

Short Review

Epigenetic Epidemiology of Complex Diseases Using Twins

Qihua Tan

Department of Epidemiology and Biostatistics, Institute of Public Health, and
Department of Clinical Genetics, Odense University Hospital/Institute of Clinical Research,
University of Southern Denmark, Odense, Denmark

Key Words

Epigenetic epidemiology · Complex diseases · Environments · Twins

Abstract

In the past decades, studies on twins have had a great impact on dissecting the genetic and environmental contributions to human diseases and complex traits. In the era of functional genomics, the valuable samples of twins help to bridge the gap between gene activity and environmental conditions through multiple epigenetic mechanisms. This paper reviews the new developments in using twins to study disease-related epigenetic alterations, links them to lifetime environmental exposure with a focus on the discordant twin design and proposes novel data-analytical approaches with the aim of promoting a more efficient use of twins in epigenetic studies of complex human diseases.

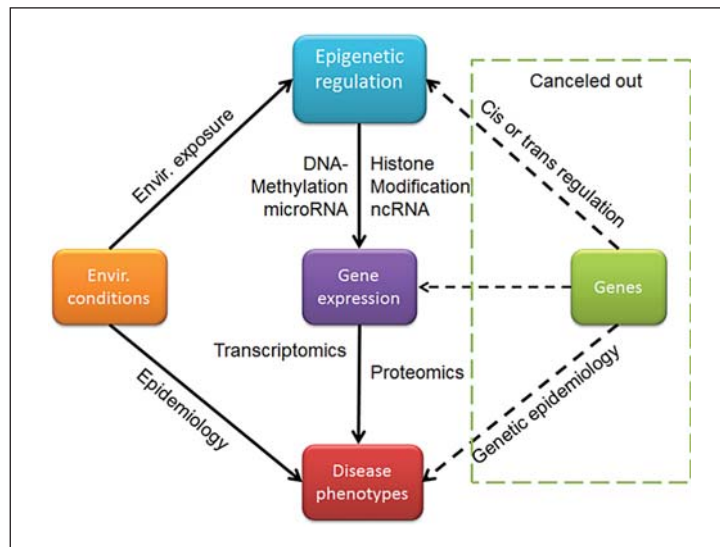
© 2013 S. Karger AG, Basel

Introduction

By definition, epigenetics deals with heritable and acquired changes in gene function that occur without a change in the DNA sequence. In a broad sense, the epigenetic control over gene activity involves multiple molecular mechanisms including the binding of small molecules to specific sites in DNA or chromatin, noncoding RNAs, microRNAs, incorporation of histone variants into chromatin, etc., all of which act as ‘volume controls’ that up- or down-regulate a gene’s expression without changing its DNA sequence. Figure 1 illustrates the coverage of epigenetic epidemiology and its relation to traditional epidemiology and genetic epidemiology. Among the various epigenetic mechanisms, DNA methylation is the major form of epigenetic modification. It is very robust and readily measurable using high-throughput techniques [1], which enable massive epigenetic profiling on a genome-wide scale. Although

Qihua Tan
Department of Epidemiology and Biostatistics, Institute of Public Health
University of Southern Denmark
J.B. Winsløvs Vej 9b, DK–5000 Odense C (Denmark)
E-Mail qtan@health.sdu.dk

Fig. 1. The diagram displays the relationship among traditional epidemiology, genetic epidemiology and epigenetic epidemiology. With the discordant MZ twin design, the genetic part to the right that serves as the confounder in EWAS is canceled out (modified from Tan et al. [7]). ncRNA = Non-coding RNA.



the molecular evidence is interesting and current techniques allow genome-wide epigenetic (DNA methylation) profiling, identifying and understanding the epigenetic patterns under a given genetic predisposition and environmental exposure impose new challenges to traditional epidemiology both in experimental design and in methodological issues.

Like any complex trait, the epigenetic regulation of gene activity is under control of both genetic and environmental factors. In fact, recent studies have shown that the impact of environmental factors can be acquired via the epigenome or genome in epigenetics [2–6], which is one of the hot topics in cancer and complex disease studies that is currently drawing active research. Fueled by the rapid development in epigenomic analysis using next-generation sequencing or array-based technologies, the newly emerging epigenetic epidemiology is serving as a bridge linking gene activity and environmental conditions. In this regard, twins are useful samples that can help with dissecting the genetic and environmental components in epigenetic regulation [7]. In particular, identical or monozygotic (MZ) twins discordant for a disease or trait of interest are unique samples for associating epigenetic alteration with disease and environmental conditions.

