Epigenetics as a unifying principle in the aetiology of complex traits and diseases

Arturas Petronis[1]

Epigenetic modifications of DNA and histones might be crucial for understanding the molecular basis of complex phenotypes. One reason for this is that epigenetic factors are sometimes malleable and plastic enough to react to cues from the external and internal environments. Such induced epigenetic changes can be solidified and propagated during cell division, resulting in permanent maintenance of the acquired phenotype. In addition, the finding that there is partial epigenetic stability in somatic and germline cells allows insight into the molecular mechanisms of heritability. Epigenetics can provide a new framework for the search of aetiological factors in complex traits and diseases.

The nature-versus-nurture debate was one of the most important themes of biomedical science in the twentieth century. Researchers resolved it by conceding that both factors have a crucial role and that phenotypes result from the actions and interactions of both, which often change over time. Most 'normal' phenotypes and disease phenotypes show some degree of heritability, a finding that formed the basis for a series of molecular studies of genes and their DNA sequences. In parallel to such genetic strategies, thousands of epidemiological studies have been carried out to identify environmental factors that contribute to phenotypes. In this article, I consider complex, non-Mendelian, traits and diseases, and review the complexities of investigating their aetiology by using traditional — epidemiological and genetic — approaches. I then offer an epigenetic interpretation that cuts through several of the Gordian knots that are impeding progress in these aetiological studies.

Environment

Considerable effort has been dedicated to uncovering measurable environmental factors that contribute to the observed variation in normal traits or that alter the risk of acquiring a disease. The epidemiological search for risk factors is based on circumstantial evidence that what humans eat, drink and breathe, as well as the rest of their biological and psychological environment, contributes to the development of many severe illnesses. This is formally supported by studies of adopted individuals, families and monozygotic twins, which are carried out with the implicit assumption that any variation that is not attributable to genetic factors must stem from the environment[1]. However, although numerous epidemiological studies have been made, there are only a few well-proven examples of specific environmental factors that substantially affect illness: for example smoking and lung disease, and sunlight and skin cancer. There are several reasons that epidemiological studies have not yielded clearer findings.

The main methodological problem lies in the nature of observational epidemiological studies. It is difficult to estimate objectively the duration, intensity and frequency of a large variety of multidirectional environmental influences[2]. Even strong associations between an environmental factor and a disease do not necessarily prove that the environmental factor has caused the disease[2]. For example, the measured association between cannabis use and schizophrenia might result from individuals medicating themselves in an attempt to dull their pre-psychotic symptoms rather than from the cannabis itself triggering the schizophrenia[3]. One way around this would be to carry out a prospective, randomized controlled trial, but this is clearly unethical because it would involve deliberately exposing people to putative disease-causing agents. Controlling human environments in a way that eliminates the biases that confound epidemiological studies is also not possible[2]. Such designs might be possible in animal studies, but adequate animal models are available for only a small proportion of human conditions. All of these difficulties mean that data from epidemiological studies cannot be interpreted definitively and must be supplemented by other experiments or data.

Another difficulty comes from the observation that environmental risk factors can often have a 'heritable' component embedded in what seems to be purely environmental[4]. For example, smoking is a major environmental risk factor for lung cancer[5]; however, the tendency to smoke regularly has a heritable component (accounting for about 60% of the variance in liability to regular tobacco use in a study of twins born in Sweden)[6]. Another case of 'contamination' of environment with heritability involves stressful life events: these have traditionally been thought to increase the risk of depression, but genetics also affects an individual's tendency to select high-risk environments[7]. After the heritable component has been subtracted, can it be assumed that the remaining stressors are purely environmental? Perhaps there are multiple layers of heritability, like a Russian *matryoshka* doll, in which successively smaller versions are concealed within the previous figurine. When researchers think that they are investigating the effects of environment, they might in fact be dealing with hidden heritability.

