

Epigenetics of Complex Diseases: From General Theory to Laboratory Experiments

A. Schumacher · A. Petronis (✉)

The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health,
Rm 28, 250 College St., M4T 1R8 ON, Toronto, Canada
Arturas_Petronis@camh.net

1	Introduction	2
2	Epigenetics and Complex Disease	3
2.1	Discordance of Monozygotic Twins, Environmental Impact, Stochasticity	4
2.2	Sex Effects and Critical Age	7
2.3	Parent-of-Origin Effects	8
2.4	Familiality and Sporadicity in Complex Disease	9
2.5	The Epigenetic Model of Complex Disease	10
3	Strategies for Detection of Epigenetic Differences in Complex Genomes	11
3.1	Methodological Issues in Epigenetic Studies of Complex Disease	11
3.2	Experimental Techniques in DNA Methylation Analysis	13
3.2.1	Bisulphite Modification-Based Approaches	13
3.2.2	Methylation-Sensitive Restriction Enzyme-Based Approaches	20
3.3	Methods for Large-Scale DNA Modification Analysis: Microarrays	23
4	Study of the Epigenetic Norm	25
	References	27

Abstract Despite significant effort, understanding the causes and mechanisms of complex non-Mendelian diseases remains a key challenge. Although numerous molecular genetic linkage and association studies have been conducted in order to explain the heritable predisposition to complex diseases, the resulting data are quite often inconsistent and even controversial. In a similar way, identification of environmental factors causal to a disease is difficult. In this article, a new interpretation of the paradigm of “genes plus environment” is presented in which the emphasis is shifted to epigenetic misregulation as a major etiopathogenic factor. Epigenetic mechanisms are consistent with various non-Mendelian irregularities of complex diseases, such as the existence of clinically indistinguishable sporadic and familial cases, sexual dimorphism, relatively late age of onset and peaks of susceptibility to some diseases, discordance of monozygotic twins and major fluctuations on the course of disease severity. It is also suggested that a substantial portion of phenotypic variance that traditionally has been attributed to environmental effects may result from stochastic epigenetic events in the cell. It is argued that epigenetic strategies, when applied in parallel with the traditional

genetic ones, may significantly advance the discovery of etiopathogenic mechanisms of complex diseases. The second part of this chapter is dedicated to a review of laboratory methods for DNA methylation analysis, which may be useful in the study of complex diseases. In this context, epigenetic microarray technologies are emphasized, as it is evident that such technologies will significantly advance epigenetic analyses in complex diseases.

1 Introduction

The possibility of understanding the molecular basis of human diseases is one of the most exciting perspectives of contemporary biomedical research. Since most (if not all) diseases exhibit inherited predisposition, significant research effort has been dedicated to identification of heritable risk factors. Despite major technological and computational developments, the progress in elucidation of aetiological DNA sequence variants in the overwhelming majority of human disease, primarily complex non-Mendelian disease, has been slow. The problems in understanding the molecular origin of complex diseases could be due to limitations in the current research strategy, which is focused primarily on DNA sequence variation (e.g. mutations, polymorphisms). As a rule, such DNA sequence variants are thought to be located in the coding or regulatory part of a gene, and this expectation originates from a series of discoveries in simple Mendelian disorders such as sickle cell anaemia, thalassaemia, phenylketonuria, Duchenne muscular dystrophy and cystic fibrosis. The idea of the essential role of DNA sequence variation has been generalized and extrapolated to the “fundamentally different group of diseases” (Risch 1990), namely complex non-Mendelian diseases. Complex diseases, unlike simple ones, exhibit irregular (non-Mendelian) mode of inheritance, discordance of monozygotic (MZ) twins, possible role of environmental factors, sexual dimorphism, a fluctuating course of disease and parental origin effects, among other features. Some methodological changes have been required in order to fit complex diseases into the already developed schemes of analyses. Such modifications basically consist of treating the genes as “predisposing” factors instead of “causative” factors and put some emphasis on environmental effects. It has to be admitted that to some extent the current paradigm has been successful in some complex diseases, especially familial cases, and a series of gene mutations has been identified in colon cancer, breast cancer and Alzheimer’s disease, to name a few. The overwhelming proportion of non-Mendelian pathology, however, remains unexplained. In this context, epigenetics—with its multifaceted role in regu-

lation of various genomic functions—arrives as a new frontier of molecular studies of complex disease. By definition, epigenetics refers to regulation of various genomic functions that are brought about by heritable, but potentially reversible changes in DNA modification (more specifically, methylation of cytosines) and chromatin structure (modifications of chromatin proteins such as histone acetylation, methylation or phosphorylation; Henikoff and Matzke 1997). Genes, even the ones that carry no mutations or disease predisposing polymorphisms, may be useless or even harmful if not expressed in the appropriate amount, at the right time of the cell cycle or in the right compartment of the nucleus. There is increasing evidence that cells can operate normally only if both DNA sequence and epigenetic components of the genome function properly. Thus far, the role of epigenetic factors has been primarily investigated in rare paediatric syndromes, such as Prader–Willi, Angelman (Nicholls 2000; Nicholls and Knepper 2001), Beckwith–Wiedemann (Feinberg 1999; Walter and Paulsen 2003; Weksberg et al. 2003) and Rett syndrome (Amir et al. 1999), and also the malignant transformation of cells in cancer (Baylin and Herman 2000; Jones and Laird 1999). This study describes the advantages of the epigenetic interpretation of common non-Mendelian complexities as well as epigenetic re-interpretation of a series of clinical and molecular findings in complex disorders. In the second part of the chapter, laboratory methods for DNA methylation analysis, which may be useful in the study of complex diseases, are reviewed and recommendations for their applications are provided.

2 Epigenetics and Complex Disease

The epigenetic theory of complex disease is based on three premises:

1. The epigenetic status of genes and genomes is far more dynamic in comparison to the DNA sequence and is subject to changes under the influence of developmental programs and the internal and external environment of the organism (Cooney et al. 2002; Sutherland and Costa 2003; Waterland and Jirtle 2003; Weaver et al. 2004). After mitotic division, the daughter chromosomes do not necessarily carry identical epigenetic patterns in comparison to the parental chromosomes. Over time, substantial epigenetic differences may be accumulated across the cells of the same cell line or the same tissue. In addition, quite significant epigenetic changes may occur even in the absence of evident environmental differences, i.e. due to stochastic reasons. In tissue culture, fidelity of maintenance methylation

in mammalian cells was detected to be between 97% and 99.9% and de novo methylation activity was as high as 3%–5% per mitosis (Riggs et al. 1998). It is important to note that epigenetic patterns are not established chaotically; there is a significant continuity of epigenetic patterns during mitotic divisions.

2. Some epigenetic signals can be transmitted along with DNA sequence across the germline generations, i.e. such signals exhibit partial meiotic stability. Although it has been generally accepted that, during the maturation of the germline, gamete's epigenetic signals are erased and new epigenetic profiles are established (Li 2002), it is now becoming clear that not all epigenetic signals are removed during gametogenesis, and epigenetically determined traits can be transmitted from one generation to another (Rakyan and Whitelaw 2003; Rakyan et al. 2001, 2002).
3. Epigenetic signals are critically important for the regulation of various genomic functions (Henikoff and Matzke 1997), and epigenetic misregulation may be as detrimental to an organism as are mutant genes. In addition to regulation of gene activity (Constancia et al. 1998; Ehrlich and Ehrlich 1993; Jones et al. 1998; Nan et al. 1998; Razin and Shemer 1999; Riggs et al. 1998; Siegfried et al. 1999), epigenetic factors play an important role in numerous other genomic functions (Bestor and al. 1994; Riggs and Porter 1996) including genetic recombination (Petronis 1996) and DNA mutability (Yang 1996).

The scientific value of the epigenetic model of complex disease lies in the possibility of integrating a variety of apparently unrelated data into a new theoretical framework, which provides the basis for new hypothesis and experimental approaches. The below overview is primarily based on the epigenetic re-analysis of various non-Mendelian irregularities in three major psychiatric diseases: bipolar disorder, schizophrenia and major depression.

2.1

Discordance of Monozygotic Twins, Environmental Impact, Stochasticity

Common phenotypic differences (discordance) in identical twins have been one of the hallmarks of complex non-Mendelian disease. For example, proband-wise, MZ concordance for major depression is 31% for male and 48% for female MZ twins (Kendler and Prescott 1999), 62%–79% in bipolar disorder (Bertelsen et al. 1977), and 41%–65% in schizophrenia (Cardno and Gottesman 2000) [for concordance rates in other diseases, in both MZ and dizygotic (DZ) twins, see Fig. 1]. The phenomenon of differential susceptibility to disease in genetically identical twins was identified decades

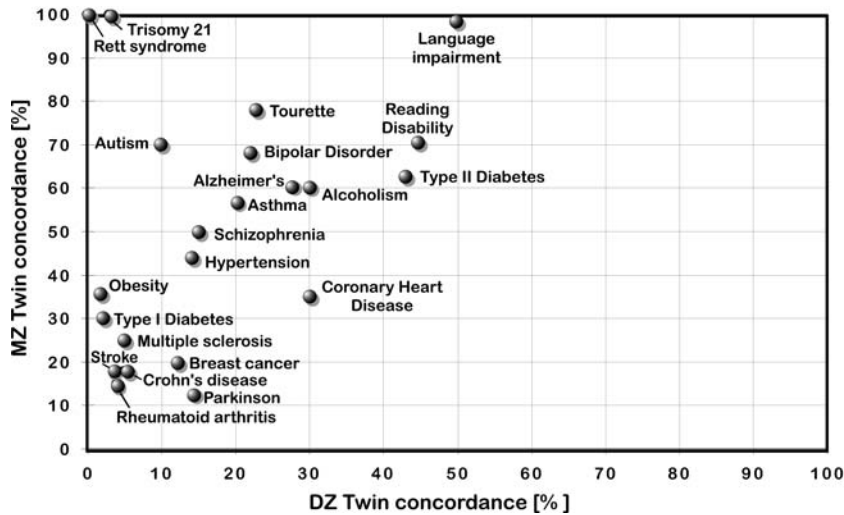


Fig. 1 Concordance of MZ and DZ twins for different disorders. As a rule, the degree of concordance in MZ twins is lower than 100% for nearly all complex diseases but substantially higher in comparison to the concordance rate in DZ twins

ago; however, the causes of such differences remain unknown. The traditional explanation for only one MZ twin having a clinical disease consists of the so-called “non-shared” environmental effects (Reiss et al. 1991), which supposedly produce disease in one of the two genetically predisposed co-twins. Several attempts to identify DNA sequence differences in MZ twins discordant for psychiatric diseases were carried out, but they did not detect any systemic DNA sequence differences in the tested twins (Deb-Rinker et al. 1999, 2002; Lavrentieva et al. 1999; Polymeropoulos et al. 1993; Tsujita et al. 1998; Vincent et al. 1998). Following the epigenetic model of complex disease (see Sect. 2.5 below), phenotypic differences in MZ twins result from their epigenetic differences. Due to the partial stability of epigenetic signals, a substantial degree of epigenetic dissimilarity can be accumulated over millions of mitotic divisions of cells in genetically identical organisms. This is well illustrated in inbred animals (Rakyan et al. 2002) and Beckwith–Wiedemann MZ twins (Weksberg et al. 2002), as well as in schizophrenia MZ twins (Petronis et al. 2003).

