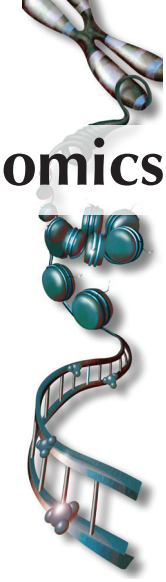
Review

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics



Using epigenomic studies in monozygotic twins to improve our understanding of cancer

Leonie Roos[1], Timothy D Spector[1] & Christopher G Bell*,[1]

[1]Department of Twin Research & Genetic Epidemiology, St Thomas' Hospital, King's College London, London, SE1 7EH, UK

*Author for correspondence: christopher.g.bell@kcl.ac.uk

Cancer is a set of diseases that exhibit not only genetic mutations but also a profoundly distorted epigenetic landscape. Over the last two decades, great advances have been made in identifying these alterations and their importance in the initiation and progression of cancer. Epigenetic changes can be seen from the very early stages in tumorigenesis and dysregulation of the epigenome has an increasingly acknowledged pathogenic role. Epigenomic twin studies have great potential to contribute to our understanding of complex diseases, such as cancer. This is because the use of monozygotic twins discordant for cancer enables epigenetic variation analysis without the confounding influence of the constitutive genetic background, age or cohort effects. It therefore allows the identification of susceptibility loci that may be sensitive to modification by the environment. These studies into cancer etiology will potentially lead to robust epigenetic markers for the detection and risk assessment of cancer.

Keywords: cancer • discordant monozygotic twins • DNA methylation • epigenetics • environment • twin study

Over the last two decades, great advances have been made in our understanding of the pathogenesis of cancer. Disease development is acknowledged as a multistep process, whereby the cancer cell acquires new biological capabilities that enable tumor growth and metastatic dissemination. These include sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastatic pathways [1,2]. These cellular adaptations can be driven through genetic mutations, and critical cancer-related genes have been identified by the co-segregation of mutations in families with rare hereditary cancer syndromes [3–5]. However, it is also well known from epidemiological studies that certain environmental factors have a strong influence on cancer risk, for example tobacco smoking and UV sunlight exposure. In order to quantify this nongenetic effect, the unique power of human twin studies has been fundamental [6].

These classic twin studies have greatly contributed to our knowledge of complex diseases as they allow the separate contribution of genetic and environmental factors to be estimated. Human identical, or monozygotic (MZ), twins account for only one in approximately 250 pregnancies worldwide. These MZ twins originate from an aberrant separation of two or more daughter cells of a single fertilized ovum and thus share an identical genome sequence. The timing of this separation can occur at a developmental stage as late as 14 days after fertilization [7]. Nonidentical, or dizygotic (DZ), twins on the other hand, originate when two ova are present that are independently fertilized and therefore are as genetically similar as normal siblings sharing, on average, 50% of segregating DNA sequence variation [8]. The division of the zygote leading to MZ twins is generally viewed as a random process and not heritable, although some clustering is seen in a few families [9]. MZ twins are extremely similar due to not only their shared genetic code ('virtual

Future Medicine part of fsg

clones'), but also due to their shared early environment. By contrast, DZ twins share most of these environmental factors, but with a comparatively much-reduced genetic component.

Detailed examination of these MZ individuals has revealed that they have surprisingly high rates of disease discordance for many common disorders, including metabolic disease, autoimmune disease, and cancer [10–13]. This difference in disease occurrence between MZ twins is typically interpreted as the result of environmental factors, although this component will also include stochastic effects [14]. What should be kept in mind is that in order to establish the most accurate disease concordance rates, a sufficient observation window is required. This is to diagnose correctly whether an individual is healthy or affected, and should be as near to the entire life span as possible. This is particularly the case for predominately late-onset diseases, such as cancer. The concordance rate, for example, was observed to be below 0.10 in MZ twins with cancer, by a large twin multi-country analysis. In this study, the youngest cohorts diagnosed with prostate, colorectal and breast cancer, were followed up until 70, 63 and 56 years, respectively [15]. For these, the most common types of cancer, the contribution of the nongenetic component was estimated to be 60–70% (Table 1). The precise mechanism of how these external factors contribute to the development of cancer is unknown; however, a prime candidate is via changes in the epigenome [16].

Compared with a generally static DNA sequence throughout life, the epigenome represents a dynamic landscape that undergoes specific changes at various stages in development [17,18], as well as modification at identified aging-related loci and random 'drift' throughout the lifetime of an individual [19–22]. This change over time is believed to be not only due to intrinsic factors, but also environmental effects [23]. It has been shown that the epigenome can be influenced by prenatal and early postnatal environmental influences, such as maternal behavior in rodents [24] and maternal diet in humans [25,26]. Therefore epigenetic epidemiology may have the potential to link changes in gene activity directly or in combination with genetic risk factors, influencing penetrance and expressivity, with environmental conditions and exposures. Due to its essential role in gene expression, abnormalities in the epigenome have been proposed to contribute to the spectrum of complex diseases [27] and especially to human cancers [28,29]. Furthermore, the high level of genetic mutations recently identified in genes that direct the epigenome also points towards this level of regulation as being a significant pathogenic player in cancer [30,31], for example, *ARID1A* in ovarian clear cell carcinoma [32].

In this article, we will discuss primarily the patterns of the most well-studied epigenomic mark, DNA methylation, in healthy and cancer subjects in the general population, the contribution of epigenetic twin studies in complex diseases, and finally the use of these studies in cancer research.

DNA methylome in healthy & cancer tissues

In the past few years, the ability to characterize the epigenome in both normal and cancer tissues has been

Table 1. The concordance rate and effects of heritable and nonshared environmental factors in cancer at various sites.