A further problem inherent in epidemiological investigations of environmental risk factors for disease pertains to the idea of the 'non-shared environment', which refers to the variation that cannot be attributed to heritable factors. In this process, environmental influences that result in similarities in family members are not considered to be important, whereas aspects of the environment that make these individuals different are proposed to account for all of the non-genetic variation[8]. It is now generally accepted that environmental factors often operate through mechanisms that make offspring in a family different rather than similar[9]. Shared factors, such as socio-economic status, parental education, child-rearing practices and marital quality, are assumed to affect siblings similarly and therefore to have little causal effect for the trait or disease in question. Meta-analysis of multiple sources of data revealed that the

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

proportion of the variation in personality traits that is attributable to non-shared environment among individuals is 45–60%, whereas that attributable to shared environment is nearly zero[10].

The non-shared environment also makes an important contribution to the risk of developing complex diseases. For certain cancers, it has been estimated to have an effect that is twice as large as that of heritability. In a Scandinavian twin study, the heritability of breast cancer was estimated at 27% (95% confidence interval (CI) 4–41%; that is, 27% of the risk of developing breast cancer is heritable); and ovarian cancer, 22% (CI 0–41%)[11]. By contrast, the non-shared environment accounted for 67% (CI 59–76%) and 78% (CI 59-99%), respectively, of the variation in whether individuals in the population developed these diseases[11]. Beyond academic circles, science writers and the general public have also discussed how the theory of non-shared environment applies to everyday lives. For example, in one popular science book, it was concluded that parents play a minor part in the mental and emotional development of their offspring and that peer groups, instead, are the strongest environmental influence on personality development[12].

The concept of non-shared environment, however, can be criticized in several ways. First, it is not possible to rationalize how siblings (that is, their tissues or cells) react differently to a specific environmental event. Second, the identification of specific non-shared environmental events is extremely difficult. A review of 43 studies concluded that, although more than 50% of the phenotypic variance in behavioural outcomes could be attributed to non-shared environment, the objectively defined non-shared events could cumulatively explain less than 2% of phenotypic variance at best[13]. Third, non-shared environment is limited to humans, yet inbred animals and cloned animals show considerable phenotypic differences, despite the absence of detectable environmental variation[14–17].

In conclusion, the effort to uncover the role of environmental factors in complex traits and diseases has diverged into several disparate, narrow questions, each of which is mostly investigated in isolation. It is not possible to design an epidemiological study that would account for all of the complexities of environmental contribution as they are currently conceived. In addition, some newer concepts are incompatible with traditional epidemiological approaches, for example the idea that there are heritable influences on non-shared environment. A breakthrough in this field cannot be expected until researchers have a clear understanding of the factors for which they are searching.

Heritability

Heritability is a central concept in biology and was thought for many years to hold the key to unravelling the molecular aetiology of human disease. Among twin, family and adoption studies of heritability, twin studies are thought to provide the most elegant approach and have been used widely to estimate the relative roles of genetic and environmental factors in phenotypic variation[18,19]. In these studies, heritability is estimated by comparing the concordance rates or intraclass correlations of monozygotic twins for particular traits with those of dizygotic twins for the same traits, given that monozygotic twins have identical nuclear DNA and that dizygotic twins have in common (on average) 50% of segregating DNA sequence variation (but not 50% of their genes or DNA in common, as is often mistakenly stated, even in respected scientific publications). Over the past 50 years, thousands of heritability studies have shown that nearly all human traits — whether normal or abnormal — are heritable to a certain extent, including even surprising behavioural traits such as divorce[20], sports participation and religious beliefs[19]. Interpreting heritability is, however, not always straightforward. Heritability can fluctuate across different stages of life and can differ markedly in different environments[21,22]. For example, in depression, heritability was found to be lower in married women than in unmarried women in ~2,000 twin pairs from the Australian Twin Registry[23]. The geneticist Irving Gottesman encapsulated these issues using the words of physicist Paul Hansma[24], writing that heritabilities are "like snapshots of a ballerina. They won't tell you about the ballet"[25].

All of these observations make it difficult to visualize how highly stable DNA sequences can account for heritability, which is malleable and context-dependent. Furthermore, it is becoming evident that the heuristic value of quantitative estimates of heritability is limited, as these estimates do not provide information on the number of genes involved, the presence or absence of major genes or the effect sizes of the genes (the magnitude of the contribution of each gene to variation in the phenotype)[22]. Genome-wide association studies have uncovered dozens of DNA polymorphisms and haplotypes that are associated with particular diseases. But, as a rule, both the individual and the cumulative effects of these genetic variants are small and do not approach the size needed to explain the initial estimates of heritability. An informative example is height: height is 80–90% heritable, but the 40 loci that have been detected in genome-wide association studies together account for less than 5% of heritability for height[26]. Although most studies have searched only for common genetic variants, it is doubtful that rare DNA sequence variants will close the gap of the 'missing heritability'.