Discordant Twin Design

Similar to genome-wide association studies, an epigenome-wide association study (EWAS) can be done using a simple case-control design [8]. Data created from such studies can easily be handled by Student's t test or a simple regression analysis. It should be noted that in genome-wide association studies we associate disease status or phenotype with static genotypes (phenotype-genotype association), whereas in EWAS, we correlate disease with a molecular functional phenotype (e.g. DNA methylation level) which can be under the control of a large number of both genetic and environmental factors (phenotype-phenotype association). The latter can be low powered by uncertainty in the molecular phenotype brought about by genetic and nongenetic confounding.

With the engagement of the discordant MZ twin design, the genetic part of the confounder is removed (fig. 1) because the genetic background is perfectly matched out within each pair [7], enabling the identification of differential DNA methylation triggered by environmental

exposure. With this design, identical twins discordant for a phenotype of interest are sampled from independent families with which intrapair genome-wide epigenetic differences can be determined by high-throughput techniques for measuring differential epigenetic modifications.

In the literature, the design has been applied in epigenetic studies for more than 10 years [9]. As shown in table 1, more and more epigenetic studies have used the discordant twin design, which has experienced considerable increases in sample sizes in recent years. Meanwhile, coverage in these studies has expanded from focused genes or regions to genome-wide analysis. Early epigenetic studies adopting the discordant twin design used descriptive analysis, perhaps due to very small sample sizes (table 1). Even with increasing sample sizes in recent years, most studies rely on a simple paired t test for data analysis. Analyzing and interpreting the high-dimensional data has thus become a challenging issue.

Data Analysis Strategies

Although the paired t test seems to fit the design nicely, there are important issues that need to be considered. First, it ignores twin-pair-specific covariates such as age and sex that could affect within-pair epigenetic differences. The effect of age on within-twin-pair epigenetic variation was illustrated by Fraga et al. [2] who found that young twins are epigenetically more similar than old twins, suggesting epigenetic modification induced by accumulated stimulation by internal and external factors including environmental exposure. We calculated within-pair correlations on genome-wide DNA methylation profiles for identical twins measured 10 years apart (longitudinal samples) and for twins at young and old ages (cross-sectional samples). Both showed obvious declines in the age pattern of epigenetic correlation in MZ twins [unpubl. data]. Furthermore, the paired t test is unable to include environmental exposure variables such as smoking and drinking. Those exposure variables are valuable for epidemiological analysis. This means that the univariate analysis only captures the association between epigenetic variation and disease, but leaves the environmental involvement in the association unassessed. Apart from the fixed effects of twin-pair-specific and environmental variables, there are also random factors such as sample location on the array, laboratory environment and variations due to time and technicians during handling, the so-called batch effect, which need to be considered because they introduce systematic influence on experimental outcomes. Taking all of these factors into consideration, we propose a mixed-model approach specifically designed for EWAS on discordant twins with the following formula:

$$\log \left[\frac{Me(+)}{Me(-)} \right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + 1|y_1 + 1|y_2 + \dots + 1|y_m$$

Here, $Me(+)$ and $Me(-)$ are DNA methylation levels measured in the affected and unaffected twin of the same pair, β_s are the slopes for fixed-effect variables (x_1 to x_n) and y_1 to y_m are the random-effect variables including batch effects. In this model, the intercept α is an important parameter because it stands for the mean fold change in the DNA methylation level between affected and unaffected twins. By testing the null hypothesis of $\alpha = 0$, we are able to test if the disease is associated with hypermethylation ($\alpha > 0$) or hypomethylation ($\alpha < 0$) at the CpG site being tested. Except for the twin-pair-specific variables, e.g. age and sex, the rest of the fixed effects are due to differential exposure to specific environments. If the β of a differential exposure variable is significantly different from zero, the corresponding environment causes hypermethylation ($\beta > 0$) or hypomethylation ($\beta < 0$) at the CpG site in the affected twins, establishing the link between environmental exposure and epigenetic regulation and disease.

Table 1. Summary of epigenetic studies using MZ twin pairs discordant for disease or trait