Site or type	MZ concordance rate (men/women)	Heritable factors (95% CI)	Nonshared environmental factors (95% CI)
Stomach	0.08/0.10	0.28 (0.00–0.51)	0.62 (0.49–0.76)
Colorectum†	0.09/0.16	0.35 (0.10–0.48)	0.60 (0.52–0.70)
Pancreas	0.06/0.03	0.36 (0.00–0.53)	0.64 (0.47–0.86)
Lung	0.11/0.09	0.26 (0.00–0.49)	0.62 (0.51–0.73)
Breast†	0.00/0.14	0.27 (0.04–0.41)	0.67 (0.59–0.76)
Cervix	–/0.02	0.00 (0.00–0.42)	0.80 (0.57–0.97)
Uterus	–/0.02	0.00 (0.00–0.35)	0.82 (0.64–0.98)
Ovary	–/0.05	0.22 (0.00–0.41)	0.78 (0.59-0.99)
Prostate†	0.21/–	0.42 (0.29–0.50)	0.58 (0.50–0.67)
Bladder	0.06/0.00	0.31 (0.00–0.45)	0.69 (0.53–0.86)
Leukemia	0.07/0.00	0.21 (0.00–0.54)	0.66 (0.45–0.88)

†Statistical significant effects of heritable factors.
–: Not applicable; MZ: Monozygotic.
Adapted with permission from [15].

revolutionized and made genome-wide analysis possible. The complexity of tumorigenesis involves both genetic and epigenetic changes that lead to abnormal gene expression patterns regulating the tumor phenotype [33,34]. In the progression of cancer development, cells acquire new characteristics that overcome normal regulation of differentiation, proliferation and cell death. These changes in the transcriptome include processes such as activation of oncogenes and deactivation of tumor suppressor genes [2]. Epigenetic changes, such as DNA methylation, can be seen in early neoplastic and even in precancerous tissues [35,36]. For example, a key early tumor adaptation to frequently encountered hypoxic conditions is accompanied by epigenomic changes [37,38]. Recent data have shown that in cancerous cells, epigenetic changes affect more genes than genetic mutations and can often act as surrogate modifications to commonly altered cancer genes [39]. This further indicates the importance of epigenetic alterations in the initiation and progression of cancer development. Dysregulation of the epigenome is thus a mechanism that enables the cancer cell to gain a growth advantage [40].

DNA methylome in healthy tissues

DNA methylation comprises the covalent addition of a methyl group to the five-carbon position of the cytosine almost exclusively in CpG dinucleotides in differentiated cells in the mammalian genome. CpG dinucleotides are found at a much lower frequency than expected due to the hypermutability of methylated cytosines, with the majority of these randomly dispersed throughout the genome and predominately methylated (70–80%) [41]. However, a small proportion of the total genomic CpG dinucleotides (~7%) typically cluster together in specific regions known as CpG islands (CGIs), half of which are located at established gene promoters, near transcription start sites and first exons. The remainder are evenly split between intragenic and intergenic regions of the genome [42], with the former being potential alternate isoform promoters [43,44]. Approximately 70% of genes in the human genome have CGIs in their promoter region and the majority of these promoter CGIs remain unmethylated [45]. However, methylation in these regions commonly correlates with transcriptional repression and this has been established in genomic imprinting, X inactivation and suppression of transposable elements [46]. Inversely, high levels of DNA methylation within the gene body region is generally associated with higher levels of transcriptional activation [47–49]. Aberrant DNA methylation patterns, both global trends and promoter specific, are a hallmark of cancer [50].

DNA methylome in cancer

Global hypomethylation

The first epigenetic alteration identified in cancer was the global reduction of methylation levels in cancer patients compared with healthy subjects in primary human tumor tissue [51,52]. Recent studies have revealed that the substantial loss of methylation involves large hypomethylated blocks covering more than half of the genome at consistent locations [28,53]. These are primarily located in partially methylated domains comprising a significant number of developmental genes. Surprisingly, hypomethylation of the blocks is accompanied by the formation of repressive chromatin marks associated with lamin nuclear regions and silencing of the genes within these regions [54]. Neighboring these hypomethylated blocks are focal hypermethylation regions at certain promoter CGIs. Further indicating the complexity of integrating expression with methylation, these were associated with gene activation and are termed long-range epigenetic activation regions [34]. Even though only a minority of CGI promoters are methylated in normal somatic cells, there are a number of oncogenes identified that show a loss of methylation at their promoters and associated activation of gene expression in tumor tissues. These include key genes in tumorigenesis such as *RRAS* [55], *MAGE1* [56], *XAGE1A* [57] and *MASPIN* [58]. Overall, the global loss of methylation is a key abnormality in cancer cells and is associated with repressive chromatin marks and silencing of genes within these regions. Together with neighboring long-range epigenetic activation regions, it highlights the importance of chromatin boundaries in healthy cells.

Promoter-specific CGI hypermethylation

The most extensively studied modification is the gain of methylation at normally unmethylated gene promoter CGIs, which are classically associated with transcriptional repression [29,59,60]. Large numbers of genes have been found to be hypermethylated in different types of cancer tissues at various stages of tumorigenesis. This can affect at least 5% of all promoter CGIs that are normally unmethylated in somatic cells [61]. The hypermethylation at individual genes is generally associated with stable gene silencing and can be tumor-type specific [62]. Many tumor suppressor genes have been identified in the last two decades that are hypermethylated and generally repressed in multiple types of cancer including *RB1* [63], *CDKN2A* [64], *MLH1* [65], *VHL* [66] and *BRCA1* [67]. More recently, promoter hypermethylation was found to disrupt miRNA pathways associated with upregulated target oncogenes, such as *BCL6* [68]. Promoter-specific hypermethylation is found in significant cellular pathways, such as DNA repair, cell cycle control, apoptosis and metastasis [69].