There are also controversial issues surrounding heritability in the context of DNA sequence variation. Coefficients of genetic similarity among relatives, which decrease by one-half with each degree of genetic relatedness (for example one-half and one-quarter for first-degree and second-degree relatives, respectively), were established in the early twentieth century. Only recently was it determined that the difference in DNA sequence among unrelated individuals is very low (<1%)[27,28]. Therefore, two unrelated individuals and two first-degree relatives have on average >99% and >99.5% DNA sequence similarity, respectively. The coefficients of genetic relatedness have not been revised accordingly, and this creates the problem illustrated in the following thought experiment. Assume that there is a population in which two randomly selected individuals have, on average, 98% DNA sequence similarity. The first-degree relatives in this hypothetical population will have, on average, 99% of their DNA in common, which is the same as unrelated individuals in the human population. Would children in this population show phenotypic similarities because they were born to the same parents and have 99% of their DNA in common? Alternatively, would such sisters and brothers show no phenotypic similarities, because they are as different genetically (1%) as unrelated individuals in the human population? Both of these predictions seem to be formally correct, but they are incompatible. The only possible compromise is if unrealistically high degrees of epistasis (when several genes interacting in a non-additive manner contribute to a phenotype) are involved, which would explain the differences between sharing 99% of DNA by descent and sharing the same amount by state. .

According to the current paradigm, DNA sequence variation is the sole substrate and carrier of heritability (Box 1). In their widely used textbook on human genetics, Friedrich Vogel and Arno Motulsky stated, "it is our goal to trace down genetic differences to the DNA level"[29]. This DNA-centric model has allowed scientists to uncover the molecular genetic origins of Mendelian traits and diseases successfully. But many traits and diseases are non-Mendelian, and these complex diseases differ considerably in their epidemiological, clinical and molecular parameters from single-gene, Mendelian, diseases. I argue that taking an epigenetic perspective allows a different interpretation of the irregularities, complexities and controversies of traditional environmental and genetic studies.

Epigenetic solutions

There is already fragmentary experimental evidence that epigenetics can account for some of the variation that had previously been attributed to environmental and heritable effects[30,31]. However, there are few such studies, and the observations seem exotic. The overall perception is that such findings are exceptions to the rule rather than hallmarks of a new and fundamentally different model of non-Mendelian genetics and biology. In this section, an alternative to the traditional model of phenotypic variation is presented, in which the importance of epigenetic factors is central. The main idea is that epigenetic stability or instability — that is, both rigid and plastic epigenetic regulation of genomes — can largely replace the genetic and environmental components in traditional models, and that inherited or acquired epigenetic

regulation or misregulation can be a core unifying molecular mechanism of complex, non-Mendelian, traits and diseases.

Epigenetics and the environment

Epigenetics is relevant to phenomena that have traditionally been attributed to the environment in two ways. First, certain environmental events, including maternal behaviour[32] and physical exercise[33], can induce plastic epigenetic changes[31]. Epigenetic factors are at the interface between environmental stimuli and long-lasting molecular, cellular and behavioural phenotypes (see page 728) that are acquired during periods of developmental plasticity[34]. The advantage of taking an epigenetic perspective is that, especially in humans, it is much easier to identify epigenetic differences than to carry out traditional epidemiological studies. Second, epigenetic factors can 'mutate' in the absence of a detectable environmental influence. For example, during mitosis, DNA methylation patterns are transmitted from maternal chromatids to daughter chromatids, and the degree of fidelity in this transmission is about three orders of magnitude lower than that of DNA sequence (an error rate of 1 in 10^6 and 1 in 10^3 for DNA sequences and DNA modification, respectively)[35]. This stochastic epigenetic instability can result in significant epigenetic differences accumulating over time across cells, despite the DNA sequence identity of these cells[36].

