somatic lineages have been reported as potential sources of discordance in MZ twins[3, 37]. Slight differences in eye or hair colour as well as fingerprint profiles, cases of mirror twinning (affecting up to 25% of MZ twins) ranging from occipital hair whorls and handedness to situs inversus as well as major malformations could constitute examples of stochastic developmental discordance. Alternatively, they could arise through gene-in utero environment interactions[20]. Stochasticity could lead to differential results from shared environmental exposures in twins.
Certain cases of differential allelic expression (DAE), including random monoallelic gene expression, could result in random mosaic expression patterns[65, 66]. These are primarily associated with X-inactivation, however the phenomenon also affects autosomal genes, primarily those with olfactory, pheromone receptor as well as immune functions that undergo allelic exclusion, and can constitute a mechanism for stochastically-driven phenotypic discordance in MZ twins[20, 67-70]. A study by Cheung et al. estimated that about 50% of heterozygous loci are subject to DAE within MZ twin pairs[71]. These inter-allelic differences in expression were similar for at least 30% of heterozygous loci, indicating some genetic control but making stochastic gene expression differences likely [71]. In contrast, a comprehensive whole genome expression experiment conducted by Baranzini et al. (2010) produced different results[72]. Their findings, based on a single pair, indicated that only 1.9% of heterozygous coding loci showed significant evidence for DAE, but out of these, 57% were concordant within between the co-twins, still leaving room for stochastic effects.

Epigenetics

Epigenetics was initially a term coined by developmental biologists with no immediate link to the issues of epidemiology and heredity [73, 74]. Rather, the term rather described the way in which gene-environment and gene-gene interactions shape a phenotype during development. The concept was developed as an argument for a complex relation between genes and a phenotype. Today, epigenetics is used to describe alterations in genomic function, mainly mitotically heritable changes in gene expression that occur through chemical modifications to the structure of chromatin without altering the DNA sequence[25, 26, 73-76]. There is evidence for transgenerational inheritance of epigenetic changes but the scope and mechanisms are under study.

Cytosine methylation is one of the most well studied epigenetic alterations found in vertebrates[25, 73-76]. It occurs at approximately 4-6% genomic cytosines, depending on the cell type[77]. DNA methylation typically occurs in the context of CpG dinucleotides, although this depends on cell type. In foetal fibroblasts, 99.98% of methylated cytosines are located in CpG dinucleotides[77]. In contrast, for embryonic stem cells the proportion reaches about 75%, which highlights the importance of non-CG methylation for gene expression in pluripotent stem cells. Cytosine methylation is mediated by a family of proteins
called DNA methyltransferases[73, 75, 78]. This form of modification is generally associated with transcriptional inactivation. It both physically prevents transcription factors from binding to the DNA, and can also recruit additional factors like methyl-CpG binding domain proteins, which can promote repressive histone modifications[25, 73, 75, 76]. Some CpGs however, remain unmethylated[73]. The precise mechanism by which the differential methylation of CpG dinucleotides occurs, is unknown[73]. Recently also hydroxymethylation of cytosines has been discovered, however its biological significance is not yet known.

The second category of epigenetic modifications characteristic of all eukaryotes are covalent alterations to histone proteins[25, 73-76]. These affect the N-terminal histone tails and, depending on the position and the type of alteration, can either repress or promote active chromatin conformation[73, 76]. The two most important types are acetylation and methylation[76]. The various types of histone modifications led to the hypothesis of an epigenetic histone code that moderates transcription in response to developmental cues and the environment[76]. Cytosine methylation, histone modifications and other types of chromatin remodeling all act together in concert, and either reinforce or disable each other through feedback loops[73].