Epigenetic differences in identical twins may reflect differential exposure to a wide variety of environmental factors (that are very difficult to investigate directly) (Taubes 1995). There is an increasing list of such environmental factors that may have an impact on the epigenetic status of the genomes and individual

genes (Jablonka and Lamb 1995; Ross 2003; Sutherland and Costa 2003). Epigenetic changes induced by diet have been of particular interest. For example, intake of folic acid affects both the global methylation level in the genome and regulation of imprinted genes (Ingrosso et al. 2003; Wolff et al. 1998). “Street” drugs may also modify epigenetic regulation: Recent studies showed that methamphetamine that causes psychosis in humans alters DNA methylation as well as expression of genes in brain regions that are thought to be involved in schizophrenia (Numachi et al. 2004). This effect may be mediated via misregulation of DNA methylation enzymes, such as DNA methyltransferase, DNMT1, which was detected to be upregulated in the brain of schizophrenia patients (Veldic et al. 2004). During pregnancy, maternal dietary methyl supplements increase DNA methylation and change methylation-dependent epigenetic phenotypes in mammalian offspring (Cooney et al. 2002; Waterland and Jirtle 2003). Particularly interesting was the finding that pup licking and arched-back nursing by rat mothers altered the offspring’s DNA methylation and histone modifications at a glucocorticoid receptor gene promoter in the hippocampus (Weaver et al. 2004).

In addition, quite significant epigenetic changes may occur even in the absence of evident environmental differences, i.e. due to stochastic reasons (Fig. 2). After mitotic division, the daughter chromosomes do not necessarily carry identical epigenetic patterns in comparison to the parental chromosomes, and this takes place without any specific environmental input. Over time, substantial epigenetic differences may be accumulated across the cells

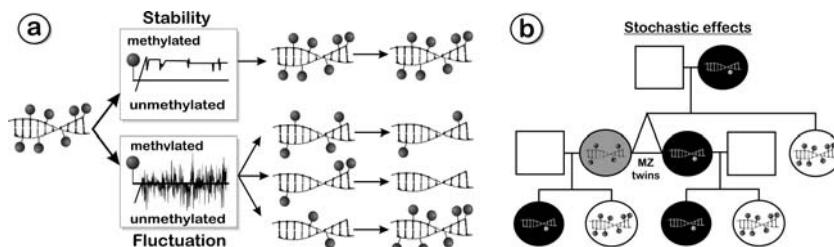


Fig. 2a, b Stochastic fluctuations in the methylation content of a genomic DNA fragment. **a** *Top*: Fully methylated regions may be particularly stable. The methylation level barely varies over time (*line*). *Bottom*: Epigenetic metastability of a short genomic region, indicated by fluctuations in the methylation level over time. The loss or gain of methyl groups at specific CpG dinucleotides results in different epigenotypes in the cells with identical genotypes. **b** An epimutation transmitted to identical twins causes the disease in only one co-twin (due to epigenetic differences in the brains), but the disease risk to the offspring of such discordant MZ twins is similar (due to the epigenetic similarities in their germline; Petronis 2004)

of the same cell line or the same tissue. As has been mentioned already, experiments performed in tissue culture (where the genomes were identical and environment was fully controlled), fidelity of maintenance methylation in mammalian cells was detected to be a few percentage points short of 100; plus, evidence for some degree of de novo methylation activity was identified (Riggs et al. 1998). Such stochastic events may add up over the numerous mitotic divisions in two identical twins and result in quite substantial epigenetic differences in some genes and genomic regions, which results in phenotypic discordance.

2.2

Sex Effects and Critical Age

One of the important peculiarities of complex disease is sexual dimorphism—differential susceptibility to a disease in males and females—that cannot be explained by genetic risk factors on the sex chromosomes. For example, women experience a lifetime episode of major depression about twice as often as men (Piccinelli and Wilkinson 2000). Additionally, data of longitudinal studies suggest that depressed women have longer episodes of depression than men and a lower rate of spontaneous remission (Weissman and Olfson 1995). The onset of the gender gap in depression occurs between the ages of 11 years and 13 years, when a precipitous rise in depression rates for adolescent girls far exceeds that in adolescent boys, and by 15 years of age females are twice as likely as males to have experienced a major depressive episode (Cyranski et al. 2000). These findings argue that changes in the endocrine milieu, as females progress from pre-puberty to post-puberty, might explain gender-linked increases in major depression (Warren and Brooks-Gunn 1989). In schizophrenia, the first episode usually occurs in one of three critical ages: adolescence/early adulthood, in the late 40s in women and in the sixth decade in both males and females (Howard et al. 2000), which evidently coincides with the periods of major hormonal rearrangements in the organism. Generally, intracellular effects of hormones are very consistent with the basic idea of epigenetic misregulation. Various hormones, including sex hormones, have an impact on epigenetic regulation. This is achieved by changing chromatin conformation (Csordas et al. 1986; Jantzen et al. 1987; Pasqualini et al. 1989; Truss et al. 1992), the local pattern of gene methylation (Saluz et al. 1986; Yokomori et al. 1995), or both. Hormone-induced epigenetic changes in critical genes may precipitate the onset of illness, and may also contribute to the differential susceptibility of the two sexes to complex diseases and the multiple age peaks seen in the onset of major psychosis. The findings of sex-specific effects in major depression, such as male-only linkage on 12q22-q23.2

(Abkevich et al. 2003) and female-only linkage on 2q33-35 (Zubenko et al. 2003), may be mediated via sex hormone-specific epigenetic modifications of the genes in these regions. Such findings argue that in some cases, to become a disease risk factor a gene (haplotype) must be epigenetically modified by oestrogens or androgens.

2.3

Parent-of-Origin Effects

In some psychiatric diseases, risk to offspring depends on the sex of the affected parent. For example, the risk of developing bipolar disorder is higher in the offspring whose mother is affected rather than the father (McMahon et al. 1995). Parent of origin-dependent clinical differences were also detected in schizophrenia (Crow et al. 1989; Ohara et al. 1997). Genetic linkage studies, although rarely performed in sex-specific fashion, also reveal parental origin effects in major psychosis (McMahon et al. 1997; Petronis et al. 2002; Schulze et al. 2003). One of the most common mechanisms of parent-of-origin effects is genomic imprinting (Hall 1990). The essence of genomic imprinting consists of the differential epigenetic modification of genes depending upon their parental origin (Barlow 1995). Disruption of the normal imprinting pattern often causes diseases that affect cell growth, development and behaviour (Pfeifer 2000). Animal studies investigating biological effects of genomic imprinting shed some new light on the impact of disrupted imprinting patterns on the development of the brain. Chimeric mice containing normal and uniparental cells have shown that parthenogenetic (Pg, complete maternal disomy) and androgenetic (Ag, complete paternal disomy) cells contribute differentially to specific regions of the brain (Allen et al. 1995; Schumacher 2001). In early development, Ag cells contribute substantially to those parts of the brain that are important for primary motivated behaviour (e.g. hypothalamus, septum and preoptic area) and proliferate extensively in the medio-basal forebrain. By contrast, Pg cells accumulate in the developing neocortex, striatum and telencephalic structures, where Ag cells are excluded (Keverne 1997). These results suggest an important role of epigenetic processes such as parent-of-origin effects or genomic imprinting in brain development. Such epigenetic events result in known aberrations of brain development. For example, Angelman syndrome—which presents with paroxysms of laughter, seizures, attention deficit, hyperactivity and aggressive behaviour—frequently shows anomalous cortical growth, resulting in cortical atrophy, microencephaly and ventricular dilation (Leonard et al. 1993; Schumacher 2001; Williams et al. 1989).

2.4 Familiality and Sporadicity in Complex Disease

While the explanation of all the above non-Mendelian features was based on the partial epigenetic stability in somatic cells, there is also an interesting perspective on the role and outcomes of the partial epigenetic stability during the maturation of the germline (Fig. 3). The meiotic epigenetic metastability allows for re-thinking on the issue of familiality (minor proportion of all cases) and sporadicity (overwhelming majority of the cases) in complex disease. Extrapolating from a single experimental finding of intergenerational dynamics of epigenetic regulation (Allen et al. 1990), it can be hypothesized that disease epimutations may develop in two possible ways: (1) regress towards the norm in the germline of an affected individual, and his/her offspring will not be affected, or (2) may persist across generations and become

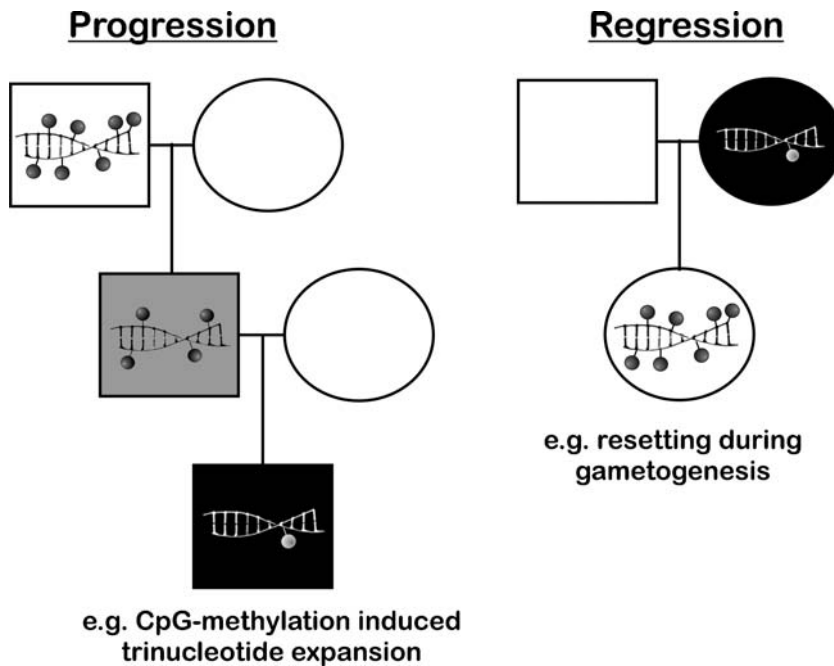


Fig. 3a, b Epigenetic perspective on the familial and sporadic cases of psychiatric disease. **a** Some epimutations may persist across generations and become even more pathogenic, which results in increased clinical severity and earlier age of onset. In some diseases it can occur that the symptoms get progressively worse every generation. **b** Other epimutations may regress towards the normal in the germline of a psychiatric patient, and his/her offspring will not be affected

even more pathogenic (Petronis 2004). Such meiotically persistent and progressing epimutations result in increasing clinical severity and earlier age of onset, which is characteristic of genetic anticipation (Petronis 2004). Genetic anticipation is a pattern of inheritance of genetic diseases where each successive generation seems to contract a more severe form of the genetic disease. Genetic anticipation has been widely investigated in psychiatric diseases (McInnis 1996; Petronis and Kennedy 1995) but is very difficult to prove due to various ascertainment biases (Heiman et al. 1996; Hodge and Wickramaratne 1995). Epigenetic studies of the intergeneration epigenetic dynamics relevant to the disease may also shed a new light on the issue of genetic anticipation.