Non-CGI methylation variation

Methylation changes in cancer are not restricted to promoter-specific CGIs. In 2009, a study by Irizarry *et al.* showed that the CGI 'shores', the regions approximately 2 kb upstream and downstream of the island, undergo significant methylation changes and also contribute to gene dysregulation. In fact, these differentially methylation regions (DMRs) occurred more frequently here than within the CGIs themselves and were associated with not only cancer-specific, but also tissue-specific and reprogramming-specific changes [70,71]. In addition, hypomethylation was also identified at shores near cell cycle-related genes that were associated with overexpression in cancer cells [28]. This highlighted that in cancer, the loss of the strict boundary in methylation change occurring at the CGI edges is crucial. The importance of CGI shore regions has also been shown during the course of development where they exhibit a tidal-like change with variable narrowing of the hypomethylated region in a lineage-specific fashion during the process of cell differentiation [72]. DMRs within these regions have been found to define tissue specificity more strongly than within the CGI themselves [73]. Furthermore, gene distal moderately concentrated CpG regions, termed low methylation regions [74] or intergenic hypomethylation regions [75] indicate potential tissue-specific enhancer loci. These intergenic hypomethylation regions were found to be more predictive of nearby gene expression than the promoter state itself [75].

Recent advances have further indicated the complexity of interpreting the epigenome and that genome-wide methylation data are required to fully explore and unravel the complete significance of these changes. Analysis of the methylation state of repetitive elements is also required due to their latent regulatory potential and paired-ended sequencing data enables improved access into these genome regions [76]. DMRs can be further delineated for functionality by overlap with genomic segmentation data, derived from coalesced chromatin signals [77], as well as transcription factor-binding motifs, such as for CTCF [78] and other complex interactions [79].

Twin model in epigenetics

MZ twins provide the unique opportunity to investigate the contribution of epigenetic variation between genetically identical individuals in complex diseases. They are ideally matched for genetic factors, age, cohort effects, maternal influences and common environment [80]. The dynamic nature of the epigenome can lead to discordance in these markers between MZ twins. Studies have shown that gene expression profiles in MZ twins already display differences at birth, highlighting the effect of the *in utero* environment [81,82]. Furthermore, developmental changes continue to occur, within, for example, the immune system where differences have been identified longitudinally until at least 5 years of age [83]. Epigenetic studies in twins give valuable insight into the heritability and stability of the epigenome and the epigenetic variation associated with disease discordance [80].

Heritability of the epigenome

Several studies have examined DNA methylation profiles between MZ and DZ pairs to identify the extent of epigenetic variation in genetically identical individuals and the heritability of the epigenome. Heritability is an estimate of the proportion of phenotypic variance that can be attributed to genetic effects in a population. The classic twin method of estimating the heritability compares MZ and DZ concordance rates or intraclass correlations. This method can be used to estimate the proportion of epigenetic (i.e., DNA methylation) variance that can be attributed to genetic effects per single locus [84]. Early array studies investigated methylation profiles between MZ twin pairs [85] and estimated heritability at multiple genomic locations across different tissues to be approximately 15% [85,86]. These findings revealed that MZ twins have very similar epigenetic profiles. However, a greater discordance was observed between older MZ twin pairs [20,85].

More recent analyses have estimated heritability of the methylome on a genome-wide scale using DNA methylation arrays such as the Illumina Infinium Human Methylation 27 Beadchip (Illumina 27k) comprising approximately 27,000 probes and the Illumina Infinium Human Methylation 450 (Illumina 450k) comprising approximately 480,000 probes. First, a study by Bell *et al.* examined heritability using the Illumina 27k comparing methylation rates within 33 MZ and 43 DZ twin pairs. They estimated a mean genome-wide heritability of 18% across 26,690 CpG sites [87]. Subsequently, Grundberg *et al.* performed a large-scale heritability analysis with the Illumina 450k array, including 97 MZ and 162 DZ twins and estimated the heritability for 344,092 methylation sites. In concordance with previous estimates, the overall heritability was estimated to be 19% [88].

DNA methylation profiles have also been shown to diverge over time within MZ twin pairs [85,86]. Recently, DNA methylation levels were assessed both globally (LINE1) and at nine candidate gene loci for common diseases for similarity in 230 MZ twin pairs aged 19–89 years old [19]. The results were similar to those of earlier studies, in that older pairs were found to have double the discordance of younger twins. A longitudinal approach, albeit in five specific loci,

identified a similar pattern of greater within-pair discordance after 10 years, as well as individual-specific methylation changes.

Together these findings confirm that the DNA methylation profile is heritable to a large extent, but does change over the lifetime of an individual, influenced by genetic variation, stochastic factors, and the environment.

Discordant MZ twin model in complex disease

The use of the discordant MZ twin design should elucidate genes that are sensitive to environmental factors in complex diseases. As they are an ultimately matched case–control study, they will be superior to population-based case–control approaches in detecting changes in the epigenome without significant confounding by genetic, aging, and cohort factors [84]. Over the past two decades, a number of studies have identified methylation differences in MZ twins associated with multiple diseases and phenotypes. The earliest studies were characterized by limited twin pair numbers and a focus only on specific candidate gene promoters. A first report found that the *DRD2* gene in lymphocytes had greater methylation differences in the discordant than the concordant pair in association with schizophrenia [89]. Other studies identified methylation differences in single gene promoters in MZ twins discordant for birth weight using buccal swabs and caudal duplication anomaly in peripheral blood mononuclear cells [90,91]. One of the better-known examples of methylation differences is the loss of imprinting at the *KCNQ1OT1* gene, identified in skin fibroblasts, of the affected MZ twin of a pair discordant for Beckwith-Wiedemann syndrome [92].