There is substantial evidence in support of epigenetic components in defining human phenotypic variation. Most of it comes from the studies of defects in genetic imprinting – an asymmetric, sex-dependent epigenetic moderation of paternal versus maternal gene expression that manifests in monoallelic expression[75]. For imprinted genes, epigenetic modifications are set in the germline each generation anew, according to sex[75]. Occasional abnormal cytosine methylation can result in epimutations, which, just like DNA mutations, can deactivate or cause both copies of the imprinted gene to be transcriptionally active[75]. Some epimutations have a direct genetic cause and are secondary to a DNA mutation in cis or in trans. Other epimutations are primary with no sequence alteration [75, 79] (Figure 1). This latter category could be further divided into stochastic epimutations, caused by inherently error-susceptible molecular machinery, and epimutations which are environmental in origin. The ratio between primary and secondary epimutations are not well defined. There might also exist an intermediate type of “facilitated epimutations” whereby the likelihood of a stochastic epimutation at certain locus is increased by genetic determinants[79]. Theoretically any gene can be targeted by epimutation giving rise to abnormal conditions, however, imprinted genes represent here a category which is more sensitive on account of their monoallelic expression.
If differences in DNA methylation can be sufficient to cross the boundary between normal and disease phenotype in imprinted, monoallelically expressed genes, then it is also probable that epimutations might affect expression of any pair of genes. Of course, with genes which are expressed in a normal biallelic manner, the effects would presumably be less severe. Consequently, as the complexity and polygenicity of a trait increases, the effect of an epimutation would most likely decrease. But each potential target for an epimutation would increase the variability of that trait, and since epimutations may arise during development, it follows that organisms can be subjected to epigenetic variegation and ought to be heterogeneous for their epigenomes[42, 80]. Thus epigenetic mosaicism between twins could be yet another source of phenotypic variation.

The source as well as significance of epimutations in twin phenotypic variability and discordance is currently a subject of debate. Stochastic epimutations (primary or secondary) offer a logical explanation for the variation which is both non-heritable and non-environmental[26, 72, 73, 81]. Their random character is entailed by de novo faults in the maintenance of DNA methylation, whose fidelity is estimated to be at the level of 97-99.9% in mammalian tissue culture with 3-5% de novo methylation per mitosis[26, 73]. Such an unstable methylome could account for phenotypic variability, although it has to be stressed that epimutation rates are lower in cells in vivo[25, 35, 80].

Studies which employ methylome profiling, locally or globally, offer a direct method of evaluating the specific contribution of epigenetics to a phenotype. Some of the main questions concern the nature of this contribution, namely the ratio between hereditary, environmentally-triggered and stochastic changes. The extent of epigenetic changes and epigenome heritability is disputable. A thorough cross-sectional study of epigenetic profiles in the lymphocytes of 80 MZ twins, aged between 3 and 74, revealed a trend of steadily accumulating changes to the epigenome with age [29]. By estimating total genomic 5-methyl cytosine content and histone H3 and H4 acetylation, 65% of the twins had almost identical epigenomes while the remaining 35% were found to be variably discordant. Both the histone acetylation and DNA methylation profiles become progressively discordant with age or different lifestyle and medical history. More importantly, the same pattern of epigenetic discordance was observed with buccal epithelial, intra-abdominal fat and skeletal muscle...
The authors proposed that the epigenome is highly heritable at birth, but epimutations arise and accumulate throughout a lifetime and their origin may be both stochastic (a process called “epigenetic drift”) and environmental. In a similar study, Kaminsky et al (2009) used CpG island microarrays to screen about 6000 loci (as compared to 1800 loci investigated by Fraga et al), in a cohort of 114 MZ and 80 DZ twins, in search for methylation differences[82]. Some discordance was demonstrated in white blood cells and replicated in buccal and gut tissue. Estimates based on 20 MZ and 20 DZ pairs indicated that methylation heritability was very low (0.014) in white blood cells, but rose to about 0.3 in buccal tissue (findings based on 19 MZ and 20 DZ pairs), and up to 0.7 when dichorionic twins only were considered. The fact that buccal epithelial tissues of MC MZ twins were significantly more discordant than in DC twins signals chorionicity as an important environmental factor influencing the epigenome and was not taken into account in the Fraga et al. study. The sample was not stratified according to age, which precludes inferences on early discordance. The findings of Kaminsky et al. (2009) could suggest that late twinning can predispose to skewed environmental conditions, or, alternatively, to more discordant epigenetic profiles. In contrast to Fraga and co-workers, Kaminsky et al. argue that stochasticity is “much more important” than environment in driving epigenetic differentiation in genetically identical organisms, however they base their claims not on their own findings, but on earlier surveys showing similar concordance for personality and social attitudes in MZ twins raised apart to hose raised together[82].

