Update TRENDS in Genetics Vol.22 No.7 July 2006

Research Focus

Epigenetics and twins: three variations on the theme

Arturas Petronis

The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, 250 College Street, Toronto Ontario, M5T 1R8, Canada

Twin studies have had a key role in the evaluation of heritability, a population-based estimate of the genetic contribution to phenotypic variation. These studies have led to the revelation that most normal and disease phenotypes are to some extent heritable. Recently, interest has shifted from phenomenological heritability to the identification of trait-specific genes. The era of twin studies, however, is not over: recent epigenetic and global gene expression studies suggest that the most interesting findings in twin-based research are still to come. The increasing realization of the influence of epigenetics in phenotypic outcomes means that the molecular mechanisms behind phenotypic differences in genetically identical organisms can be explored. Analyses of epigenetic twin differences and similarities might yet challenge the fundamental principles of complex biology, primarily the dogma that complex phenotypes result from DNA sequence variants interacting with the environment.

Introduction

Twin-based designs provide an estimate of the relative contribution of genetic and non-genetic factors to a specific phenotype. The basic principle is simple: monozygotic (MZ) twins have identical genomes and dizygotic (DZ) twins share only half of their segregating DNA, and therefore the contribution of genetic factors to a specific trait should be twice the difference of concordance rates between MZ twins and DZ twins. By contrast, phenotypic differences in genetically identical twins are usually treated as a proof of an environmental contribution. Although detecting formal evidence for environmental contributions is relatively trivial, uncovering the specific environmental factors has been hampered by methodological complexities, primarily unclear cause–effect relationships between a specific environmental factor and a phenotype [1]. Although epidemiological studies, as a rule, depend on the assumptions and biases of researchers, large-scale epigenetic and gene expression studies of twins might become a productive means of understanding the impact of the environment on the organism, cell and genome.

There is now increasing evidence that DNA and chromatin modifications react to various types of environmental effects. In plants, flowering by prolonged cold – vernalization – involves changes in histone modification in the discrete domains of genes that encode repressors of flowering [2]. In *Drosophila*, environmental stress or drugs, via chaperone Hsp90, increase the activity of a histone H3 lysine 4 methyltransferase, which activates the chromatin of target genes and exposes previously hidden morphological phenotypes [3]. A series of experiments has demonstrated an epigenetic impact of carcinogenic agents on modifications of DNA and histones [4]. Interestingly, increased pup licking, grooming and arched-back nursing by rat mothers altered the DNA methylation and histone modifications in the promoter of a glucocorticoid receptor gene, expressed in the hippocampus of their offspring [5]. These studies suggest that epigenetic modifications can be a molecular substrate for the impact of the endogenous and exogenous environment. The attempt to dissect complex multidirectional environmental effects can be substituted by a novel indirect approach that would first aim to uncover the molecular epigenetic impact of such effects, and then identify the specific causal factors behind such epigenetic changes. In this article, I discuss three relevant aspects: epigenetic differences in twins; epigenetic similarities in twins; and quantitative epigenetics.

Epigenetic differences in twins

A comparison of identical twins is an ideal design for testing environmental epigenetics, because DNA sequence differences that would be abundant in a singleton-based study is not a confounding factor. There are already several studies searching for epigenetic differences between identical twins. A disease-specific DNA methylation difference, an imprinting defect at *KCNQ1OT1*, was detected in MZ twins affected with Beckwith–Wiedemann syndrome but not in their healthy co-twins [6]. A traditional epidemiological study attempting to identify environmental risk factors in this disease would be almost impossible because of the rarity of the syndrome and the timing of pathological events that take place in early embryogenesis. By contrast, molecular epigenetic studies can identify the disease-specific molecular defect and also provide some insights into the pathogenic mechanism (i.e. lack of maintenance of DNA methylation at a key stage of preimplantation during development [6]. Localized DNA methylation differences in MZ twins, although not associated with any specific phenotype, have been identified in the regulatory regions of genes encoding the dopamine D2 receptor [7] and catechol-o-methyltrasferase [8].