In the last few years, the advent in high-throughput DNA methylation microarrays have given rise to a number of epigenome-wide association studies in discordant MZ twins. These studies highlight the value of the discordant MZ twin design by unraveling new candidate genes for diseases using whole blood samples or blood-derived cells, such as bipolar disorder [93], systemic lupus erythematous [94], scleroderma [95], autism [96,97] and schizophrenia [98]. Although these studies have small numbers of discordant MZ twin pairs and will require replication in order to be confirmed, they have revealed an interesting new view on DNA methylation changes largely independent of genome sequence variation in a large number of regions across the genome.

In contrast to previous discordant MZ twin studies, a study on pain sensitivity in 25 discordant MZ twin pairs and 50 unrelated subjects assayed genome-wide DNA methylation using methylated DNA immunoprecipitation followed by second-generation sequencing. In a case–control approach, nine genome-wide significant pain DMRs were identified that were subsequently validated by deep bisulfite pyrosequencing, as well as by Illumina 450k array when CpG probes were available in or near the DMR sites (52% of the top 100 pain DMRs). Using a pure discordant MZ twin pair design, the results included several of the top 100 ranked pain DMRs from the case–control approach. These distinguished regions in the genome for pain sensitivity that were nongenetic and may be due to environmental effects [99].

Now with increased availability of base-resolution whole-genome bisulfite sequencing data, meta-analyses have started to determine where in the genome the most variable or 'dynamic' DNA methylation regions are located [41]. These results indicate that the current DNA methylation arrays and even techniques such as reduced-representation bisulfite sequencing are poorly capturing this dynamic methylome (27k: 0.7%; 450k: 8.9%; reduced-representation bisulfite: 11.5%). Whole-genome bisulfite sequencing remains the gold standard, but is extremely costly and inefficient with approximately 70–80% of the reads generated having either no CpGs present or only fully methylated nonvariant CpGs [41].

The epigenomic landscape varies significantly across different tissue types and is a prerequisite consideration when designing any type of epigenetic study. The cell-type that is most relevant to disease will potentially exhibit cell-specific pathogenic epigenetic changes. Unfortunately, some disease-relevant tissues are not easily available and can only be procured via post-mortem material or invasive methods. Many studies have therefore been performed using peripheral tissues, such as whole blood, as an alternative to try to identify surrogate epigenetic changes associated with disease. Any findings arising should therefore be cross-validated in disease-related tissue. Nevertheless, several studies have shown evidence that some epigenetic changes in blood were similar to changes seen across different tissues [100,101].

The rise of discordant MZ twin epigenome-wide association studies has the potential to identify genes and loci sensitive to the environment that are associated with the range of complex diseases, but there is an increasing need for high-resolution studies, with sufficient numbers of samples, and replication/validation of results to yield convincing results.

Discordant MZ twin studies in cancer

At present, two studies have focused solely on DNA methylation differences in cancer discordant MZ twins. Apart from familial cancers (5–10%), in which highly penetrant genetic mutations play a major role [102], the

contribution of common inherited genetic factors identified in sporadic cancers is still moderate. The recent surge of genome-wide association studies since 2007 have identified multiple common genetic variants in different cancer phenotypes, although they individually have relatively modest effect sizes (~1.1–1.3 relative risk), and even in combination can only account for a total of approximately 10% of the heritability [14]. As the DNA methylome can be, to a large extent, heritable due to the influence of CpG-SNPs and the existence of various methylation quantitative trait loci, MZ twins can help to unravel the potential environmental component of the epigenetic mutations seen in cancer (Figure 1).

Galetzka *et al.* described one MZ twin pair discordant for childhood leukemia and secondary thyroid cancer [103]. They assessed DNA methylation at the promoter sites of six tumor suppressor genes including the well-studied *BRCA1* gene in skin tissue. The BRCA1 protein is significantly involved in the maintenance of genome integrity and mutations at the *BRCA1* gene causes most hereditary breast and ovarian cancers [104]. At the *BRCA1* promoter, an increase in methylation was shown constitutively in the affected twin, when compared with her healthy co-twin. Additionally, basal expression levels of BRCA1 protein were lower in the affected twin. They proposed that the difference in DNA methylation may have occurred during early embryogenesis, after the blastocyst fission and predisposes the individual to tumor development later in life [103]. It is an interesting finding, however limited as they could only study one MZ pair and six tumor suppressor genes. Hypermethylation of the *BRCA1* promoter has been previously identified in 10–15% of sporadic nonfamilial breast cancer [105,106] and, thus, strongly supports the hypothesis that epigenetic, as well as genetic modifications, contribute to cancer [107].

The second study by Heyn *et al.* assessed DNA methylation differences in whole blood on an epigenome-wide scale in 15 MZ twin pairs discordant for breast cancer [108]. The Illumina 450k was used to uncover key genes involved in the development of cancer. They detected 403 differentially methylated positions, of which the vast majority of loci were hypomethylated in the affected twin. Interestingly, 27% were located in CGI shore regions. Of these results, 14 CpG sites were examined in a validation cohort of 21 discordant MZ twin pairs and the gene *DOK7*, was shown to be hypermethylated in the cancer-affected twins. This was consistent across four neighboring CpG sites, forming a cancer-related DMR. The increased methylation was also shown in a subset of four whole blood samples taken 4–5 years prior to diagnosis and was confirmed in primary tumor samples and cancer cell lines, suggesting that *DOK7* and similar genes could be promising epigenetic blood biomarkers for early breast cancer diagnosis [108].