The study of Fraga et al. (2005) stressed the significance of age in studies DNA methylation twin discordance, with the youngest twin pair having identical methylation levels. In 2010, Saffery et al. sampled four different tissue types from 56 MZ and 35 DZ twins pairs at birth, and analysed them for CpG methylation at four differentially methylated regions (DMRs) associated with the IGF2/H19 locus[83]. Within MZ pairs, the absolute methylation difference for all DMRs and tissues was in general small and ranged between 3-4%, however the difference varied depending on tissue type and specific CpG tested. An effect of chorionicity was confirmed. The largely similar epigenetic profiles at birth would support the Fraga et al study but conclusions are limited by the investigation of just four genetic regions. A longitudinal study of CpG methylation in MZ twin cohorts, sampled from birth using several different tissues with single CpG dinucleotide resolution would be of invaluable use to estimate the levels of epigenetic discordance and make inferences upon its nature. In one of the first longitudinal twin methylation studies, Wong et al. (2010) attempted to address the
issue of epigenetic heritability and stochastic versus environmental epigenetic change[84]. They examined 46 MZ and 45 DZ twin pairs for methylation at three chosen loci relevant to psychology, first at the age of 5 and then at the age of 10. They found variable discordance in all pairs. Not all loci were equally prone to temporal epigenetic change although alterations were observed for all three genes. By comparing MZ to DZ concordance rates, no significant differences were detected and the authors concluded that alterations were weakly heritable, and thus predominantly attributable to the environment – both shared (indicated by high concordance) and non-shared. Interestingly, despite low heritability, the intraclass correlation coefficients (ICC) of MZ twins remained stable or increased, which seems contradictory to the epigenome-wide findings of Fraga et al. (2005). This might be a consequence of the narrow locus-specific scope of the study [84]. The authors do not give any explanation for the phenomenon and conclude that the epigenome is dynamic and subject to changes and environmental influence. It can differ between MZ twins even in early childhood and, depending on the locus, non-shared as well as common familial environment can significantly affect its methylation profile. However the authors acknowledge that when ICC and heritability is low, indicating little familial environment and genetic contribution, stochastic epimutations can provide an alternative explanation for the discordance. Further confirmation of the authors’ conclusions are required. Table 2 sums up the main methylome stability and discordance studies conducted in MZ twins.

TABLE 1

Altogether, these findings suggest that methylation patterns can be to a large extent genetically-determined and heritable, yet do not remain stable over lifetime. Variable degrees of epigenetic discordance can be observed in MZ twins and it is evident that sample size, age, tissue as well as CpG island selection can all significantly influence its estimates. There is substantial locus-to-locus and inter-individual variation in temporal methylation dynamics. So far there is conflicting evidence on early epigenetic discordance in MZ twins, but age should be a crucial factor in all future studies of methylome changes. Chorionicity appears to be an important factor altering discordance and heritability estimates, therefore studies investigating methylome concordance in twins should also take this into account, although this information is often lacking. One potential problem which can affect findings in longitudinal studies is re-sampling from epigenetically different cellular subpopulations[84]. The environmental influence on the epigenome is relevant and global epigenome studies in
human MZ twins alone cannot resolve the sources of epigenetic discordance. Since intrauterine environment, postnatal shared and non-shared environmental factors as well as sequence polymorphisms acting in cis and trans can all be partly responsible for the methylation discordance, evaluating the significance of the intrinsic, stochastic epigenetic drift poses a methodological obstacle difficult to surmount[2, 29, 73, 83-86]. Estimating the exact proportion of stochastically-determined differences will require a deeper knowledge of the ways in which the non-shared environment shapes the methylome as well as specific mechanisms responsible for the drift[29]. This may be difficult to investigate in humans.

**Methylation Studies and Human Disease**

A number of studies have investigated methylation differences in MZ twins in relation to disease or different phenotypic conditions (Table 2). Initially, studies utilised bisulphite conversion combined with sequencing of pre-selected candidate loci. In some of the very first twin methylation studies Petronis et al (2003) found differences in CpG methylation of a regulatory sequence of the dopamine D2 receptor that was greater in a schizophrenia-discordant pair than in a concordant one[87]. Another early study using bisulphite sequencing found methylation discordance at two regions of the COMT gene promoter in a sample of six MZ twin pairs, all discordant for birth weight[88]. Oates et al. (2006) bisulphite sequenced the AXIN1 promoter in a MZ pair discordant for caudal duplication anomaly[89]. More recently, a survey of CpG methylation at six chosen tumour suppressor genes (ATM, BRCA1, BRCA2, MLH1, RAD51C and TP53) in a single MZ twin pair discordant for childhood leukemia and secondary thyroid carcinoma, revealed increased BRCA2 methylation [90]. The proband had a significantly more methylated promoter than the healthy co-twin. The main limitations of these studies are the small sample sizes and narrow-scope, limiting the number of potential associations to be found.