A recent study by Fraga *et al.* investigated global and locus-specific differences of DNA methylation and histone

Corresponding author: Petronis, A. (arturas_petronis@camh.net).
Available online 11 May 2006

modification in a cohort of 40 identical twins aged 3 to 74 years [9]. DNA samples from peripheral blood cells and, in several cases, from epithelial mouth cells, intra-abdominal fat and skeletal muscle biopsies were subjected to various molecular technologies, including bisulfite sequencing-based mapping of methylated cytosines, restriction landmark genomic scanning and chromatin immunoprecipitation experiments. The authors concluded that there is age- dependent epigenetic 'drift'. Intuitively, it makes sense that there should be an increasing number of incongruent epigenetic changes in the aging co-twins. It is important to note, however, that the degree of twin discordance for various complex diseases does not necessarily correlate with age of onset. For example, concordance of Alzheimer's disease, which manifests predominantly in the eight decade of life can be as high as 83% [10], whereas concordance rates for early onset (first and second decade) type 1 diabetes can be as low as 23% [11] (Figure 1). This issue requires further investigation to dissect the DNA sequence contribution versus the epigenetic contributions to the concordance or discordance of a given disease. It is also necessary to move from a course-scale genome-wide analysis to locus-specific epigenetic studies because it is likely that the epigenetic status at different loci can exhibit different age dynamics. There are some other important questions that future studies should attempt to address. Why do only 35% of the twins in Fraga *et al.*'s study have detectable epigenetic differences in their methylated cytosine genomic content and histone H3 and H4 acetylation levels, and how do these twins differ from the remainder of the twin sample? Why do some older twins exhibit a greater density of methylated cytosines in Fraga *et al.*'s study (e.g. 50 year old twins had a greater density of methylated cytosines than 3 year old twins), whereas other studies showed that global DNA methylation decreases with age [12]? To what extent do the detected differences in the twins reflect genuine molecular epigenetic 'drift' rather than an epiphenomenon induced by white blood cell heterogeneity [13]?

The phenomenon of epigenetic differences has been well known in various isogenic organisms, which, like MZ twins, contain identical genomes [14]. Of particular interest are epigenetic studies of inbred animals, which have revealed links between partially stable, or metastable, epigenetic regulation and different phenotypic outcomes [15,16]. It has been shown that certain phenotypes, including various morphological traits and predisposition to tumors, are partially epigenetically inherited [17–19]. Without any specific environmental influence these inherited epigenetic patterns could be further modified by stochastic events affecting the performance of the DNA methylation maintenance and histone modification machinery of the organism. An example of such epigenetic stochasticity is the occasional failure of DNA methyltransferases to methylate some hemi-methylated cytosines following DNA replication, resulting in the non-identical epigenetic patterns of daughter chromosomes. In addition, DNA methyltransferases can exhibit *de novo* activity and can target unmethylated cytosines without any evident reason for doing so [20]. These observations suggest that, in some cases, epigenetic and phenotypic differences in MZ twins are possible without any specific environmental cause. The stochastic – but not environmental – scenarios of twin discordance could be more common than is generally assumed. In general, the increasing realization of the value and potential of epigenetics in phenotypic outcomes gives hope that the molecular mechanisms behind phenotypic differences in genetically identical organisms can be understood. The paradigmatic equation: P (phenotype) =G (genes) +E (environment) might require modification to P=G+E+ EpiG (epigenetics).