Future perspective

The power to detect methylation differences in cancer using the discordant MZ design will depend on many factors, other than improving sample size. Most studies use whole blood and this analysis contains a mixture of the constituent blood subtypes' epigenomes and, thus, is a composite representation that needs to be taken into account. The sensitivity of DNA methylome analysis means cell-type deconvolution is now possible, and will improve with increasingly higher resolution data [109], although subtle cell-type changes themselves could be pathognomonic. Exposure-specific non-shared environmental risk factors may also be present across cell types, but only be involved in pathogenesis in particular cell types in which the disease-related gene is expressed [110], thus further adding a surrogate potential of blood methylation findings.

As twin cohorts are often longitudinal, they contain a wealth of information, making it possible to have DNA methylation samples even prior to diagnosis. The opportunity to study the predisease methylation state in MZ twins could identify methylation differences that provide the 'first hit', making individuals more susceptible to the development of cancer and therefore serve as a predictive biomarker. Using nonprimary tissue samples taken after cancer diagnosis could identify markers that are related to systemic epigenetic events that reflect the consequence of this disease or treatment.

One of the major limitations for a cancer discordant MZ twin design is the collection of samples. Although there are over 60 twin registries worldwide [111], few except for the larger European collections are focused on cancer and have DNA saved either from blood or tissue. Furthermore, substantial numbers are needed for a powerful discovery set and therefore pooling of resources is required. TwinsUK, together with six other twin registries from Europe and Australia, are involved in the set up of a consortium of this nature (EUrodiscotwin), with the purpose of collecting informative discordant MZ twin pairs. This, in combination with replication in vast available case–control data, could make a major impact in advancing knowledge for many types of cancers.

Genome-wide epigenetic-modifying agents are already licensed for some hematological cancers [29] and trialed in others [112]; however, the future possibilities of targeted 'epidrugs', due to the malleability of epigenetic changes, could be considerable. In the coming years, these modifications will require functional