With the advance of microarray and next-generation sequencing technology, it became possible to study methylation changes on a genome-wide scale. In a study of discordant risk-taking attitudes in a single MZ twin pair, Kaminsky et al (2007) looked at CpG methylation of about 12,192 CpG loci and found differences in methylation of the DLX1 gene, implicated in stress-response[91]. A methylation-sensitive–representational difference analysis study on a MZ pair discordant for bipolar disorder by Kuratomi et al. (2008) yielded four DMRs and
one candidate gene, also confirmed to be differentially expressed[92]. Using an Illumina GoldenGate array, Javierre et al (2009) looked at methylation of 1505 CpG sites in 807 gene promoters across 5 MZ twins discordant for systemic lupus erythematosus (SLE), 5 twins discordant for rheumatoid arthritis (RA) and 5 twins discordant for dermatomyositis (DM), and discovered significant differences at 49 loci between the SLE-affected twins and their healthy co-twins, not seen in the RA and DM discordant twins [93]. The SLE cases had lower methylation levels and higher expression in several genes with immune functions. In a more recent analysis, Baranzini et al (2010) looked for differences in methylation state of approximately 2 million CpG dinucleotides in three MZ twin pairs discordant for Multiple Sclerosis using reduced representation bisulfite sequencing (RRBS)[72]. The authors used high thresholds for methylation differences which reduced the number of differentially methylated loci to 2, 10 and 176 between the different twin pairs. The differences were inconsistent between the three twin pairs leading the authors to conclude that methylation differences could not explain twin discordance. The study’s small sample size and its heterogeneous character (twins of European and African American descent) constituted perhaps the greatest limitation reducing the power to detect significant methylation differences[94]. A recent analysis of methylation in 3 pairs of MZ twins discordant for autism using a 8.1 K CpG microarray yielded 73 differently methylated CpG islands and two candidate genes[95]. The first genome-wide study using the Illumina 27k array (covering ~27000 CpG sites) of MZ twin methylomes in schizophrenia and bipolar disorder in a cohort of discordant 22 pairs revealed a number of disease associated DMRs including **GGN, SLC117A, SMUG1, SOX1** and **TCF7L2 which have been implicated by a previous study** [96]. Methylome profiling with the Illumina 27k array in a MZ twin cohort discordant for psoriasis failed to identify any significant DMRs, however showed that methylation correlates to the levels of expression at some disease-associated loci including **IL13, ALOX5AP, PTHLH and TNFSF11** [97]. A slightly different approach was adopted by Mastreoni et al. (2009), who looked at whole tissue methylation using immunohistochemistry[32]. Different methylation levels in temporal neocortex neuronal nuclei were found in 2 MZ twins discordant for Alzheimer Disease, with hypomethylation in the affected twin[32].

**TABLE 2**

Whilst genome-wide studies have enabled discovery of more DMRs, studies are still in their infancy and face multiple issues [98]. So far, most studies have investigated methylation in
small samples of one to a dozen twin pairs; use of larger discordant MZ twin cohorts will increase the power to detect potentially causal differentially methylated regions. Increasing the size of the twin sample might be challenging for rare diseases and study designs involving longitudinal sampling [98]. Improvements to study designs in the future will likely require sampling from multiple tissues, particularly those that might be relevant to disease as variation of the epigenome varies significantly across different cell types. However, some tissues are not easily accessible and sampling across different tissues might involve biopsies and post-mortem material [98]. Currently, the use of several technologies and platforms makes cross-comparisons difficult [98, 99]. Comparisons between MZ and DZ as well as MC and DC twins should provide insights into the role of genetics and intrauterine environment in shaping epigenetic variation.