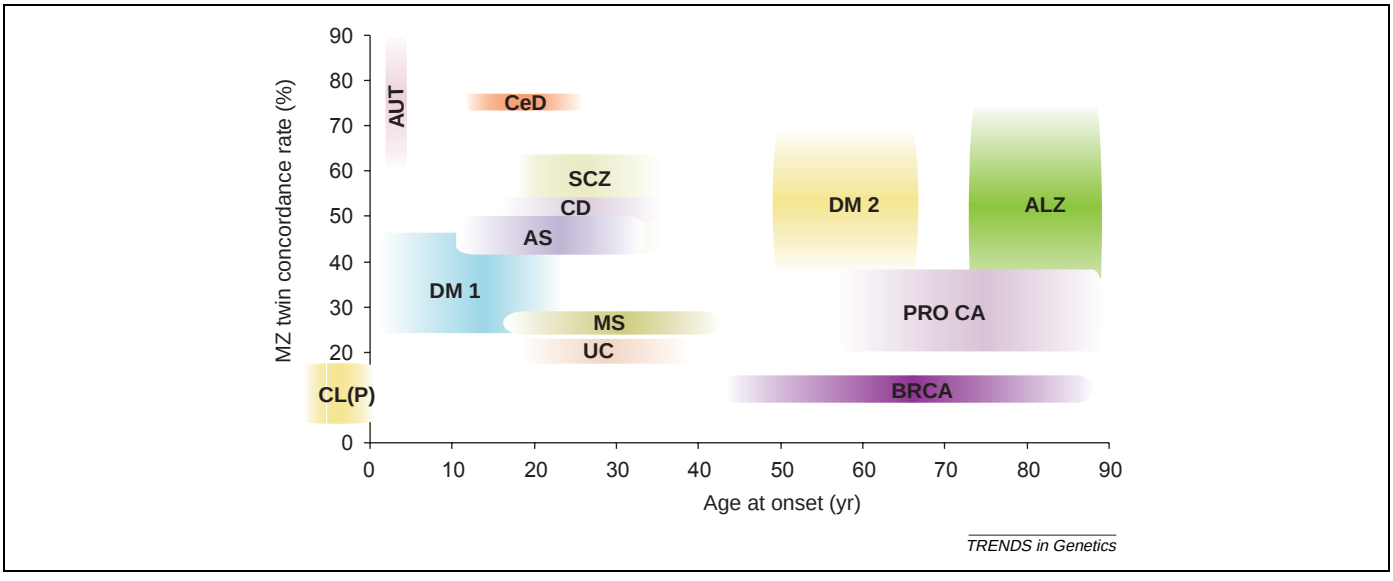

Figure 1. The age at disease onset and concordance rates of MZ twins. The left and right margins of each shape represent the range of age at onset, and the upper and lower margins represent the lowest and the greatest concordance rates of each disease. The data is taken from least two independent twin studies. Cleft lip and palate [CL(P)] occurs during embryogenesis and therefore placed to the left of the graph. Abbreviations: AUT, autism; CeD, celiac disease; SCZ, schizophrenia; CD, Crohn's disease; AS, asthma; DM1, type 1 diabetes mellitus; DM2, type 2 diabetes mellitus; MS, multiple sclerosis; UC, ulcerative colitis; BRCA, breast cancer; PRO CA, prostate cancer; ALZ, Alzheimer's disease. Primary references of twin concordance rates are available from the author.

Epigenetic similarities in twins

The other side of the 'coin' – epigenetic similarities in identical twins – has not been explored yet, although this is an equally interesting and important aspect of complex non-mendelian biology. Although traditionally all phenotypic similarities in MZ twins are explained by their DNA sequence identity, there is no reason to exclude epigenetic contributions. If epigenetic differences can result in different phenotypes, could epigenetic similarities be the molecular reason for some phenotypic similarities? As mentioned earlier, the Fraga *et al.* study showed that two-thirds of tested twins (65%) exhibited similar patterns in their epigenetic characteristics. Similarly, some cytosines in the genes encoding dopamine D2 receptor and catechol-o-methyltransferase also exhibit a high degree of symmetry in MZ twins [7,8].