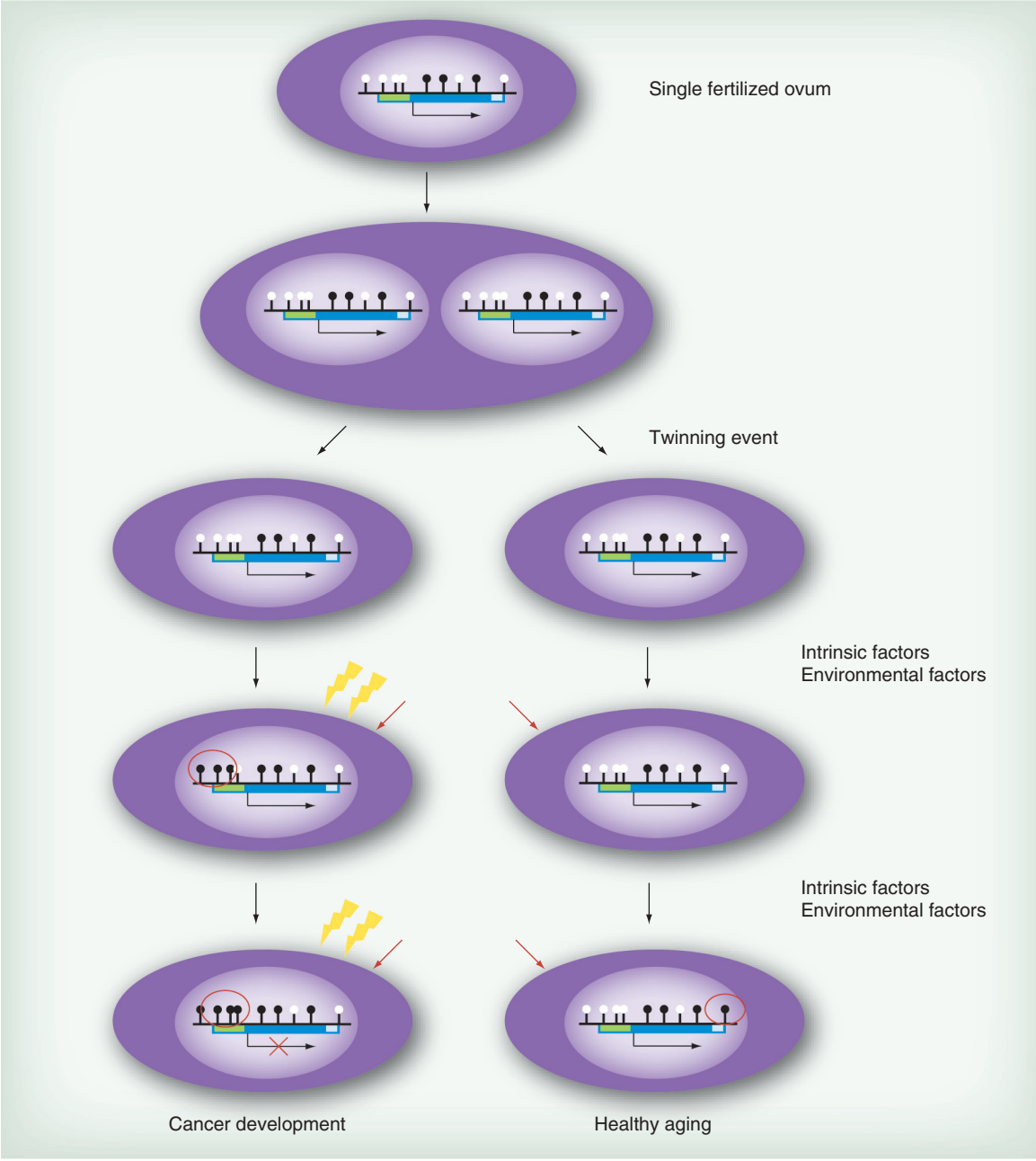


Figure 1. Contribution of epigenetic variation to cancer discordance within monozygotic twin pairs. One actively transcribed tumor suppressor gene in a single cell is used to illustrate how epigenetic changes could lead to cancer discordance over a lifetime within a monozygotic twin pair. The white and black circles represent unmethylated and methylated CpG sites respectively, across different regions of the gene. At the time of the twinning, both twins still have same methylation pattern. Over time, due to intrinsic and environmental factors, the pattern becomes markedly different. The one twin on the left acquires methylation on the promoter site of the gene, which may influence expression and predispose this twin to cancer development, while the other healthy twin on the right, depicts a healthy aging process with just minor changes.

assessment through rapidly emerging sequence-specific technologies, such as TALENs [113] and CRISPR [114]. Thus, if through the exceptional biology of MZ twins convincing variation can be first identified, these epigenetic changes have the potential to move from markers of disease to targets of future therapeutics.

Financial & competing interests disclosure

Research in the Department of Twin Research & the TwinsUK cohort are funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007–2013), as well as support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility,

and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. T Spector is holder of an ERC Advanced Principal Investigator award. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Epigenome

- The epigenome has a crucial role in the regulation of gene expression.
- It represents a dynamic landscape that possesses a degree of plasticity to react to the environment.
- It has the potential to link gene activity with environmental conditions and exposures.

DNA methylome in cancer

- Cancerous cells have a profoundly distorted epigenetic landscape.
- Recent data have shown that, in cancerous cells, more genes are affected by epigenetic changes than genetic changes.

Discordant monozygotic twin studies in cancer

- The discordant monozygotic design offers the unique opportunity to study epigenetic variation without significant confounding by genetic factors.
- They have the potential to quantify environmental risk factors through the measurement of epigenetic modifications.
- Epigenetic studies in cancer discordant monozygotic twins will enable novel insights in cancer etiology, improving prediction, diagnosis and prognosis.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 100(1), 57–70 (2000).

2 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 144(5), 646–674 (2011).

3 Klein AP, Brune KA, Petersen GM *et al.* Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer Res.* 64(7), 2634–2638 (2004).

4 Geary J, Sasieni P, Houlston R *et al.* Gene-related cancer spectrum in families with hereditary non-polyposis colorectal cancer (HNPCC). *Fam. Cancer* 7(2), 163–172 (2008).

5 Casadei S, Norquist BM, Walsh T *et al.* Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. *Cancer Res.* 71(6), 2222–2229 (2011).

6 Van Dongen J, Slagboom PE, Draisma HH, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat. Rev. Genet.* 13(9), 640–653 (2012).

7 Hall J. Twinning. *Lancet* 362(9385), 735–743 (2003).

8 Hall JG. Twins and twinning. *Am. J. Med. Genet.* 61(3), 202–204 (1996).

9 Hamamy HA, Ajlouni HK, Ajlouni KM. Familial monozygotic twinning: report of an extended multigeneration family. *Twin Res.* 7(3), 219–222 (2004).

10 Kendler KS, Prescott CA. A population-based twin study of lifetime major depression in men and women. *Arch. Gen. Psychiatry* 56(1), 39–44 (1999).

11 MacGregor AJ, Snieder H, Rigby AS *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* 43(1), 30–37 (2000).

12 Sadovnick AD, Armstrong H, Rice GP *et al.* A population-based study of multiple sclerosis in twins: update. *Ann. Neurol.* 33(3), 281–285 (1993).

13 Bogdanos DP, Smyk DS, Rigopoulou EI *et al.* Twin studies in autoimmune disease: genetics, gender and environment. *J. Autoimmun.* 38(2–3), J156–169 (2012).

14 Manolio TA. Bringing genome-wide association findings into clinical use. *Nat. Rev. Genet.* 14(8), 549–558 (2013).

15 Lichtenstein P, Holm NV, Verkasalo PK *et al.* Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* 343(2), 78–85 (2000).

16 Feinberg AP. An epigenetic approach to cancer etiology. *Cancer J.* 13(1), 70–74 (2007).

17 Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447(7143), 425–432 (2007).

18 Kota SK, Feil R. Epigenetic transitions in germ cell development and meiosis. *Develop. Cell* 19(5), 675–686 (2010).

19 Talens RP, Christensen K, Putter H *et al.* Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* 11(4), 694–703 (2012).

20 Wong CC, Caspi A, Williams B *et al.* A longitudinal study of epigenetic variation in twins. *Epigenetics* 5(6), 516–526 (2010).

21 Rakyan VK, Down TA, Maslau S *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.* 20(4), 434–439 (2010).

22 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.* 20(4), 440–446 (2010).

23 Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* 13(2), 97–109 (2011).

24 Weaver IC, Cervoni N, Champagne FA *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7(8), 847–854 (2004).

25 Anderson OS, Nahar MS, Faulk C *et al.* Epigenetic responses following maternal dietary exposure to physiologically relevant levels of bisphenol A. *Environ. Mol. Mutagen.* 53(5), 334–342 (2012).

26 Tobi EW, Lumey LH, Talens RP *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* 18(21), 4046–4053 (2009).

27 Liu Y, Aryee MJ, Padyukov L *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* 31(2), 142–147 (2013).

28 Hansen KD, Timp W, Bravo HC *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43(8), 768–775 (2011).

29 Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. *Nat. Rev. Cancer* 11(10), 726–734 (2011).