2.5

The Epigenetic Model of Complex Disease

The epigenetic model of complex disease could be imagined as a result of a chain of aberrant epigenetic events that begins with a pre-epimutation, a primary epigenetic problem that takes place during the maturation of the germline. Yet, pre-epimutation increases the risk for the disease but is not sufficient to cause the disease per se. The epigenetic misregulation can be tolerated to some extent, and the age of disease onset may depend on the multidirectional effects of tissue differentiation, stochastic factors, hormones and probably some external environmental factors (nutrition, infections, medications, addictions, etc.; Jaenisch and Bird 2003; Sutherland and Costa 2003; Petronis 2004). It may take decades until the epigenetic misregulation reaches a critical threshold beyond which the cell (tissue, structure) is no longer able to function normally. The phenotypic outcome depends on the overall effect of the series of pre- and post-natal impacts on the pre-epimutation. Only some predisposed individuals will reach the “threshold” of epigenetic misregulation that causes the phenotypic changes that meet the diagnostic criteria for a clinical disorder (Fig. 4). Severity of epigenetic misregulation may fluctuate over time, which in clinical terms is called remission and relapse. In some cases, “ageing” epimutations may start slowly regressing back to the norm. For example, in major psychosis, this is seen as fading psychopathology or even partial recovery, which is consistent with age-dependent epigenetic changes in the genome (Fuke et al. 2004).

In conclusion, it can be argued that epigenetic mechanisms have the potential to explain a number of non-Mendelian features of complex disease. The advantages of the epigenetic scenario in comparison to the DNA sequence-based model is that the former is consistent with long years of ostensible health, critical susceptibility periods, fluctuating course and even clinical improvement after decades of being affected with a debilitating disease. The

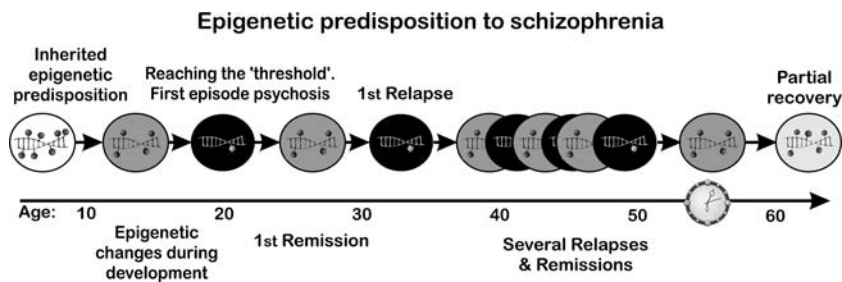


Fig. 4 Disease-development due to pre-epimutations. A pre-epimutation changes during development and may be influenced by various factors such as cell differentiation, intra- and extracellular environment, hormones and stochastic factors. Reaching a certain epigenetic threshold results in psychosis

epigenetic theory does not reject the role of DNA sequence variation but rather suggests that, in complex diseases, contribution of epigenetic factors may be substantial, and DNA sequence variation within genes should be investigated in parallel with the epigenetic regulation of genes.

3

Strategies for Detection of Epigenetic Differences in Complex Genomes

3.1

Methodological Issues in Epigenetic Studies of Complex Disease

The aforementioned theoretical speculations provide the basis for numerous working hypotheses of epigenetic differences between individuals affected with complex disease and controls. Despite significant heuristic value, epigenetic studies of complex disease are confounded by a number of factors which should be taken into account when designing experiments or interpreting data.

First, for an epigenetic study, specific tissues—the primary sites of the disease manifestation—are required, unlike the traditional genetic (linkage and association) studies where the cell/tissue source of the DNA sample is not critical. Disease-associated epigenetic differences are more likely to be detected in the disease-related tissue than the unrelated one, so in the case of major psychiatric disease, post-mortem brain samples are necessary.

Second, when dealing with complex organs such as brain, cellular heterogeneity should be taken into account as well, because there is little doubt that different types of neurons and glial cells may exhibit epigenetic differences—or colon epithelial cells versus the whole tissue. An ideal epigenetic experiment would investigate homogeneous cells collected by laser capture microdissec-

tion or fluorescence-activated cell sorting. However, the microdissection approach can only yield several hundred cells (and nanograms of DNA), which may limit the application of further methods (that require large amounts of DNA) for mapping of epigenetic differences.

Third, human tissues usually become available after a relatively long (12 h–30 h) post-mortem interval, which causes degradation of the nuclear epigenomic components, primarily histones. The same applies to the archived pathology samples, such as paraffin-embedded or formalin-fixed tissue sections. In comparison to histone modification, DNA modification is relatively stable (this is one of the main reasons why the below review of specific methods and the majority of epigenetic research to date is focused on DNA methylation analysis).

The fourth complexity is related to our ignorance of what specific genes and genomic loci could be aetiologically important in epigenetic studies of complex disease. It is not clear, for example, if in major psychosis the genes encoding dopamine or serotonin receptors (that are among the most popular molecules in psychiatric research) are really the best targets for epigenomic analyses. The focus on such genes may just be a reflection of our biased and superficial understanding of the etiopathogenic mechanisms of the disturbed brain.

Fifth, it is not clear what size of epimutation—the disease-specific epigenetic change—can be expected in complex disease: Is it a major difference—“black or white” case (such as seen in imprinted genes)—between the affected individuals and controls, or rather the more “shades of grey” scenario, where the epigenetic differences between the affected and unaffected subjects are rather subtle.

Sixth, detected associations of epigenetic changes and disease phenotype do not automatically imply the cause–effect relationship, as disease process can be caused not only by epigenetic changes but could be the cause of some epigenetic changes. Longitudinal studies, especially the ones that include the premorbid conditions (before the presentation of the clinical symptoms) are practically limited to experimental animals. In human studies, analysis of the tissues that are not involved in the disease process may provide some insights on the cause–effect relationship. The expectation is that at least some epigenetic changes can also be detected in other tissues, which may reveal tissue non-specific epigenetic changes reflecting events that took place before the tissues were formed, such as in early embryogenesis. Precedents for this type of research are evidenced by the epimutations at *IGF2* (Cui et al. 2003) in lymphocytes of colon cancer patients, and at *KCNQ1OT1* in the lymphocytes and skin fibroblasts of Beckwith–Wiedemann syndrome individuals (Weksberg et al. 2002).

3.2

Experimental Techniques in DNA Methylation Analysis

A myriad of techniques exist for the identification of methylated cytosines (Table 1). Despite their diversity, most of the analytical approaches can be divided into the bisulphite modification-based techniques and the ones that use methylation-sensitive restriction enzymes (MSREs).

3.2.1

Bisulphite Modification-Based Approaches

The bisulphite modification method has been one of the most significant developments in methylation analysis. The key advantage of this method is sensitivity, as the degree of methylation in each position of cytosines can be identified with great precision. Although there are different permutations of the bisulphite technology, all of them are based on the conversion of cytosine into uracil under conditions where 5-mC remains unaltered (Frommer et al. 1992). A number of published protocols differ in the way the chemical modification is performed; however, the approach remains one of the more demanding techniques of molecular biology. The most commonly encountered artefact arises due to the high salt concentrations in the bisulphite reaction, which favours reannealing of DNA and, in turn, inhibits the sulphonation. The incomplete conversion will then be detected as false methylated cytosines (Rein et al. 1998). In addition, a small portion of 5-methylcytosines (5-mC) may be converted to thymine, which results in false negatives (Thomassin et al. 1999). Furthermore, treatment with bisulphite, especially at high temperatures, leads to DNA degradation due to partial acid-catalysed depurination. Consequently, a high proportion of the template DNA is too fragmented to be analysed. This problem is important when only limited amounts of starting material is available, e.g. when using post-mortem brain samples, small amounts of body fluids or microdissected tissues. This predicament intensifies further if the remaining DNA fragments are lost during the subsequent purification (desalting), which has to be extremely stringent. Since sodium bisulphite is a very effective pH buffer, residual bisulphite will prevent the complete alkalization of the DNA solution during desulphonation. However, if the reaction intermediate uracil-sulphonates are not converted, any DNA polymerase will be unable to replicate the template. Additionally, due to the 3-dimensional nature of the single-stranded (ss)DNA template, it may occur that some cytosines are not converted because they are included in hairpin structures that contain double-stranded regions (Rother et al. 1995). When analysing hypermethylated sequences, this effect can be even more severe,