What is the ratio of environment-induced changes to stochastic epigenetic changes? Most probably, stochastic epigenetic changes are more common than environment-induced changes. It is these stochastic epigenetic changes in somatic cells — rather than the non-shared environment — that might account for the observed discordance between monozygotic twins, the degree of which is independent of whether the twins were reared together or apart[37]. In addition, stochastic epigenetic variation — and not the non-shared environment — can also explain why phenotypic variation in populations of inbred animals is as large as that in outbred animals, despite both being raised in tightly controlled environments[38]. Therefore, epigenetic stochasticity seems to be a major mechanism that leads to phenotypic differences among genetically identical organisms. This concept is also consistent with the observation that the strict standardization of laboratory environments does not have a major effect on the inter-individual variability of inbred animals compared with the tremendous environmental variability in a natural setting[39].

In addition to somatic epigenetic instability, another source of 'non-shared environment' is epigenetic variation in the germ line. Numerous epigenetically different zygotes can be produced by the same parents. More specifically, DNA methylation profiles differ significantly across sperm and oocytes derived from the same individuals[40,41]. Fine mapping of the methylated cytosine bases in the CpG islands of six disease-associated genes (*BRCA1*, *BRCA2*, *PSEN1*, *PSEN2*, *DM1* and *HD*) showed that each sperm cell has a unique DNA methylation profile and that the variation in epigenetic marks greatly exceeds that in DNA sequence[41]. Although the functional impact of each modified and unmodified cytosine position is unknown, such epigenetically variable germ cells provide different epigenetic starting points for offspring of the same parents. After fertilization, the zygote is epigenetically reprogrammed, raising the question of whether germ-cell-specific epigenetic differences are retained[42,43]. Several observations suggest that this retention is possible. First, there are two notable examples of parental epigenetic marks 'surviving': in the mouse alleles agouti viable yellow (A^{vy}) and axin fused ($Axin1^{Fu}$). After fertilization, despite a temporary loss of the DNA methylation profiles of the parental germ line[40], these epigenetic marks persist during development and result in predictable phenotypic outcomes. Second, cloned animals have considerable epigenetic (and phenotypic) differences from their single parent, despite the DNA sequence identity of offspring and parent. These epigenetic differences are vestiges of the different epigenetic signatures in two zygotes: one highly artificial, and the other natural[44]. This would not be expected if the parental epigenetic profile had been completely erased and a new epigenetic profile established. Third, DNA methylation profiles are more similar between monozygotic twins than between dizygotic twins, also supporting the idea that zygotic epigenetic signals survive, at least in part[45]. This observation is also relevant to epigenetic heritability and is discussed further in the next subsection.

Box 1 | Origin of the current paradigm in the biology of human disease

The roots of the current paradigm of the molecular basis of human disease can be traced back to more than 60 years ago, when Linus Pauling and colleagues found that haemoglobin protein that had been extracted from individuals with sickle-cell anaemia had an abnormal electrophoretic mobility[72]. Their study showed that sickle-cell anaemia is a 'molecular disease', and they postulated that the disease was caused by a defect in the globin protein. It was soon determined that a single amino-acid substitution is the specific chemical change that distinguishes haemoglobin in individuals with sickle-cell anaemia from 'normal' haemoglobin. A series of fundamental discoveries in molecular genetics — including the elucidation of the structure of DNA, the deciphering of the genetic code and the advent of recombinant DNA technology — paved the road for the detection of the first molecular genetic defects in the variant of the haemoglobin gene that causes sickle-cell anaemia and thalassaemia. Eventually, DNA-sequence-based strategies led to genetic mutations being uncovered in people with other genetic diseases, such as phenylketonuria, Duchenne muscular dystrophy and cystic fibrosis. In most affected individuals, a genetic defect was located in the coding sequence of a gene, and the detected mutations were found to change the structure and function of the encoded protein. The mechanism by which such diseases arise is straightforward: if there are mutations in both alleles (for autosomal recessive diseases or in a single allele on the X chromosome in males), there is loss of function of a specific protein. If there are no proteins that can substitute for the non-functioning protein in the cell, then disease occurs.