**Disease Studies and Causality**

The associations yielded by various methylation studies emphasise the need to develop methods which establish causality [98, 99]. Traditionally, in genetic studies, this was done by demonstrating perfect co-segregation of putative causal alleles with affected individuals in families, as well as alterations to expression or structure of the protein encoded by the allele[100, 101]. In non-Mendelian, complex diseases with a significant environmental component and no clear-cut disease segregation, causality is mainly investigated through case-control association studies by sorting candidate gene using p-value thresholds which minimise false positive errors, and optimally replicating the results in independent cohorts[102-104]. Of course, owing to linkage disequilibrium, population stratification, type I and type II errors or mere chance, association does not equate to causality until proven by functional work [103, 105]. Proving causality is, just like the definition, ultimately always context-dependent, and there is no uniform agreement on what constitutes adequate evidence, however, most authors are clear that some physical, biological link ought to be established[104, 106-111]. Epigenetic alterations at promoter sites should affect transcriptional activity [91, 95]. However, the problem is that epigenetic differences could be in fact side-effects of disease or treatment. Studies investigating epigenetic changes are potentially prone to false conclusions due to reverse causation or confounding[112]. Since the nature of the epigenome is dynamic and most epimutations arise throughout lifetime, the key to addressing causality might be in their timing[94]. A longitudinal approach assaying for
epigenetic discordance at birth or in early infancy and resampling at regular intervals should produce a timeline for methylation changes and help to sort the potentially causal alterations from the secondary, side-effect ones. Epigenetic differences identified in twins could be further investigated in longitudinal cohorts as a part of a two-stage study design[112]. Ultimately, studies of disease-associated epigenetic changes should be followed up with the aim of establish a link with biological function [112]. To date, only one longitudinal epigenome wide study has been conducted investigating single CpG methylation differences in a panel of MZ twins discordant for type 1 diabetes. [113] 132 methylation variable positions associated with disease status were discovered, some of which were replicated in an independent group of 9 type 1 diabetes affected singletons prior to disease diagnosis. [113]

**Conclusion**

The plausible assumption made by Galton that twin discordance can be explained by differential environmental exposures after birth is no longer tenable. Genetics, the *in utero* environment, stochasticity and epigenetics can all potentially play a role in determining phenotypic discordance. The field of epigenetics is in its infancy. There is very strong evidence for the direct role and relevance of epigenetics in shaping human phenotypic variability. The role of the epigenome can be both as a mediator of genetic and environmental effects or as an independent stochastic factor. Currently, the significance of primary epimutations in twin discordance is unknown. Furthermore, it is not 100% clear to what extent the epigenome is heritable and whether monozygotic twins are epigenetically identical at birth. Further studies are required to address these important questions. Ultimately, longitudinal studies with repeated sampling may be required to fully understand the nature of monozygotic twin discordance.

**List of Abbreviations:**

DZ: dizygotic
DC: dichorionic
DA: diamniotic
MZ: monozygotic
MC: monochorionic
MA: monoamniotic
IUGR: intrauterine growth restriction
TTTS: twin-to-twin transfusion syndrome
CNVs: copy number variants
DAE: differential allelic expression
DMRs: differentially methylated regions
SLE: systemic lupus erythematosus
RA: rheumatoid arthritis
DM: dermatomyositis
RBS: reduced representation bisulfite sequencing

Competing Interests: None

Authors’ Contribution:
Review design and concepts: Witold Czyz, Dr Sreeram V Ramagopalan
Manuscript draft: Witold Czyz
Critical revision of the manuscript: Dr Julia Morahan
Study supervision: Dr Sreeram V Ramagopalan, Professor George Ebers

Funding: This work was funded by the Wellcome Trust [090532/Z/09/Z].
Reference:


**Figure 1:** A) The mechanism of a secondary epimutation. A DNA alteration at an imprinting centre (IC) indirectly influences and alters the methylation pattern (represented by black and white triangles) at another locus, which could be in cis or in trans. B) In primary epimutation, the external stimulus (whether environmental or stochastic) alters the methylation directly.

**Table 1:** Studies of CpG methylation discordance in MZ twins (* denotes that not all pairs were used for each estimate in the study):
having differential susceptibility to shared and non-shared exposures.