Table 1 Methylation profiling technologies

Category	Method	BS	MSR	PCR	Gene Diff	Global	Resolution ^a	Reference ^a
BS-based	Standard bisulphite genomic sequencing	•	-	•	-	-	High	Frommer et al. 1992
	RNaseI T1-MALDI-TOF	•	-	•	-	-	Medium	Schatz et al. 2004
	MethylQuant	•	-	•	-	-	Very high	Thomassin et al. 2004
	MethylLight/HeavyMethyl-MethylLight/ConLight-MSP	•	-	•	-	-	High	Cottrell et al. 2004; Eads et al. 2000; Rand et al. 2002
Pyrosequencing	Pyrosequencing	•	-	•	-	-	Very high	Dupont et al. 2004; Uhlmann et al. 2002
	Chloroacetaldehyde assay	•	-	-	-	•	Very low	Oakeley et al. 1999
MALDI-mass spectrometry	MALDI-mass spectrometry	•	-	•	-	-	Very high	Tost et al. 2003
	Quantitative analysis of methylated alleles (QAMA)	•	-	•	-	-	Medium	Zeschmigk et al. 2004
MSP-DHPLC	Methylation-specific PCR (MSP)	•	-	•	-	-	High	Herman et al. 1996
	MSP-DHPLC	•	-	•	-	-	High	Baumer et al. 2001
Primer extension and ion pair reverse phase HPLC	Methylation-sensitive single-nucleotide primer extension (Ms-SnuPE)	•	-	•	-	-	High	Gonzalzo and Jones 2002
	Primer extension and ion pair reverse phase HPLC	•	-	•	-	-	High	Matin et al. 2002
Methylation-sensitive dot blot assay (MS-DBA)	Quantitative real-time PCR	•	-	•	-	-	High	Lo et al. 1999
	Methylation-sensitive dot blot assay (MS-DBA)	•	-	•	-	-	Medium	Clement and Benhattar 2005
Microplate-based quantitative methylation assay (MANIC)	Microplate-based quantitative methylation assay (MANIC)	•	-	•	-	-	Low	Yamamoto et al. 2004

Table 1 (continued)

Category	Method	BS	MSR	PCR	Gene Diff	Global	Resolution ^a	Reference(s)
	Enzymatic regional methylation assay (ERMA)	●	-	●	●	-	Medium	Galm et al. 2002; Zhang et al. 2004
	Denaturing gradient gel electrophoresis (DGGE)	●	-	●	●	-	Very low	Guldberg et al. 2002
	In-tube melting curve analysis and melting curve methylation specific PCR (McMSP)	●	-	●	●	-	Very low	Akey et al. 2002; Guldberg et al. 2002; Worm et al. 2001
	Methylation-sensitive single-strand conformation analysis (MS-SSCA)	●	-	●	●	-	Low	Dobrovic et al. 2002
	Methylation-specific oligonucleotide microarray	●	-	●	●	●	High	Adorjan et al. 2002; Balog et al. 2002; Gitan et al. 2002
	Combined bisulphite restriction analysis (COBRA)	●	●	●	●	-	High	Xiong and Laird 1997
	Melting curve combined bisulphite restriction analysis (McCOBRA)	●	●	●	●	-	Medium	Akey et al. 2002
MSR-based	Arbitrarily primed (AP)-PCR/restriction fingerprinting (MSRF)	-	●	●	●	-	Low	Gonzalzo et al. 1997; Huang et al. 1997; Liang et al. 2002
	<i>Mcr</i> BC-PCR	-	●	●	●	-	Low	Chotai and Payne 1998
	MS-RDA/methylated CpG island amplification (MCA)-RDA	-	●	●	●	-	Low	Toyota et al. 1999; Ushijima et al. 1997
	Amplification of inter-methylated sites (AIMS)	-	●	●	●	●	Low	Frigola et al. 2002

Table 1 (continued)

Category	Method	BS	MSR	PCR	Gene Diff	Global	Resolution ^a	Reference ^a
	Methylation-sensitive amplicon subtraction (MS-AS)	-	•	•	-	-	Low	Mueller and Doerfler 2000
	MS-AFLP	-	•	•	-	-	Low	Xiong et al. 1999
	Electrochemical sensing LM-PCR	-	•	•	-	-	Low	Hou et al. 2003a
	MS-amplicon epigenetic-oligonucleotide microarray	-	•	•	•	•	High	Schumacher et al. 2006
	NotI-CODE microarrays	-	•	•	-	•	Very low	Li et al. 2002
	NotI-MS-AFLP microarray	-	•	•	•	•	Very low	Yamamoto and Yamamoto 2004
	Hypermethylated fraction amplicon microarray	-	•	•	•	•	Medium	Hatada et al. 2002; Yan et al. 2002
	Methylation-sensitive restriction (MSR)-Southern	-	•	-	-	•	High	Rein et al. 1998
	McrBC-Southern	-	•	-	-	•	Very low	Sutherland et al. 1992
	Restriction landmark genomic scanning (RLGS)	-	•	-	•	-	Very low	Hayashizaki et al. 1993
	Photocrosslinking oligonucleotide hybridization assay	-	•	-	-	-	Medium	Peoples et al. 2004
	Size fractionation MSR microarray	-	•	-	•	•	High	Tompa et al. 2002
Other	High-performance liquid chromatography (HPLC)	-	-	-	-	•	Very low	Kuo et al. 1980
	High-performance or micellar electrokinetic capillary electrophoresis (HPCE)/(MECE)	-	-	-	-	•	Very low	Fraga et al. 2002; Stach et al. 2003

Table 1 (continued)

Category	Method	BS	MSR	PCR	Gene Diff	Global	Resolution ^a	Reference(s)
	Thin-layer chromatography (TLC)	-	-	-	-	•	Very low	Schmitt et al. 1997
	SssI acceptance assay	-	-	-	-	•	Very low	Wu et al. 1993
	5 ^m C-antibody immunochemical assay	-	-	-	-	•	Low	Oakeley et al. 1997
	Nearest-neighbour [³ H-CH ₃]-analysis	-	-	-	-	•	Low	Hubrich-Kuhner et al. 1989
	Hydrazine (N ₂ H ₄)-sequencing	-	-	•	-	-	High	Rev. Thomassin et al. 1999
	Permanganate (MnO ₄ ⁻)-sequencing	-	-	•	-	-	High	Rev. Thomassin et al. 1999
	Ligation-mediated (LM)-PCR	-	-	•	-	-	High	Pfeifer et al. 1989
	Methyl-CpG binding domain (MBD) column assays/ICEAMP	-	-	-	•	-	Very low	Brock et al. 2001; Cross et al. 1994

BS, bisulphite conversion; MSR, methyl-sensitive restriction enzymes; gene, method is suitable to interrogate larger DNA stretches or whole genes; Diff, enables the direct genome-wide comparison of two test-samples without a priori knowledge of the target; global, enables assessment of the total methylation level of a genome ^aVery high, quantitative single CpGs; high, nucleotide level; medium, several CpGs; low, many CpGs; very low: global

since methyl groups stabilize the double-helix structure. Finally, the amplification of bisulphite-treated DNA may result in preferential amplification of either the methylated or unmethylated alleles. The “classical” bisulphite modification approach requires cloning and sequencing of the PCR products amplified from the bisulphite-treated DNA. Since the average number of such clones to be sequenced is at least 20, it is evident that this step makes the entire bisulphite modification approach a very labour-intensive procedure. Several approaches that by-pass the cloning and sequencing step by means of direct interrogation of C/T ratio (^{met}C/C before bisulphite modification) in the amplicon have been developed over the last decade: for example, pyrosequencing and methylation-sensitive single-nucleotide primer extension.

Pyrosequencing This method is based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each deoxynucleotide (dNTP) upon DNA-chain elongation. In the first step, a primer is hybridized to a bisulphite-treated single-stranded, PCR-amplified template DNA and incubated with dNTP in the presence of exonuclease-deficient Klenow DNA polymerase (see Fig. 5a). Nucleotides are sequentially added to the reaction mix in a predetermined order. If a nucleotide is complementary to the template base and thus incorporated, PPi is released. PPi is then used as a substrate by an ATP sulphurylase, which converts quantitatively adenosine 5'-phosphosulphate (APS) to adenosine triphosphate (ATP). This ATP drives a luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The emitted light is detected by a charge-coupled device (CCD) and finally results in a peak, indicating the number and type of nucleotide incorporated in the form of a pyrogram.

Pyrosequencing can be applied to quantitate several CpG dinucleotides in one reaction and it works well with very short PCR fragments (Dupont et al. 2004; Uhlmann et al. 2002). This is particularly important in the analysis of samples that may contain moderately degraded DNA, such as paraffin-embedded tissues or post-mortem brain-samples. Pyrosequencing is a relatively rapid analysis once bisulphite-converted DNA is prepared. Additionally, cytosines outside of the analysed CpG positions can be used as internal controls to confirm completion of the bisulphite treatment. Nevertheless, an inherent problem with the described method is sequencing of polymorphic regions in heterogeneous DNA material. Polymorphisms may cause the sequencing reaction to become out of phase, making the interpretation of the succeeding sequence difficult. Another problem is the difficulty in determining the number of incorporated nucleotides in homopolymeric regions, which are present at a high frequency in bisulphite-treated DNA, due to the

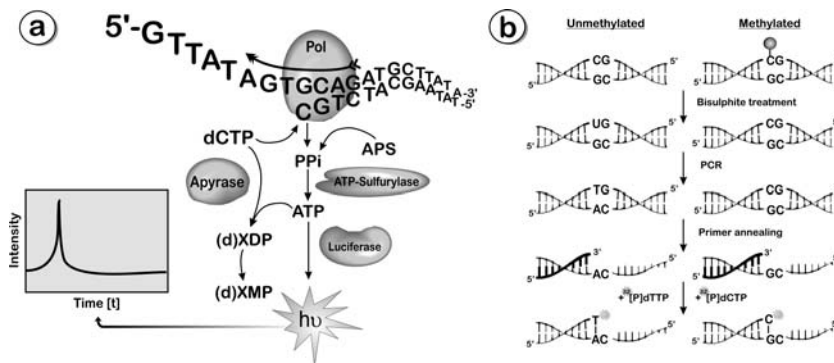


Fig. 5a, b Fine-mapping and confirmation of methylation abnormalities. **a** Pyrosequencing of bisulphite DNA. In this reaction, dCTP (the nucleotide added) is incorporated by a DNA polymerase complementary to the next unpaired nucleotide (G) on the bisulphite-treated template DNA. The pyrophosphate (PP_i) released is converted to ATP and then to a light signal via an enzyme cascade, including ATP-sulphurylase and luciferase. Before the addition of the next nucleotide starts, any excess of nucleotide and ATP is degraded by apyrase, which regenerates the reaction solution. **b** Schematic outline of methylation-sensitive single-nucleotide primer extension (MS-SNuPE). After bisulphite treatment, the DNA is amplified with a strand-specific primer that does not discriminate between methylated and unmethylated alleles. Then, a primer is annealed upstream of the target sequence, immediately terminating 5' to the original CpG. Finally, a single-nucleotide extension reaction is performed using radioactively or fluorescently labelled triphosphates, and the incorporation of the nucleotides is quantified

nonlinear light response following incorporation of more than 5–6 identical nucleotides.