Study	Year of publ.	Sample size, pairs	Coverage	Laboratory method	Statistical method	Discordant trait	Finding
Weksberg et al. [9]	2002	10	KCNQ10T gene	Southern blotting	descriptive	BWS	loss of imprinting predisposes to BWS
Petronis et al. [11]	2003	2	DRD2 gene	Bis-seq	descriptive	schizophrenia	high epigenetic similarity in affected twins
Oates et al. [12]	2006	1	AXIN1 gene	Bis-seq	descriptive	caudal duplication	high methylation in the promoter region of the AXIN1 gene in the affected twin
Mill et al. [13]	2006	12	COMT gene	Bis-seq	descriptive	birth weight	epigenetic variation in MZ twins may play a key role in the etiology of psychopathology
Kuratomi et al. [14]	2008	1	genome-wide	MS-RDA	descriptive	bipolar disorder	altered DNA methylation of PPIEL is associated with bipolar disorder
Kaminsky et al. [15]	2008	1	genome-wide	array	descriptive	risk-taking behavior	some DNA methylation differences may have developmental and behavioral implications
Mastroeni et al. [16]	2009	1	genome-wide	immunohistochemistry	t test	Alzheimer's disease	low levels of DNA methylation are observed in temporal neocortex neuronal nuclei of diseased twin
Javierre et al. [17]	2010	5	genome-wide	array	paired t test	systemic lupus erythematosus	discordant twins feature widespread changes in DNA methylation
Baranzini et al. [18]	2010	3	genome-wide	array	descriptive	multiple sclerosis	there are very few changes in CpG methylation between siblings
Nguyen et al. [19]	2010	3	genome-wide	array	SAM	autism	epigenetic contributions to autism are confirmed
Harder et al. [20]	2010	8	NF1 regions	Bis-seq	paired t test	NF1	DNA methylation of the normal NF1 allele is responsible for the modification of the NF1 phenotype
Tierling et al. [21]	2011	1	candidate regions	Bis-seq	descriptive	BWS	KvDMR1 is exclusively hypomethylated in all cell types of the affected BWS twin
Souren et al. [22]	2011	8	candidate regions	Bis-seq	paired t test	BMI	DNA methylation variability at nine regions does not contribute to discordancy in BMI
Dempster et al. [23]	2011	22	genome-wide	Infinium array	paired t test	bipolar disorder	further evidence for DNA methylation in the etiology of major psychoses
Gervin et al. [24]	2012	27	genome-wide	Infinium array	paired t test	psoriasis	epigenetic alteration potentially contributes to the development of psoriasis
Runyon et al. [25]	2012	21	selected loci	Bis-seq	paired t test	asthma	differential functions of T cell subsets are regulated by changes in DNA methylation
Wong et al. [26]	2013	50	genome-wide	Infinium array	paired t test	autism	numerous differentially methylated regions are associated with autism
Souren et al. [27]	2013	17	genome-wide	Infinium array	rank test	birth weight	no distinguishable genome-wide DNA methylation differences due to birth weight discordance

BWS = Beckwith-Wiedemann syndrome; NF1 = neurofibromatosis type 1; SAM = significance analysis of microarrays.

Twin Modeling in Epigenetic Epidemiology

The classical twin model has made valuable contributions to genetic epidemiology by enabling genetic and environmental contributions to be explored in human diseases. By treating DNA methylation level as a continuous phenotype, the twin model can be applied to the study of epigenetic epidemiology. For example, Bell et al. [10] recently estimated heritability of the DNA methylation phenotype using twin methods and reported a genome-wide heritability estimate of 18%. The twin model, when applied to each region or CpG site, estimates both genetic and environmental contributions to the epigenetic variation at that location. Here, the environmental part can be further divided into the common environment shared by twin pairs (the rearing environment) and the nonshared unique environments for each twin. The former can be interesting because it links the early-life environment with the later-life epigenetic status. Preliminary results from our twin modeling of genome-wide DNA methylation data show that while some genes are highly under genetic control, there are lots of genes whose activities are predominantly regulated by unique environmental factors [unpubl. data]. Further bioinformatic analysis will elucidate the biological and functional profiles of these genes.

Conclusions

Unlike traditional epidemiology that associates environmental exposure with disease, epigenetic epidemiology establishes a link between environmental exposure and epigenetic regulation and disease condition, thereby providing a molecular basis of disease etiology triggered by environmental cues. Although modern technologies allow high-throughput profiling of epigenetic patterns already at the genome level, our understanding of genetic and environmental influences on epigenetic processes remains limited. With proper designs and analytical approaches, studies on twins can help us identify epigenetic marks and link them with environmental exposure. It can be expected that the valuable twin samples are going to make new contributions to revealing and understanding the genetic and epigenetic mechanisms of human diseases.

Acknowledgements

This work was jointly supported by a Novo Nordisk Foundation 2010 research grant and a research grant from Region of Southern Denmark 2010 awarded to Dr. Qihua Tan.

References

- 1 Zhang Y, Jeltsch A: The application of next generation sequencing in DNA methylation analysis. *Genes* 2010; 1:85–101.
- 2 Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suñer D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M: Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 2005;102:10604–10609.
- 3 Wong AH, Gottesman II, Petronis A: Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum Mol Genet* 2005;14:R11–R18.
- 4 Poulsen P, Esteller M, Vaag A, Fraga MF: The epigenetic basis of twin discordance in age-related diseases. *Pediatr Res* 2007;61:38R–42R.
- 5 Szyf M, McGowan P, Meaney MJ: The social environment and the epigenome. *Environ Mol Mutagen* 2008;49: 46–60.