30 Dawson MA, Kouzarides T, Huntly BJ. Targeting epigenetic readers in cancer. *N. Engl. J. Med.* 367(7), 647–657 (2012).

31 Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat. Rev. Genet.* 14(11), 765–780 (2013).

32 Jones S, Wang TL, Shih Ie M *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* 330(6001), 228–231 (2010).

33 You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* 22(1), 9–20 (2012).

34 Bert SA, Robinson MD, Strbenac D *et al.* Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell* 23(1), 9–22 (2013).

35 Garcia P, Manterola C, Araya JC *et al.* Promoter methylation profile in preneoplastic and neoplastic gall bladder lesions. *Mol. Carcinogen.* 48(1), 79–89 (2009).

36 Baylin SB, Ohm JE. Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nat. Rev. Cancer* 6(2), 107–116 (2006).

37 Shahrzad S, Bertrand K, Minhas K, Coomber BL. Induction of DNA hypomethylation by tumor hypoxia. *Epigenetics* 2(2), 119–125 (2007).

38 Harris AL. Hypoxia – a key regulatory factor in tumour growth. *Nat. Rev. Cancer* 2(1), 38–47 (2002).

39 Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7(1), 21–33 (2006).

40 Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* 13(7), 497–510 (2013).

41 Ziller MJ, Gu H, Muller F *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463), 477–481 (2013).

42 Illingworth RS, Gruenewald-Schneider U, Webb S *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6(9), e1001134 (2010).

43 Illingworth R, Kerr A, Desousa D *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 6(1), e22 (2008).

44 Maunakea AK, Nagarajan RP, Bilenky M *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466(7303), 253–257 (2010).

45 Weber M, Hellmann I, Stadler MB *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39(4), 457–466 (2007).

46 Bird A. DNA methylation patterns and epigenetic memory. *Genes Develop.* 16(1), 6–21 (2002).

47 Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. *Science* 315(5815), 1141–1143 (2007).

48 Ball MP, Li JB, Gao Y *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 27(4), 361–368 (2009).

49 Aran D, Toperoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* 20(4), 670–680 (2011).

50 Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 8(4), 286–298 (2007).

51 Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301(5895), 89–92 (1983).

52 Gama-Sosa MA, Slagel VA, Trewyn RW *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* 11(19), 6883–6894 (1983).

53 Hon GC, Hawkins RD, Caballero OL *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22(2), 246–258 (2012).

54 Berman BP, Weisenberger DJ, Aman JF *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44(1), 40–46 (2012).

55 Nishigaki M, Aoyagi K, Danjoh I *et al.* Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. *Cancer Res.* 65(6), 2115–2124 (2005).

56 De Smet C, De Backer O, Faraoni I, Lurquin C, Brasseur F, Boon T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc. Natl Acad. Sci. USA* 93(14), 7149–7153 (1996).

57 Lim JH, Kim SP, Gabrielson E, Park YB, Park JW, Kwon TK. Activation of human cancer/testis antigen gene, XAGE-1, in tumor cells is correlated with CpG island hypomethylation. *Int. J. Cancer* 116(2), 200–206 (2005).

58 Akiyama Y, Maesawa C, Ogasawara S, Terashima M, Masuda T. Cell-type-specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells. *Am. J. Pathol.* 163(5), 1911–1919 (2003).

59 Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat. Rev. Cancer* 4(2), 143–153 (2004).

60 Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* 349(21), 2042–2054 (2003).

61 Schuebel KE, Chen W, Cope L *et al.* Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet.* 3(9), 1709–1723 (2007).

62 Esteller M, Corn PG, Baylin SB, Herman JG. A gene hypermethylation profile of human cancer. *Cancer Res.* 61(8), 3225–3229 (2001).

63 Ohtani-Fujita N, Fujita T, Aoike A, Osifchin NE, Robbins PD, Sakai T. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene* 8(4), 1063–1067 (1993).

64 Gonzalez-Zulueta M, Bender CM, Yang AS *et al.* Methylation of the 5′ CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Res.* 55(20), 4531–4535 (1995).

65 Kane MF, Loda M, Gaida GM *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* 57(5), 808–811 (1997).

66 Herman JG, Latif F, Weng Y *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl Acad. Sci. USA* 91(21), 9700–9704 (1994).

67 Rice JC, Massey-Brown KS, Futscher BW. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. *Oncogene* 17(14), 1807–1812 (1998).

68 Saito Y, Liang G, Egger G *et al.* Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell* 9(6), 435–443 (2006).

69 Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum. Mol. Genet.* 16(1), R50–R59 (2007).

70 Irizarry RA, Ladd-Acosta C, Wen B *et al.* The human colon cancer methylome shows similar hypo- and hyper-methylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41(2), 178–186 (2009).

•• **The first study to identify the significant methylation changes occuring in CpG island 'shore' regions.**

71 Doi A, Park IH, Wen B *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41(12), 1350–1353 (2009).

72 Hodges E, Molaro A, Dos Santos CO *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* 44(1), 17–28 (2011).

• **Highlights the tidal nature of methylation change around CpG islands with the process of differentiation.**

73 Bock C. Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13(10), 705–719 (2012).

74 Stadler MB, Murr R, Burger L *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480(7378), 490–495 (2011).

75 Schlesinger F, Smith AD, Gingeras TR, Hannon GJ, Hodges E. *De novo* DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res.* 23(10), 1601–1614 (2013).

76 Ward MC, Wilson MD, Barbosa-Morais NL *et al.* Latent regulatory potential of human-specific repetitive elements. *Mol. Cell* 49(2), 262–272 (2013).