Methylation-Sensitive Single-Nucleotide Primer Extension This PCR-based technique allows the quantitative analysis of several CpG sites in parallel. After bisulphite conversion, the region of interest is amplified by strand-specific PCR (Fig. 5b). The MS-SNuPE assay utilizes internal primers that anneal to the amplified template and terminate immediately 5' of the single nucleotide(s) to be assayed. The annealed bisulphite-specific primers, which should not preferentially discriminate between methylated and unmethylated alleles, are extended by a DNA polymerase. Quantitation of the ratio of methylated versus unmethylated cytosine (C versus T) at the original CpG sites can then be determined by incubating the PCR product, primer(s) and DNA polymerase with either [32 P]dCTP or [32 P]TTP or fluorescent-labelled nucleotides followed by denaturing polyacrylamide gel electrophoresis or quantitation in a sequencer. Opposite-strand Ms-SNuPE primers can also

be designed which would incorporate either [³²P]dATP or [³²P]dGTP. The amount of methylation at multiple CpG sites can be analysed in a single reaction by using a multiplex oligonucleotide strategy without the need for restriction enzymes. Advantages of the Ms-SNuPE technique are that it can be performed in a quantitative manner and with a small amount of starting material. However, it is likely that cytosines which are located close to a CpG dinucleotide of interest do affect the results (Kaminsky et al. 2005). This problem is particularly important when analysing CpG islands. Until recently, many such regions have been investigated using Ms-SNuPe, but CpG dinucleotides in the area of primer design had to be avoided (Dahl and Guldberg 2003), which is a noticeable limitation of the technology. Recent attempts have been made to overcome the problem of Ms-SNuPe primer mismatch effects in order to interrogate CpG sites independent of sequence context, including GC-rich regions, using matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Tost et al. 2003).

3.2.2

Methylation-Sensitive Restriction Enzyme-Based Approaches

Methylation sensitive restriction enzymes were first applied to epigenetic studies at least three decades ago and for a long time were the primary tools for DNA methylation analysis until the fine mapping using the bisulphite modification approach was developed. The interest in MSREs is now resurging as these enzymes are the key tools for large-scale epigenomic profiling using microarrays (see Sect. 3.3). Classical examples of methods using MSREs are MS-Southern and the restriction landmark genomic scanning (RLGS), a method that was used to detect genomic regions with alterations in DNA methylation associated with tumourigenesis (Hayashizaki et al. 1993). RLGS employs direct end labelling of the genomic DNA digested with a methylation-sensitive restriction enzyme (usually *NotI*) and two-dimensional gel-electrophoresis. The status of DNA methylation can then be determined by monitoring the appearance or disappearance of spots in the gel. Other methods use the restriction endonuclease *McrBC* to compare DNA sample pairs (Chotai and Payne 1998; Sutherland et al. 1992). *McrBC* cleaves DNA containing ^mC on one or both strands but will not act upon unmethylated DNA. *McrBC* will act upon a pair of Pu^mCG sequence elements, thereby detecting a high proportion of methylated CpGs, but will not recognize *HpaII/MspI* sites (5'-C^mCGG-3') in which the internal cytosine is methylated. *McrBC* digestion was used as a diagnostic test for Prader-Willi and Angelman syndromes based on differential digestion of repressed (maternally imprinted) *SNRPN* sequences by *McrBC*, followed by PCR amplification of the *SNRPN* promoter (Chotai and Payne 1998).

Over 250 different methylation-sensitive restriction enzymes (including isoschizomers) are now available (see also Table 2), but which of these enzymes are really useful and informative for methylation profiling? Informative MSREs are defined by the number of cleavage fragments in the range of approximately 75 bp to 2,000 bp that can be ligated to adaptors and efficiently amplified, and are not lost during column-purification steps. Some enzymes, although they cut frequently in the genome, produce fewer informative fragments compared to other enzymes that do not cut as frequently. For example, the non-palindromic *Acil* (5'-CCGC-3') recognizes more than twice as many CpG sites in CpG island regions compared to *HpaII*, but on the other hand produces fewer fragments in the size range that can be detected by PCR or amplified fragment length polymorphism (AFLP) methods (Table 2). Other important enzyme features are their digestion- and ligation-efficiency, non-specific ("star") activity (e.g. *Eco72I*), costs and alternate CpG recognition sequences. An example for the latter is *TauI* (5'-GC^C/G_CGC-3') that covers approximately 11% of all CpG dinucleotides in CpG islands; however, the results can be ambiguous, since this enzyme recognizes two different CpG-containing sequences. Many other enzymes might be useful for specific purposes, but may be exchanged with enzymes of higher CpG coverage. For example, *Kpn2I* has the recognition sequence 5'-TCCGGA-3', which is already covered by the 4-base cutter *HpaII* (5'-CCGG-3'). Other MSREs, such as *Fnu4HI* (5'-GCNGC-3'), will also cut sequences that do not contain a CpG dinucleotide, hence they are relatively inadequate in methylation analysis. All of these requirements for the enzyme reduces the list of potentially useful and informative MSREs to about 17 (Table 2), which would cover up to 85% of all CpG island CpG dinucleotides but less than 50% of all CpG dinucleotides in other genomic regions. The number of CpG sites that could be interrogated would even increase dramatically if a methylation-sensitive enzyme was developed that could cut the palindromic 4-base sequence 5'-TCGA-3'.

To gain the most out of restriction analyses, it is crucial to choose the right enzyme combination for the targets to be interrogated. For example, some MSREs cut relatively frequently in CpG islands but rarely recognize a sequence outside of a CpG island region, as it is the case for *Him6I* (5'-GCGC-3') or *Bsp143II* (5'-PuGCGCPy-3'). In contrast, enzymes such as *HpyCH4IV* (5'-ACGT-3') cut predominantly outside of CpG island sequences and are less useful in the interrogation of CpG islands, for instance in CpG island microarray-based studies (see Sect. 3.3). Several other methods rely on the specific methylation-sensitive cleavage of the rare cutter *NotI* (5'-GCGGCCGC-3'), for example RLGS and AFLP methods and a couple of microarray approaches (Li et al. 2002; Yamamoto and Yamamoto 2004). However, *NotI*-sites are not well represented in the genome and will only provide a very rough overview

Table 2 Methylation-sensitive restriction enzymes

MSRE	Cut site 5'-3'	Overhang ^b	CpGs in CpG islands ^a (~%)	CpGs in non-CpG islands ^a (~%)	Fragments/kb in CpG islands ^a	Fragments/kb in non-CpG islands ^a
<i>AclI</i> (<i>SstI</i>)	CCGC, GCGG	5'-CG-3'	30.60	17.36	3.23	1.79
<i>Hin6I</i> (<i>HinP1I</i>)	GCGC	5'-CG-3'	14.40	5.05	3.98	0.61
<i>Bsh1236I</i> (<i>AccII</i>)	GCGG	Blunt end	12.06	2.45	3.57	0.25
<i>HpaII</i> (<i>BsiSI</i>)	CCGG	5'-CG-3'	11.70	9.33	3.98	1.18
<i>TauI</i>	GCSGC	3'-GGC-5'	11.30	2.07	3.25	0.18
<i>Hin1I</i> (<i>BsaHI</i>)	GRCGYC	5'-CG-3'	2.6	0.9	1.92	0.11
<i>Bsh1285I</i> (<i>BsaO</i>)	CGRYCG	3'-Py-Pu-5'	2.4	<0.05	1.82	<0.02
<i>Eco52I</i> (<i>EagI</i>)	CGGGCG	5'-GGCC-3'	1.7	<0.05	1.17	<0.02
<i>HpyCH4IV</i> (<i>TaiI</i>)	ACGT	5'-CG-3'	1.66	6.73	1.24	0.97
<i>Esp3I</i> (<i>BsmBI</i>)	CGTCTC	Blunt end	1.3	1.93	0.59	0.18
<i>CpoI</i> (<i>CspI</i>)	CGGWCCG	5'-GAC-3'	0.24	<0.05	0.1	<0.02
<i>Bsp119I</i> (<i>AsuII</i>)	TTCGAA	5'-CG-3'	0.11	0.13	0.11	<0.02
<i>PvuI</i> (<i>BspCI</i>)	CGATCG	3'-TA-5'	0.11	0.13	0.02	<0.02
<i>SalI</i>	GTCGAC	5'-TCGA-3'	0.09	0.26	<0.02	<0.02
<i>Bsu15I</i> (<i>ClaII</i>)	ATCGAT	5'-CG-3'	<0.05	0.39	<0.02	0.02
<i>NotI</i> (<i>CciNI</i>)	GCGGCCGC	5'-GGCC-3'	<0.05	<0.05	0.1	<0.02
<i>Pfl23II</i> (<i>SunI</i>)	CGTACG	5'-GTAC-3'	<0.05	0.13	<0.02	<0.02

R=A/G; Y=C/T; W=A/T; S=C/G ^aThe number of 75-bp- to 2-kb-long ("informative") fragments, derived from several Mbp of randomly selected CpG island and non-CpG island sequences on chromosomes 1, 2, 4, 5, 6, 9, 17, 19 and 20 ^bThe isochizomers may produce different overhangs

of methylation patterns. Hence, it is not advisable to include *NotI* in genome-wide analyses of complex diseases. All the above MSRE aspects are directly relevant to their application in the large-scale high-throughput microarray-based DNA methylation profiling.