- 6 Ling C, Groop L: Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes* 2009;58:2718–2725.
- 7 Tan Q, Christiansen L, Thomassen M, Kruse TA, Christensen K: Twins for epigenetic studies of human aging and development. *Ageing Res Rev* 2013;12:182–187.
- 8 Rakyan VK, Down TA, Balding DJ, Beck S: Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011;12:529–541.
- 9 Weksberg R, Shuman C, Caluseriu O, Smith AC, Fei YL, Nishikawa J, Stockley TL, Best L, Chitayat D, Olney A, Ives E, Schneider A, Bestor TH, Li M, Sadowski P, Squire J: Discordant KCNQ10T1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. *Hum Mol Genet* 2002;11:1317–1325.
- 10 Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, The MuTHER Consortium, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P: Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 2012;8:e1002629.
- 11 Petronis A, Gottesman II, Kan P, Kennedy JL, Basile VS, Paterson AD, Pependikyte V: Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr Bull* 2003;29:169–178.
- 12 Oates NA, van Vliet J, Duffy DL, Kroes HY, Martin NG, Boomsma DI, Campbell M, Coulthard MG, Whitelaw E, Chong S: Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *Am J Hum Genet* 2006;79:155–162.
- 13 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* 2006;141B:421–425.
- 14 Kuratomi G, Iwamoto K, Bundo M, Kusumi I, Kato N, Iwata N, Ozaki N, Kato T: Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins. *Mol Psychiatry* 2008;13:429–441.
- 15 Kaminsky Z, Petronis A, Wang SC, Levine B, Ghaffar O, Floden D, Feinstein A: Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res Hum Genet* 2008;11:1–11.
- 16 Mastroeni D, McKee A, Grover A, Rogers J, Coleman PD: Epigenetic differences in cortical neurons from a pair of monozygotic twins discordant for Alzheimer's disease. *PLoS One* 2009;4:e6617.
- 17 Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, Jacinto FV, Lopez-Longo FJ, Dopazo J, Forn M, Peinado MA, Carreno L, Sawalha AH, Harley JB, Siebert R, Esteller M, Miller FW, Ballestar E: Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 2010;20:170–179.
- 18 Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo SJ, Kwok P, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF: Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 2010;464:1351–1356.
- 19 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* 2010;24:3036–3051.
- 20 Harder A, Titze S, Herbst L, Harder T, Guse K, Tinschert S, Kaufmann D, Rosenbaum T, Mautner VF, Windt E, Wahlander-Danek U, Wimmer K, Mundlos S, Peters H: Monozygotic twins with neurofibromatosis type 1 (NF1) display differences in methylation of NF1 gene promoter elements, 5' untranslated region, exon and intron 1. *Twin Res Hum Genet* 2010;13:582–594.
- 21 Tierling S, Souren NY, Reither S, Zang KD, Meng-Hentschel J, Leitner D, Oehl-Jaschkowitz B, Walter J: DNA methylation studies on imprinted loci in a male monozygotic twin pair discordant for Beckwith-Wiedemann syndrome. *Clin Genet* 2011;79:546–553.
- 22 Souren NY, Tierling S, Fryns JP, Derom C, Walter J, Zeegers MP: DNA methylation variability at growth-related imprints does not contribute to overweight in monozygotic twins discordant for BMI. *Obesity* 2011;19:1519–1522.
- 23 Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M, Kravariti E, Touloupoulou T, Murray RM, Mill J: Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet* 2011;20:4786–4796.
- 24 Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, Olsen AO, Brandt I, Harris JR, Undlien DE, Lyle R: DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet* 2012;8:e1002454.
- 25 Runyon RS, Cachola LM, Rajeshuni N, Hunter T, Garcia M, Ahn R, Lurmann F, Krasnow R, Jack LM, Miller RL, Swan GE, Kohli A, Jacobson AC, Nadeau KC: Asthma discordance in twins is linked to epigenetic modifications of T cells. *PLoS One* 2012;7:e48796.
- 26 Wong CC, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, Plomin R, Mill J: Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry* 2013, DOI: [10.1038/mp.2013.41](https://doi.org/10.1038/mp.2013.41).
- 27 Souren NY, Lutsik P, Gasparoni G, Tierling S, Gries J, Riemenschneider M, Fryns JP, Derom C, Zeegers MP, Walter J: Adult monozygotic twins discordant for intra-uterine growth have indistinguishable genome-wide DNA methylation profiles. *Genome Biol* 2013;14:R44.