It would be interesting to compare the epigenetic status of identical twins and non-identical twins – this type of comparison has been used in population heritability studies. A recent microarray-based global gene expression study detected a significant correlation on the expression levels of the actively regulated genes in both MZ and DZ twins [21]. Importantly, specific mRNA levels were more similar in MZ twins (by a factor of two) compared with those in DZ twins. If MZ twins also exhibit a greater degree of epigenetic similarity compared with DZ twins, two competing (but not necessarily exclusive) hypotheses are possible. The first hypothesis argues that larger epigenetic variation in DZ twins is caused by DNA sequence differences. The second hypothesizes that epigenetic inheritance causes difference: DZ twins exhibit more epigenetic differences than MZ twins do, because the former originated from different zygotes carrying two different epigenetic profiles, whereas the latter developed from the same zygote, and therefore had the same epigenetic profile. Although there might be no straightforward approach to proving or rejecting either of the two hypotheses, a key experiment can be performed on experimental animals. Several decades ago, Gartner and colleagues, while investigating genetic and environmental contributions, compared phenotypic outcomes in MZ (generated by splitting eight-cell embryos) and multizygotic offspring of inbred mice [22]. Because both MZ and fraternal offspring are inbred, their DNA sequences are identical. If MZ and non-MZ inter-class differences are detected in uniform environments, these must be due to epigenetic differences in the zygotes. Based on several examples of epigenetic inheritance [16,23,24], there is little doubt that such epigenetic heritability does exist – the question is how widespread in the genome and how universal across different species it is? If it is a relatively common phenomenon, further revisions of P = G + E might be required, in which the G component is split into DNA sequence-based inheritance and epigenetic inheritance.

Epigenetics and twins in a wider context: quantitative epigenetics

While discussing epigenetic mechanisms of induced heritable morphological alterations in *Drosophila melanogaster*, Rutherford and Henikoff introduced the term 'quantitative epigenetics' to acknowledge the role of epigenetic inheritance in complex traits, which are often called quantitative traits in model organisms [25]. Quantitative genetics, the 'older sib' of quantitative epigenetics, has been a dynamic field and in the past several decades thousands of quantitative trait loci (QTLs) have been identified, although only several dozen have been characterized to the specific nucleotide level [26]. Could the slow progress in QTL cloning be hampered by our ignorance about quantitative epigenetics? What proportion of QTL could be of epigenetic origin? For example, if the *agouti* locus [27], was analyzed as a standard QTL (i.e. limited to DNA sequence analysis without taking into account its epigenetic status), our understanding would be far from complete because the presence of the retrotransposon at the 5′ end of the gene is a poor predictor of the phenotype.

In human studies, the epigenetic analyses of twins should not be treated separately from the rest of non-mendelian biology. Numerous other features of complex traits – relatively late age of onset, sexual dimorphism, parental origin effects, fluctuations of phenotype intensity within an individual, in addition to dramatic changes in the population rates of some diseases (e.g. asthma, autism) over short periods of time – are consistent with epigenetic mechanisms [28]. In this context, epigenetics could emerge as a unifying concept for the large variety of non-mendelian features in complex traits. The advent of microarrays and their application to large-scale global epigenetic and gene expression profiling opens unlimited possibilities for the mapping of epigenetic changes without any *a priori* knowledge of their location. Epigenetic differences identified in MZ twins accounting for discordance of a specific phenotype can be further explored in singletons with respect to epigenetic changes induced by the environment, hormones and age. In addition, epigenetic findings in singletons can be further analyzed and used for an independent verification in discordant twins. A particularly important aspect of such studies in terms of the cause-and-effect relationship between the epigenetic change and phenotype is the comparison of MZ and DZ twins, which will estimate inherited and acquired epigenetic contributions.

Concluding remarks

Studies of epigenetic differences and similarities of twins bring a new perspective to the various phenomenological, clinical and molecular aspects of non-mendelian biology, new laboratory technologies and also a new hope that the mystery of complex traits can be 'unlocked' in the near future. Interestingly, twins have been used in genetic studies for more than half a century and some epigenetic factors (e.g. methylated cytosines) were identified long before the structure of DNA was discovered. Now these two 'oldies' (epigenetics and twins) are entering into a new alliance that once again might demonstrate that 'discovery consists not in finding new lands but in seeing with new eyes' (Marcel Proust).

Acknowledgements

I thank Irving Gottesman (University of Minnesota) and Jonathan Mill (Centre for Addiction and Mental Health, Toronto) for their comments

and suggestions. This research has been supported by the grants from NIH (R01 MH074127-01), Canadian Institutes for Health Research, Ontario Mental Health Foundation, NARSAD and the Stanley Foundation.