3.3

Methods for Large-Scale DNA Modification Analysis: Microarrays

Microarrays constitute a significant advance in methylation analysis of complex disease because they may interrogate a very large number of loci in a highly parallel fashion. The principle of “epigenomic” microarrays is the same as in other kinds of arrays: Fluorescently labelled fragments of the tested nucleic acids hybridize to the complementary DNA sequences on the microarray, and intensity of fluorescent signal at each specific spot represents the amount of a specific fragment in the tested sample. Thus far, enzyme-based “epigenomic” microarray approaches have focussed predominantly on the enrichment and analysis of the hypermethylated fraction of the genome. This technology was used in several studies for the identification of abnormally methylated CpG islands in tumour cells (Hatada et al. 2002; Hou et al. 2004; Huang et al. 1999; Shi et al. 2003b; Yan et al. 2002). Using the hypermethylated DNA fragments for methylation analyses seems to be practical for detection of major epigenetic changes in some regions of the genome. However, the overall proportion of CpG dinucleotides that can be interrogated is substantially lower compared to a potential analysis using the unmethylated DNA fraction. Also, unmethylated cytosines represent a much smaller part of cytosines in comparison to the methylated one (depending on the tissue, over 70% of cytosines in the human genome are methylated). Analysis of this smaller unmethylated fraction is more sensitive to detect subtle methylation abnormalities. For example, if 20% of all CpGs in a given tissue are unmethylated, a *de novo* methylation of 10% would result in 100% (decrease of from 20% to 10%) difference in the unmethylated fraction. In the same scenario, only a 12% change (from 80% to 90%) would be detected for the hypermethylated fraction of genomic DNA. An approach using the hypomethylated DNA fraction was suggested (Tompa et al. 2002), where a fragmentation by a methyl-sensitive restriction endonuclease is followed by a sucrose gradient size-fractionation. The small fragments (<2.5 kb) will predominantly contain hypomethylated fragments and can then be labelled and hybridized to microarrays (Fig. 6c). In the original protocol, *MspI* was used for DNA cleavage; however, this enzyme is only blocked by methylation of the outer cytosine in the 5'-CCGG-3' sequence, a form of methylation that is encountered in plants but usually not in humans. Hence, to

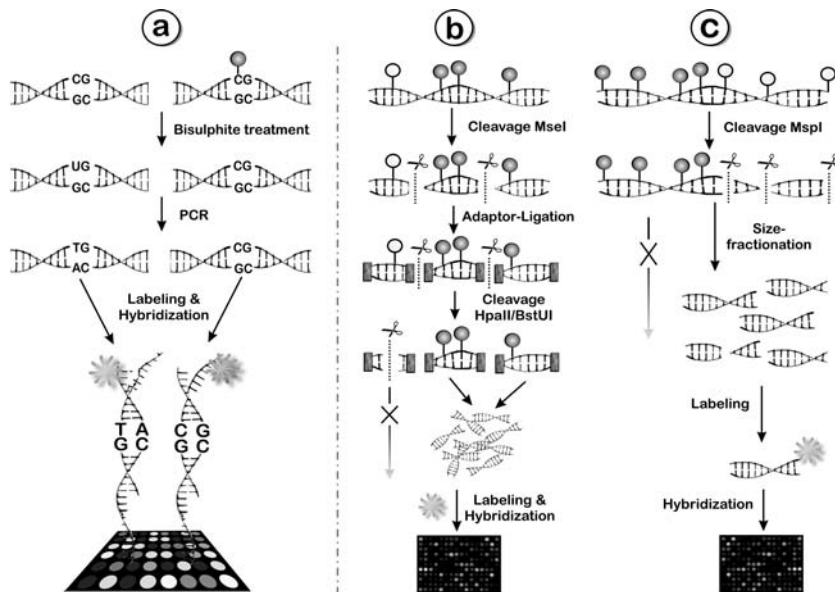


Fig. 6a–c Microarray strategies for methylation profiling. **a** Typical bisulphite approach. An amplified bisulphite-treated sample is hybridized to a set of oligonucleotides (19–25 nucleotides in length) that discriminate methylated and unmethylated cytosine at specific nucleotide positions, and quantitative differences in hybridization are determined by fluorescence analysis. **b** Restriction-based approach that uses the hypermethylated fraction of the genome. Tester and control are cleaved and adaptors specific for the cut-sites are ligated to the fragments. Unmethylated sequences are eliminated by cleavage with MSREs. The remaining hypermethylated fragments are labelled and hybridized to a microarray. **c** After methylation-sensitive cleavage, small fragments (<2.5 kb; mostly unmethylated) are size-fractionated in a sucrose gradient, labelled and hybridized

apply this technique for the study of complex human disease, the endonuclease has to be replaced by another MSRE, such as *HpaII* or *Acil* (see Table 2).

For detailed DNA methylation profiling, high-resolution oligonucleotide arrays are recommended, for example microarrays that are based on 25-nucleotide perfect match–mismatch oligomers that have been generated by Affymetrix for transcriptome studies (Kapranov et al. 2002). At this time, microarrays that cover all the non-repetitive regions of human chromosomes 21 and 22 and the regions selected for the ENCODE project (<http://www.genome.gov/page.cfm?pageID=10005107>) of the human genome are commercially available. There is a very good chance that in the next

5 years high-resolution oligonucleotides-based microarrays for the entire human genome will be manufactured.

Since restriction enzymes are used in many methylation assays, DNA sequence variation (single-nucleotide polymorphisms, SNPs) may simulate epigenetic differences. However, until now most methods used in epigenetic studies have not been differentiating between methylation changes and the presence of SNPs within the restriction sites of the applied restriction enzymes. In order to exclude the impact of DNA sequence variation, it is suggested to check the available SNP databases and identify the DNA sequence variation within the restriction sites of the used enzymes. From CpG island microarray studies, the estimate is that 10% to 30% of methylation variation detected between individuals could be in fact due to DNA sequence variation (Schumacher et al. 2006). For comparison, in a pilot study for the Human Epigenome Project (HEP), interrogation of 3,273 unique CpG sites within the human major histocompatibility complex (MHC) on chromosome 6 revealed that 101 CpGs overlapped with known SNPs, all representing sites at which the CpG was lost (Rakyan et al. 2004).

Microarrays can also be used to interrogate C→T transitions in bisulphite-modified DNA sequences (Adorjan et al. 2002; Balog et al. 2002; Gitan et al. 2002; Hou et al. 2003a, b; Shi et al. 2003a, b). Bisulphite arrays contain oligonucleotides that measure the C(G)/T(A) ratio in the bisulphite-treated DNA, which correspond to ^{met}C/C in the native DNA (Fig. 6a). Bisulphite-based microarray technologies have the advantage that they are not limited to specific recognition sequences, as in cleavage-based approaches. However, although informative and precise, microarrays can contain only a limited number of oligonucleotides because treatment with bisulphite degenerates the four-nucleotide code, which results in the loss of specificity of a large portion of the genome. For example, after bisulphite treatment all of the possible 16 permutations of a four-base sequence containing unmethylated C and T (CCCC, CTCT, CCCT, CCTT, TCTC, TTTC, TTTT, etc.) will become identical TTTT (Fig. 7). This degeneration will predominantly affect unmethylated regions, such as CpG islands.

4 Study of the Epigenetic Norm

In addition to human morbid epigenetics, research of the “normal” epigenome is also of significant interest, as such information may be crucial in understanding of the molecular mechanisms of development, ageing, tissue specificity and sex differences, among other systemic aspects in human biology.

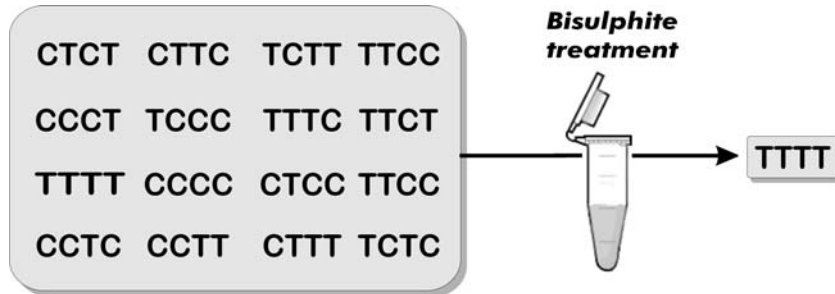


Fig. 7 After bisulphite treatment, all of the possible 16 permutations of a four-base DNA sequence containing unmethylated C and T will become identical

Documentation of normal epigenome patterns, however, is not a trivial task.

What would it be to accomplish a comprehensive annotation of the human epigenome? A conservative approach would require epigenomic profiling of DNA methylation and various kinds of histone modifications (at least 10 types) of roughly 16 million nucleosomes (as the basic structural unit of chromosomes) in approximately 260 different cell-types in the human body, at, let us say, approximately 20 different time-points (from the zygote stage through embryogenesis and post-natal development, adolescence, youth, adulthood and ageing) in 100 males and 100 females, each measurement performed in duplicate. Taken all together, one would generate $11 \times 260 \times 20 \times 200 \times 16 \times 10^6 \times 2 = 3.6 \times 10^{14}$ data points (bits). Each data point has to be referenced, which means that the chromosomal location, for example, of the modification, its time point in development, the kind of modification (methylation or acetylation) and so on, have to be stored along with it. For the storage of this raw data alone, one would require at least 0.37 PB (petabytes) of memory capacity, which is equivalent to several hundred average computer systems today. Although it is difficult to imagine how all this epigenetic information can be processed, it is very likely that there are numerous levels of redundancy (such as hypermethylated DNA regions will usually exhibit histone hypoacetylation, cells originating from the same stem should exhibit numerous epigenetic similarities).

An analogy with the game of chess illustrates the possible reduction in information content to be processed. Chess is known to have an infinite number of possible positional combinations; however, in praxis the number is finite, since specific positions and combinations of pieces would be illegal (e.g. the king cannot move into check). In fact, a “mere” 2×10^{46} moves (roughly) are theoretically possible, a number that can be mathematically approached. Out of such subsets of data, “normal” patterns, and eventually the collective

behaviours of the system, and the algorithms of the system's interaction with its environment, can be identified.

In the field of complexity theory, statistical approaches that reduce the data without losing the essential features and characteristics of the system have been developed. For example, an average "behaviour" of a large number of components can be considered rather than the "behaviour" of any individual component (e.g. the co-operation of DNA methyl groups with histone modifications), drawing heavily on the laws of probability and aiming at predictions of larger systems on the basis of the properties of their single constituents. Additionally, not all combinations of epigenetic components are unique; there are patterns present in the arrangements that allow us to classify and filter many combinations in the same way. There are also several practical interrelated approaches based on heuristic functions to study a complex system, which do not rely on static algorithms and pre-defined ideas. Heuristic approaches are self-learning or adaptive processes, based on empirical information intended to increase the probability of solving a problem. For example, "heuristic programming" would approach the problem of finding epigenetic patterns by a method of trial and error in which the success of each attempt at solution is used to improve the subsequent attempts, until a solution acceptable within defined limits is reached.

A good starting point in gathering and interpreting epigenetic data would be to understand ways of describing complex systems (especially the need for a uniform epigenetic nomenclature for multicomponent data). Second, we have to understand the interactions of the components giving rise to the pattern of behaviour and the process of formation of epigenetic information (within the cell and on the evolutionary scale). Ultimately, it is unlikely that the human epigenomic databases will consist of traditional raw data; rather, it will be user-friendly profiles, diagrams, and equations describing developmental, intra- and inter-individual variation, and epigenetic "plasticity". Altogether, this effort will provide a much "loftier view of life" (Beck et al. 1999).