77 Ernst J, Kheradpour P, Mikkelsen TS *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345), 43–49 (2011).

78 Wang H, Maurano MT, Qu H *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 22(9), 1680–1688 (2012).

79 Hu S, Wan J, Su Y *et al.* DNA methylation presents distinct binding sites for human transcription factors. *eLife* 2, e00726 (2013).

80 Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends Genet.* 27(3), 116–125 (2011).

81 Gordon L, Joo JH, Andronikos R *et al.* Expression discordance of monozygotic twins at birth: effect of intrauterine environment and a possible mechanism for fetal programming. *Epigenetics* 6(5), 579–592 (2011).

82 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* 38(1), 73–79 (2009).

83 Martino DJ, Tulic MK, Gordon L *et al.* Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* 6(9), 1085–1094 (2011).

84 Bell JT, Saffery R. The value of twins in epigenetic epidemiology. *Int. J. Epidemiol.* 41(1), 140–150 (2012).

85 Fraga MF, Ballestar E, Paz MF *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* 102(30), 10604–10609 (2005).

• **Study recognized the significant epigenetic differences that arise in monozygotic twins over time.**

86 Kaminsky ZA, Tang T, Wang SC *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* 41(2), 240–245 (2009).

•• **First study that estimated DNA methylation heritability in monozygotic and dizygotic twins on a genome-wide scale.**

87 Bell JT, Tsai PC, Yang TP *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 8(4), e1002629 (2012).

88 Grundberg E, Meduri E, Sandling JK *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Human Genet.* 93(5), 876–890 (2013).

89 Petronis A, Gottesman, Ii, Kan P *et al.* Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophren. Bull.* 29(1), 169–178 (2003).

90 Mill J, Dempster E, Caspi A, Williams B, Moffitt T, Craig I. Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 141B(4), 421–425 (2006).

91 Oates NA, Van Vliet J, Duffy DL *et al.* Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am. J. Human Genet.* 79(1), 155–162 (2006).

92 Weksberg R, Shuman C, Caluseriu O *et al.* Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. *Hum. Mol. Genet.* 11(11), 1317–1325 (2002).

93 Kuratomi G, Iwamoto K, Bundo M *et al.* Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol. Psychiatry* 13(4), 429–441 (2008).

94 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.* 20(2), 170–179 (2010).

95 Selmi C, Feghali-Bostwick CA, Lleo A *et al.* X chromosome gene methylation in peripheral lymphocytes from monozygotic twins discordant for scleroderma. *Clin. Exp. Immunol.* 169(3), 253–262 (2012).

96 Nguyen A, Rauch TA, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J.* 24(8), 3036–3051 (2010).

97 Wong CC, Meaburn EL, Ronald A *et al.* Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol. Psychiatry* 19, 495–503 (2013).

98 Dempster EL, Pidsley R, Schalkwyk LC *et al.* Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum. Mol. Genet.* 20(24), 4786–4796 (2011).

99 Bell JT, Loomis AK, Butcher LM *et al.* Differential methylation of the TRPA1 promoter in pain sensitivity. *Nat. Commun.* 5, 2978 (2014).

100 Davies MN, Volta M, Pidsley R *et al.* Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol.* 13(6), R43 (2012).

101 Horvath S, Zhang Y, Langfelder P *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 13(10), R97 (2012).

102 Garber JE, Offit K. Hereditary cancer predisposition syndromes. *J. Clin. Oncol.* 23(2), 276–292 (2005).

103 Galetzka D, Hansmann T, El Hajj N *et al.* Monozygotic twins discordant for constitutive BRCA1 promoter methylation, childhood cancer and secondary cancer. *Epigenetics* 7(1), 47–54 (2012).

104 Huen MS, Sy SM, Chen J. BRCA1 and its toolbox for the maintenance of genome integrity. *Nat. Rev. Mol. Cell Biol.* 11(2), 138–148 (2010).

105 Dobrovic A, Simpfendorfer D. Methylation of the *BRCA1* gene in sporadic breast cancer. *Cancer Res.* 57(16), 3347–3350 (1997).

106 Rice JC, Ozcelik H, Maxeiner P, Andrulis I, Futscher BW. Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens. *Carcinogenesis* 21(9), 1761–1765 (2000).

107 Stefansson OA, Esteller M. Epigenetic modifications in breast cancer and their role in personalized medicine. *Am. J. Pathol.* 183(4), 1052–1063 (2013).

108 Heyn H, Carmona FJ, Gomez A *et al.* DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis* 34(1), 102–108 (2013).

•• **The first epigenome-wide cancer discordant monozygotic twin study. This revealed a promising blood biomarker for breast cancer diagnosis.**

109 Koestler DC, Christensen B, Karagas MR *et al.* Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. *Epigenetics* 8(8), 816–826 (2013).

110 Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* 14(8), 585–594 (2013).

111 Hur YM, Craig JM. Twin registries worldwide: an important resource for scientific research. *Twin Res. Hum. Genet.* 16(1), 1–12 (2013).

112 Nabilsi NH, Pardo CE, Zajac-Kaye M, Kladde MP. Breast cancer epigenetics: biomarkers and therapeutic potential. In: *Cell and Molecular Biology of Breast Cancer*. Schatten H (Ed.). Humana Press, NY, USA, 211–259 (2013).

113 Reyon D, Maeder ML, Khayter C *et al.* Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. *Curr. Protoc. Mol. Biol.* 12, 12 (2013).

114 Ramalingam S, Annaluru N, Chandrasegaran S. A CRISPR way to engineer the human genome. *Genome Biol.* 14(2), 107 (2013).