**Table 2:** Methylation studies in MZ twins discordant for personality and disease:

<table>
<thead>
<tr>
<th>Study</th>
<th>Condition</th>
<th>Method</th>
<th>Tissue</th>
<th>MZ Pairs</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weksberg et al. 2002</td>
<td>Beckwith-Wiedemann syndrome</td>
<td>Southern Blot with DMR probes (H19, KvDMR1, SNRP)</td>
<td>Lymphocytes and Fibroblasts</td>
<td>10</td>
<td>Loss of methylation at (KvDMR1) in all probands</td>
</tr>
<tr>
<td>Petronis et al. 2003</td>
<td>Schizophrenia</td>
<td>Bis-seq (DRD2)</td>
<td>Lymphocytes</td>
<td>1</td>
<td>Discordance confirmed</td>
</tr>
<tr>
<td>Mill et al. 2006</td>
<td>Attention-deficit Hyperactivity Disorder</td>
<td>Bis-seq (COMT)</td>
<td>Buccal epithelial</td>
<td>12</td>
<td>0.1% to 52.3% discordance</td>
</tr>
<tr>
<td>Oates et al. 2006</td>
<td>Caudal Duplication</td>
<td>Bis-seq (AXIN1)</td>
<td>PBMC</td>
<td>1</td>
<td>Discordance confirmed</td>
</tr>
<tr>
<td>Kuratomi et al. 2007</td>
<td>Bipolar Disorder</td>
<td>MS-RDA</td>
<td>Lymphoblastoid cell lines</td>
<td>1</td>
<td>4 DMRs, 1 candidate gene</td>
</tr>
<tr>
<td>Kaminsky et al. 2008</td>
<td>Risk-Taking Behavior</td>
<td>MEDIP-Chip</td>
<td>PBMC</td>
<td>1</td>
<td>38 DMRs, 1 candidate gene</td>
</tr>
<tr>
<td>Mastroeni et al. 2009</td>
<td>Alzheimer Disease</td>
<td>Immunohistochemistry</td>
<td>Temporal neocortex</td>
<td>1</td>
<td>Discordance confirmed</td>
</tr>
<tr>
<td>Javierre et al. 2010</td>
<td>Systemic Lupus Erythematous</td>
<td>MEDIP-Chip</td>
<td>White Blood Cells</td>
<td>5</td>
<td>49 DMRs, 8 candidate genes</td>
</tr>
<tr>
<td>Wong et al. 2010</td>
<td>ADHD, Depression, Antisocial Behaviour</td>
<td>Quantitative High-Throughput Mass Spectrometry (DRD4, SERT, MAOA)</td>
<td>Buccal epithelial</td>
<td>46</td>
<td>Discordance confirmed</td>
</tr>
<tr>
<td>Baranzini et al. 2010</td>
<td>Multiple Sclerosis</td>
<td>RRBS</td>
<td>CD4(^+) lymphocytes</td>
<td>3</td>
<td>2-178 DMRs, no candidate</td>
</tr>
<tr>
<td>Hu et al. 2010</td>
<td>Autism</td>
<td>MEDIP-Chip</td>
<td>Lymphoblastoid cell lines</td>
<td>3</td>
<td>73 DMRs, 2 candidate genes</td>
</tr>
<tr>
<td>Tierling et al. 2010</td>
<td>Beckwith-Wiedemann syndrome</td>
<td>Bis-seq (11 DMR)</td>
<td>Peripheral blood cells</td>
<td>1</td>
<td>Hypomethylation at (KvDMR1)</td>
</tr>
<tr>
<td>Harder et al. 2010</td>
<td>Optic Glioma</td>
<td>Bis-seq (NF1)</td>
<td>Leukocytes</td>
<td>8</td>
<td>Discordance confirmed</td>
</tr>
<tr>
<td>Souren et al. 2011</td>
<td>BMI</td>
<td>Bis-seq</td>
<td>Saliva</td>
<td>8</td>
<td>Small discordance identified not correlated with BMI discordance</td>
</tr>
<tr>
<td>Rakyan et al. 2011</td>
<td>Type 1 Diabetes</td>
<td>Illumina Array</td>
<td>CD14(^+) cells</td>
<td>15 (+9 healthy control)</td>
<td>132 methylation variable positions</td>
</tr>
<tr>
<td>Study</td>
<td>Disease Type</td>
<td>Technology</td>
<td>Cell Type</td>
<td>Sample Size</td>
<td>Findings</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------</td>
<td>------------</td>
<td>--------------------</td>
<td>-------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dampster et al. 2011</td>
<td>Schizophrenia / Bipolar Disorder</td>
<td>Illumina Array</td>
<td>Whole blood</td>
<td>22</td>
<td>Disease associated DMRs, including <em>ST6GALNAC1</em> as the top candidate</td>
</tr>
<tr>
<td>Galetzka et al. 2012</td>
<td>Childhood leukemia / Secondary thyroid carcinoma</td>
<td>Bis-seq (6 tumour suppressors)</td>
<td>Skin fibroblasts</td>
<td>1</td>
<td>Increased <em>BRCA1</em> methylation</td>
</tr>
<tr>
<td>Gervin et al. 2012</td>
<td>Psoriasis</td>
<td>Illumina Array</td>
<td>CD4+ and CD8+ cells</td>
<td>27</td>
<td>No significant methylation differences. Significant correlation between some DMRs and psoriasis-associated gene expression differences.</td>
</tr>
</tbody>
</table>
Figure 1

(A) 
Mutation → Secondary epimutation

DNA strand

IC

CpG island methylation

(B) 
Primary epimutation

DNA strand

IC

CpG island methylation