References

- Abkevich V, Camp NJ, Hensel CH, et al (2003) Predisposition locus for major depression at chromosome 12q22–12q23.2. *Am J Hum Genet* 73:1271–1281
- Adorjan P, Distler J, Lipscher E, et al (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res* 30:e21
- Akey DT, Akey JM, Zhang K, Jin L (2002) Assaying DNA methylation based on high-throughput melting curve approaches. *Genomics* 80:376–384
- Allen ND, Norris ML, Surani MA (1990) Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* 61:853–861

- Allen ND, Logan K, Lally G, et al (1995) Distribution of parthenogenetic cells in the mouse brain and their influence on brain development and behavior. *Proc Natl Acad Sci U S A* 92:10782–10786
- Amir RE, Van den Veyver IB, Wan M, et al (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* 23:185–188
- Balog RP, de Souza YE, Tang HM, et al (2002) Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal Biochem* 309:301–310
- Barlow DP (1995) Gametic imprinting in mammals. *Science* 270:1610–1613
- Baumer A, Wiedemann U, Hergersberg M, Schinzel A (2001) A novel MSP/DHPLC method for the investigation of the methylation status of imprinted genes enables the molecular detection of low cell mosaicisms. *Hum Mutat* 17:423–430
- Baylin SB, Herman JG (2000) DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* 16:168–174
- Beck S, Olek A, Walter J (1999) From genomics to epigenomics: a loftier view of life. *Nat Biotechnol* 17:1144
- Bertelsen A, Harvald B, Hauge M (1977) A Danish twin study of manic-depressive disorders. *Br J Psychiatry* 130:330–351
- Bestor TH, Chandler VL, Feinberg AP (1994) Epigenetic effects in eukaryotic gene expression. *Dev Genet* 15:458
- Brock GJ, Huang TH, Chen CM, Johnson KJ (2001) A novel technique for the identification of CpG islands exhibiting altered methylation patterns (ICEAMP). *Nucleic Acids Res* 29:E123
- Cardno AG, Gottesman II (2000) Twin studies of schizophrenia: from bow-and-arrow concordances to Star Wars Mx and functional genomics. *Am J Med Genet* 97:12–17
- Chotai KA, Payne SJ (1998) A rapid, PCR based test for differential molecular diagnosis of Prader-Willi and Angelman syndromes. *J Med Genet* 35:472–475
- Clement G, Benhattar J (2005) A methylation sensitive dot blot assay (MS-DBA) for the quantitative analysis of DNA methylation in clinical samples. *J Clin Pathol* 58:155–158
- Constancia M, Pickard B, Kelsey G, Reik W (1998) Imprinting mechanisms. *Genome Res* 8:881–900
- Cooney CA, Dave AA, Wolff GL (2002) Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* 132:2393S–2400S
- Cottrell SE, Distler J, Goodman NS, et al (2004) A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res* 32:e10
- Cross SH, Charlton JA, Nan X, Bird AP (1994) Purification of CpG islands using a methylated DNA binding column. *Nat Genet* 6:236–244
- Crow TJ, DeLisi LE, Johnstone EC (1989) Concordance by sex in sibling pairs with schizophrenia is paternally inherited. Evidence for a pseudoautosomal locus. *Br J Psychiatry* 155:92–97
- Csordas A, Puschendorf B, Grunicke H (1986) Increased acetylation of histones at an early stage of oestradiol-mediated gene activation in the liver of immature chicks. *J Steroid Biochem* 24:437–442
- Cui H, Cruz-Correa M, Giardiello FM, et al (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 299:1753–1755

- Cyranowski JM, Frank E, Young E, Shear MK (2000) Adolescent onset of the gender difference in lifetime rates of major depression: a theoretical model. *Arch Gen Psychiatry* 57:21–27
- Dahl C, Guldberg P (2003) DNA methylation analysis techniques. *Biogerontology* 4:233–250
- Deb-Rinker P, Klempan TA, O'Reilly RL, et al (1999) Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. *Genomics* 61:133–144
- Deb-Rinker P, O'Reilly RL, Torrey EF, Singh SM (2002) Molecular characterization of a 2.7 kb, 12q13-specific, retroviral related sequence isolated by RDA from monozygotic twins discordant for schizophrenia. *Genome* 45:1–10
- Dobrovic A, Bianco T, Tan LW, et al (2002) Screening for and analysis of methylation differences using methylation-sensitive single-strand conformation analysis. *Methods* 27:134–138
- Dupont JM, Tost J, Jammes H, Gut IG (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 333:119–127
- Eads CA, Danenberg KD, Kawakami K, et al (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* 28:E32
- Ehrlich M, Ehrlich K (1993) Effect of DNA methylation and the binding of vertebrate and plant proteins to DNA. In: Jost J, Saluz P (eds) *DNA methylation: molecular biology and biological significance*. Birkhauser Verlag, Basel, 145–168
- Feinberg AP (1999) Imprinting of a genomic domain of 11p15 and loss of imprinting in cancer: an introduction. *Cancer Res* 59:1743s–1746s
- Fraga ME, Uriol E, Borja Diego L, et al (2002) High-performance capillary electrophoretic method for the quantification of 5-methyl 2'-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. *Electrophoresis* 23:1677–1681
- Frigola J, Ribas M, Risques RA, Peinado MA (2002) Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Res* 30:e28
- Frommer M, McDonald LE, Millar DS, et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89:1827–1831
- Fuke C, Shimabukuro M, Petronis A, 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
- Galm O, Rountree MR, Bachman KE, et al (2002) Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Res* 12:153–157
- Gitan RS, Shi H, Chen CM, et al (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res* 12:158–164
- Gonzalzo ML, Jones PA (2002) Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension (Ms-SNuPE). *Methods* 27:128–133
- Gonzalzo ML, Liang G, Spruck CH 3rd, et al (1997) Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res* 57:594–599

- Guldberg P, Worm J, Gronbaek K (2002) Profiling DNA methylation by melting analysis. *Methods* 27:121–127
- Hall JG (1990) Genomic imprinting: review and relevance to human diseases. *Am J Hum Genet* 46:857–873
- Hatada I, Kato A, Morita S, et al (2002) A microarray-based method for detecting methylated loci. *J Hum Genet* 47:448–451
- Hayashizaki Y, Hatada I, Hirotsune S, et al (1993) Restriction landmark genomic scanning (RLGS) method and its application (in Japanese). *Seikagaku* 65:109–115
- Heiman GA, Hodge SE, Wickramaratne P, Hsu H (1996) Age-at-interview bias in anticipation studies: computer simulations and an example with panic disorder. *Psychiatr Genet* 6:61–66
- Henikoff S, Matzke MA (1997) Exploring and explaining epigenetic effects. *Trends Genet* 13:293–295
- Herman JG, Graff JR, Myohanen S, et al (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 93:9821–9826
- Hodge SE, Wickramaratne P (1995) Statistical pitfalls in detecting age-of-onset anticipation: the role of correlation in studying anticipation and detecting ascertainment bias. *Psychiatr Genet* 5:43–47
- Hou P, Ji M, Ge C, et al (2003a) Detection of methylation of human p16(Ink4a) gene 5'-CpG islands by electrochemical method coupled with linker-PCR. *Nucleic Acids Res* 31:e92
- Hou P, Ji M, Liu Z, et al (2003b) A microarray to analyze methylation patterns of p16(Ink4a) gene 5'-CpG islands. *Clin Biochem* 36:197–202
- Hou P, Ji M, Li S, et al (2004) High-throughput method for detecting DNA methylation. *J Biochem Biophys Methods* 60:139–150
- Howard R, Rabins PV, Seeman MV, Jeste DV (2000) Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. The International Late-Onset Schizophrenia Group. *Am J Psychiatry* 157:172–178
- Huang TH, Laux DE, Hamlin BC, et al (1997) Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique. *Cancer Res* 57:1030–1034
- Huang TH, Perry MR, Laux DE (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* 8:459–470
- Hubrich-Kuhner K, Buhk HJ, Wagner H, et al (1989) Non-C-G recognition sequences of DNA cytosine-5-methyltransferase from rat liver. *Biochem Biophys Res Commun* 160:1175–1182
- Ingrosso D, Cimmino A, Perna AF, et al (2003) Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia. *Lancet* 361:1693–1699
- Jablonka E, Lamb M (1995) Epigenetic inheritance and evolution. Oxford University Press, New York, pp 1–360
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33 Suppl:245–254
- Jantzen K, Fritton HP, Igo-Kemenes T, et al (1987) Partial overlapping of binding sequences for steroid hormone receptors and DNaseI hypersensitive sites in the rabbit uteroglobin gene region. *Nucleic Acids Res* 15:4535–4552
- Jones PA, Laird PW (1999) Cancer epigenetics comes of age. *Nat Genet* 21:163–167

- Jones PL, Veenstra GJ, Wade PA, et al (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19:187–191
- Kaminsky ZA, Assadzadeh A, Flanagan J, Petronis A (2005) Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. *Nucleic Acids Res* 33:e95
- Kapranov P, Cawley SE, Drenkow J, et al (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916–919
- Kendler KS, Prescott CA (1999) A population-based twin study of lifetime major depression in men and women. *Arch Gen Psychiatry* 56:39–44
- Keverne EB (1997) Genomic imprinting in the brain. *Curr Opin Neurobiol* 7:463–468
- Kuo KC, McCune RA, Gehrke CW, et al (1980) Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res* 8:4763–4776
- Lavrentieva I, Broude NE, Lebedev Y, et al (1999) High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. *FEBS Lett* 443:341–347
- Leonard CM, Williams CA, Nicholls RD, et al (1993) Angelman and Prader-Willi syndrome: a magnetic resonance imaging study of differences in cerebral structure. *Am J Med Genet* 46:26–33
- Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3:662–673
- Li J, Protopopov A, Wang F, et al (2002) NotI subtraction and NotI-specific microarrays to detect copy number and methylation changes in whole genomes. *Proc Natl Acad Sci U S A* 99:10724–10729
- Liang G, Gonzalgo ML, Salem C, Jones PA (2002) Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* 27:150–155
- Lo YM, Wong IH, Zhang J, et al (1999) Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. *Cancer Res* 59:3899–3903
- Matin MM, Baumer A, Hornby DP (2002) An analytical method for the detection of methylation differences at specific chromosomal loci using primer extension and ion pair reverse phase HPLC. *Hum Mutat* 20:305–311
- McInnis MG (1996) Anticipation: an old idea in new genes. *Am J Hum Genet* 59:973–979
- McMahon FJ, Stine OC, Meyers DA, et al (1995) Patterns of maternal transmission in bipolar affective disorder. *Am J Hum Genet* 56:1277–1286
- McMahon FJ, Hopkins PJ, Xu J, et al (1997) Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. *Am J Hum Genet* 61:1397–1404
- Mueller K, Doerfler W (2000) Methylation-sensitive amplicon subtraction: a novel method to isolate differentially methylated DNA sequences in complex genomes. *Gene Funct Dis* 1:154–160
- Nan X, Ng HH, Johnson CA, et al (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393:386–389
- Nicholls RD (2000) The impact of genomic imprinting for neurobehavioral and developmental disorders. *J Clin Invest* 105:413–418

- Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet* 2:153–175
- Numachi Y, Yoshida S, Yamashita M, et al (2004) Psychostimulant alters expression of DNA methyltransferase mRNA in the rat brain. *Ann N Y Acad Sci* 1025:102–109
- Oakeley EJ, Podesta A, Jost JP (1997) Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc Natl Acad Sci U S A* 94:11721–11725
- Oakeley EJ, Schmitt F, Jost JP (1999) Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. *Biotechniques* 27:744–746, 748–750, 752
- Ohara K, Xu HD, Mori N, et al (1997) Anticipation and imprinting in schizophrenia. *Biol Psychiatry* 42:760–766
- Pasqualini JR, Mercat P, Giambiagi N (1989) Histone acetylation decreased by estradiol in the MCF-7 human mammary cancer cell line. *Breast Cancer Res Treat* 14:101–105
- Peoples R, Wood M, Van Atta R (2004) Photocrosslinking oligonucleotide hybridization assay for concurrent gene dosage and CpG methylation analysis. *Methods Mol Biol* 287:233–249
- Petronis A (1996) Genomic imprinting in unstable DNA diseases. *Bioessays* 18:587–590
- Petronis A (2004) The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol Psychiatry* 55:965–970
- Petronis A, Kennedy JL (1995) Unstable genes—unstable mind? *Am J Psychiatry* 152:164–172
- Petronis A, Popendikyte V, Kan P, Sasaki T (2002) Major psychosis and chromosome 22: genetics meets epigenetics. *CNS Spectr* 7:209–214
- Petronis A, Gottesman II, Kan P, et al (2003) Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophr Bull* 29:169–178
- Pfeifer GP, Steigerwald SD, Mueller PR, et al (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* 246:810–813
- Pfeifer K (2000) Mechanisms of genomic imprinting. *Am J Hum Genet* 67:777–787
- Piccinelli M, Wilkinson G (2000) Gender differences in depression. Critical review. *Br J Psychiatry* 177:486–492
- Polymeropoulos MH, Xiao H, Torrey EF, et al (1993) Search for a genetic event in monozygotic twins discordant for schizophrenia. *Psychiatry Res* 48:27–36
- Rakyan V, Whitelaw E (2003) Transgenerational epigenetic inheritance. *Curr Biol* 13:R6
- Rakyan VK, Preis J, Morgan HD, Whitelaw E (2001) The marks, mechanisms and memory of epigenetic states in mammals. *Biochem J* 356:1–10
- Rakyan VK, Blewitt ME, Druker R, et al (2002) Metastable epialleles in mammals. *Trends Genet* 18:348–351
- Rakyan VK, Hildmann T, Novik KL, et al (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2:e405
- Rand K, Qu W, Ho T, et al (2002) Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* 27:114–120

- Razin A, Shemer R (1999) Epigenetic control of gene expression. *Results Probl Cell Differ* 25:189–204
- Rein T, DePamphilis ML, Zorbas H (1998) Identifying 5-methylcytosine and related modifications in DNA genomes. *Nucleic Acids Res* 26:2255–2264
- Reiss D, Plomin R, Hetherington EM (1991) Genetics and psychiatry: an unheralded window on the environment. *Am J Psychiatry* 148:283–291
- Riggs A, Porter T (1996) Overview of epigenetic mechanisms. In: Russo VEA MR, Riggs AD (eds) *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 29–45
- Riggs A, Xiong Z, Wang L, JM L (1998) Methylation dynamics, epigenetic fidelity and X chromosome structure. In: Wolffe A (ed) *Epigenetics*. John Wiley and Sons, Chichester, pp 214–227
- Risch N (1990) Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol* 7:3–16; discussion 17–45
- Ross SA (2003) Diet and DNA methylation interactions in cancer prevention. *Ann N Y Acad Sci* 983:197–207
- Rother KI, Silke J, Georgiev O, et al (1995) Influence of DNA sequence and methylation status on bisulfite conversion of cytosine residues. *Anal Biochem* 231:263–265
- Saluz HP, Jiricny J, Jost JP (1986) Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis. *Proc Natl Acad Sci U S A* 83:7167–7171
- Schatz P, Dietrich D, Schuster M (2004) Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. *Nucleic Acids Res* 32:e167
- Schmitt F, Oakeley EJ, Jost JP (1997) Antibiotics induce genome-wide hypermethylation in cultured *Nicotiana tabacum* plants. *J Biol Chem* 272:1534–1540
- Schulze TG, Chen YS, Badner JA, et al (2003) Additional, physically ordered markers increase linkage signal for bipolar disorder on chromosome 18q22. *Biol Psychiatry* 53:239–243
- Schumacher A (2001) Mechanisms and brain specific consequences of genomic imprinting in Prader-Willi and Angelman syndromes. *Gene Funct Dis* 1:7–25
- Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, Virtanen C, Winegarden N, Cheng J, Gingeras T, Petronis A (2006) Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res* (in press)
- Shi H, Maier S, Nimmrich I, et al (2003a) Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J Cell Biochem* 88:138–143
- Shi H, Wei SH, Leu YW, et al (2003b) Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. *Cancer Res* 63:2164–2171
- Siegfried Z, Eden S, Mendelsohn M, et al (1999) DNA methylation represses transcription in vivo. *Nat Genet* 22:203–206
- Stach D, Schmitz OJ, Stilgenbauer S, et al (2003) Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res* 31:E2
- Sutherland E, Coe L, Raleigh EA (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J Mol Biol* 225:327–348
- Sutherland JE, Costa M (2003) Epigenetics and the environment. *Ann N Y Acad Sci* 983:151–160

- Taubes G (1995) Epidemiology faces its limits. *Science* 269:164–169
- Thomassin H, Oakeley EJ, Grange T (1999) Identification of 5-methylcytosine in complex genomes. *Methods* 19:465–475
- Thomassin H, Kress C, Grange T (2004) MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. *Nucleic Acids Res* 32:e168
- Tompa R, McCallum CM, Delrow J, et al (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr Biol* 12:65–68
- Tost J, Schatz P, Schuster M, et al (2003) Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res* 31:e50
- Toyota M, Ho C, Ahuja N, et al (1999) Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 59:2307–2312
- Truss M, Chalepakis G, Pina B, et al (1992) Transcriptional control by steroid hormones. *J Steroid Biochem Mol Biol* 41:241–248
- Tsujita T, Niikawa N, Yamashita H, et al (1998) Genomic discordance between monozygotic twins discordant for schizophrenia. *Am J Psychiatry* 155:422–424
- Uhlmann K, Brinckmann A, Toliat MR, et al (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23:4072–4079
- Ushijima T, Morimura K, Hosoya Y, et al (1997) Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proc Natl Acad Sci U S A* 94:2284–2289
- Veldic M, Caruncho HJ, Liu WS, et al (2004) DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. *Proc Natl Acad Sci U S A* 101:348–353
- Vincent JB, Kalsi G, Klempan T, et al (1998) No evidence of expansion of CAG or GAA repeats in schizophrenia families and monozygotic twins. *Hum Genet* 103:41–47
- Walter J, Paulsen M (2003) Imprinting and disease. *Semin Cell Dev Biol* 14:101–110
- Warren MP, Brooks-Gunn J (1989) Mood and behavior at adolescence: evidence for hormonal factors. *J Clin Endocrinol Metab* 69:77–83
- Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23:5293–5300
- Weaver IC, Cervoni N, Champagne FA, et al (2004) Epigenetic programming by maternal behavior. *Nat Neurosci* 7:847–854
- Weissman MM, Olfson M (1995) Depression in women: implications for health care research. *Science* 269:799–801
- Weksberg R, Shuman C, Caluseriu O, et al (2002) Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. *Hum Mol Genet* 11:1317–1325
- Weksberg R, Smith AC, Squire J, Sadowski P (2003) Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development. *Hum Mol Genet* 12 Spec No 1:R61–68
- Williams CA, Hendrickson JE, Cantu ES, Donlon TA (1989) Angelman syndrome in a daughter with del(15)(q11q13) associated with brachycephaly, hearing loss, enlarged foramen magnum, and ataxia in the mother. *Am J Med Genet* 32:333–338

- Wolff GL, Kodell RL, Moore SR, Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *Faseb J* 12:949–957
- Worm J, Aggerholm A, Guldberg P (2001) In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin Chem* 47:1183–1189
- Wu J, Issa JP, Herman J, et al (1993) Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc Natl Acad Sci U S A* 90:8891–8895
- Xiong LZ, Xu CG, Saghai Maroof MA, Zhang Q (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Mol Gen Genet* 261:439–446
- Xiong Z, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* 25:2532–2534
- Yamamoto F, Yamamoto M (2004) A DNA microarray-based methylation-sensitive (MS)-AFLP hybridization method for genetic and epigenetic analyses. *Mol Genet Genomics* 271:678–686
- Yamamoto T, Nagasaka T, Notohara K, et al (2004) Methylation assay by nucleotide incorporation: a quantitative assay for regional CpG methylation density. *Biotechniques* 36:846–850, 852, 854
- Yan PS, Efferth T, Chen HL, et al (2002) Use of CpG island microarrays to identify colorectal tumors with a high degree of concurrent methylation. *Methods* 27:162–169
- Yang AS JP, Shibata A (1996) The mutational burden of 5-methylcytosine. In: Russo V, Riggs A (eds) *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 77–94
- Yokomori N, Moore R, Negishi M (1995) Sexually dimorphic DNA demethylation in the promoter of the Slp (sex-limited protein) gene in mouse liver. *Proc Natl Acad Sci U S A* 92:1302–1306
- Zeschnigk M, Bohringer S, Price EA, et al (2004) A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. *Nucleic Acids Res* 32:e125
- Zhang Z, Chen CQ, Manev H (2004) Enzymatic regional methylation assay for determination of CpG methylation density. *Anal Chem* 76:6829–6832
- Zubenko GS, Maher B, Hughes HB 3rd, et al (2003) Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *Am J Med Genet* 123B:1–18