References

1 Taubes, G. (1995) Epidemiology faces its limits. *Science* 269, 164–169
2 Amasino, R. (2004) Vernalization, competence, and the epigenetic memory of winter. *Plant Cell* 16, 2553–2559
3 Ruden, D.M. *et al*. (2005) Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethylstibesterol on uterine development and cancer. *Hum. Mol. Genet.* 14, 149–155
4 Sutherland, J.E. and Costa, M. (2003) Epigenetics and the environment. *Ann. N. Y. Acad. Sci.* 983, 151–160
5 Weaver, I.C. *et al*. (2004) Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7, 847–854
6 Weksberg, R. *et al*. (2002) Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith–Wiedemann syndrome. *Hum. Mol. Genet.* 11, 1317–1325
7 Petronis, A. *et al*. (2003) Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr. Bull.* 29, 169–178
8 Mill, J. *et al.* 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.* (in press)
9 Fraga, M.F. *et al*. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10604–10609
10 Bergem, A.L. *et al*. (1997) The role of heredity in late-onset alzheimer disease and vascular dementia. A twin study. *Arch. Gen. Psychiatry* 54, 264–270
11 Kaprio, J. *et al*. (1992) Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia* 35, 1060–1067
12 Fuke, C. *et al*. (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann. Hum. Genet.* 68, 196–204
13 Martin, G.M. (2005) Epigenetic drift in aging identical twins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10413–10414
14 Matzke, M.A. and Matzke, A.J. (2000) Cloning problems don't surprise plant biologists. *Science* 288, 2318
15 Rakyan, V.K. *et al*. (2002) Metastable epialleles in mammals. *Trends Genet.* 18, 348–351
16 Blewitt, M.E. *et al*. (2004) How the mouse got its spots. *Trends Genet.* 20, 550–554
17 Cui, H. *et al*. (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 299, 1753–1755
18 Morgan, H.D. *et al*. (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318
19 Suter, C.M. *et al*. (2004) Germline epimutation of MLH1 in individuals with multiple cancers. *Nat. Genet.* 36, 497–501
20 Ushijima, T. *et al*. (2003) Fidelity of the methylation pattern and its variation in the genome. *Genome Res.* 13, 868–874
21 Tan, Q. *et al*. (2005) Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Hum. Genet.* 117, 267–274
22 Gartner, K. and Baunack, E. (1981) Is the similarity of monozygotic twins due to genetic factors alone? *Nature* 292, 646–647
23 Chong, S. and Whitelaw, E. (2004) Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* 14, 692–696
24 Rakyan, V. and Whitelaw, E. (2003) Transgenerational epigenetic inheritance. *Curr. Biol.* 13, R6
25 Rutherford, S.L. and Henikoff, S. (2003) Quantitative epigenetics. *Nat. Genet.* 33, 6–8
26 Flint, J. *et al*. (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* 6, 271–286
27 Argeson, A.C. *et al*. (1996) Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics* 142, 557–567
28 Petronis, A. (2001) Human morbid genetics revisited: relevance of epigenetics. *Trends Genet.* 17, 142–146

0168-9525/$ - see front matter
doi:10.1016/j.tig.2006.04.010

Genome-wide association: a promising start to a long race

David M. Evans and Lon R. Cardon

The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

A recent study by Cheung *et al.* demonstrates how to identify expression quantitative trait loci (eQTLs) underlying gene expression phenotypes through a combination of genome-wide linkage analysis and subsequent fine mapping or by genome-wide association (GWA) analysis. This study emphasizes the complexity of human traits, highlighting the challenges faced by investigators – in particular, insufficient linkage disequilibrium between the trait and marker variant, genetic heterogeneity and correcting for multiple testing will all adversely impact the power to detect loci by association. These issues must be considered carefully if the GWA approach is to succeed in mapping complex phenotypes.

Corresponding author: Cardon, L.R. (lon.cardon@well.ox.ac.uk).
Available online 19 May 2006

GWA analysis of gene expression levels in humans

Recently, after much anticipation, the first genome-wide association (GWA) studies in humans are beginning to appear in the literature [1,2]. Cheung and colleagues recently published the first GWA analysis of gene expression levels in a human population [3]. The idea behind their approach, which has been termed 'expression genetics' [4], is to subject levels of gene expression to the