This successful identification of genetic defects in Mendelian diseases provided the basis for the current paradigm of human morbid genetics. The idea of the importance of DNA mutations has been generalized and extrapolated to a "fundamentally different group of diseases"[73], namely complex, non-Mendelian, diseases. The model was revised to fit complex diseases into the schemes of analyses that had already been developed. But the modifications essentially consisted of treating the genes as predisposing factors instead of causative factors and putting a stronger emphasis on environmental effects. At present, the evidence that diseases are heritable implies the presence of DNA mutations or polymorphisms that predispose an individual to be affected by a particular disease.

The realization that defining the DNA variants associated with diseases could lead to a breakthrough in disease diagnostics and personalized treatment stimulated the rapid development of powerful tools for analysing DNA (for example DNA microarrays and next-generation sequencing technologies) and comprehensive DNA databases (for example haplotype maps, and annotations of single-nucleotide polymorphisms and copy-number variants). The overall value and limitations of the DNA-sequence-based paradigm for understanding the mechanism of complex diseases and traits in humans will be clear when DNA-analysis technologies become inexpensive enough to make it feasible to sequence and compare the whole genomes of thousands of individuals.

Epigenetics and heritability

This epigenetic variation in the germ line and partial epigenetic stability in somatic cells could shed light on the epigenetic mechanisms of heritability. The term heritability is used here because it mirrors twin-based studies of 'phenotypic heritability', but it does not imply transmission between generations. My research group recently assessed the DNA methylation profiles of buccal cells from 20 sets of monozygotic twins and 20 sets of dizygotic twins. Matched monozygotic co-twins had significantly higher intraclass correlations than dizygotic co-twins[45]. This greater epigenetic discordance in dizygotic twins can, to some extent, be secondary to DNA sequence differences[46]. But our computational analyses of single-nucleotide polymorphisms and comparison

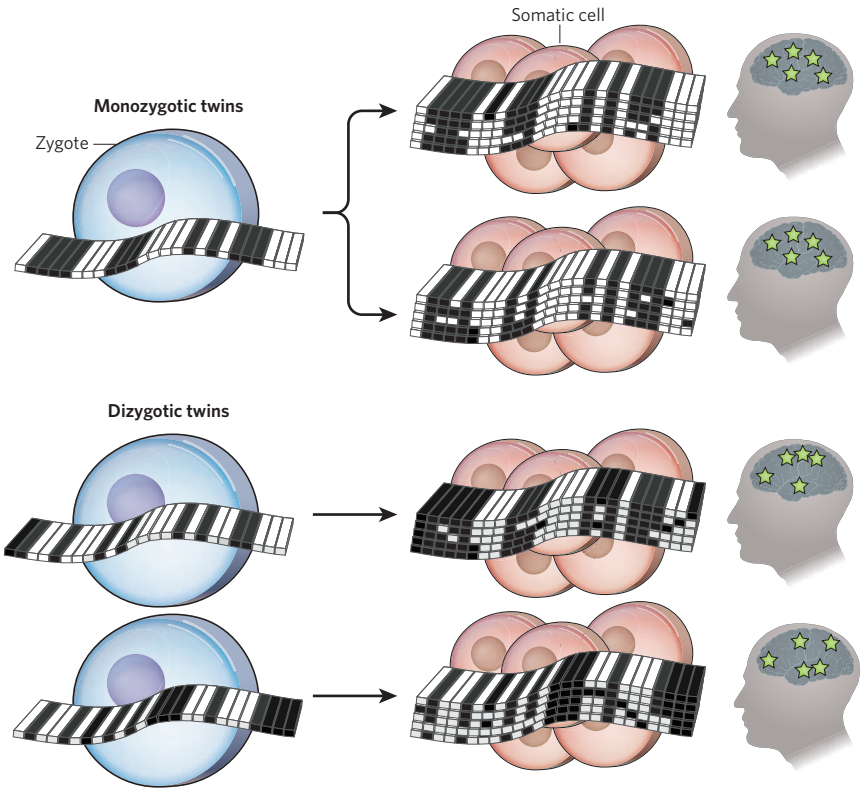

Figure 1 | Twin-based epigenetic heritability. DNA methylation profiles are presented as black and white keys in the germ line or zygote (one layer) and somatic cells (multiple layers). Black denotes, for example, methylated cytosine, and white denotes, for example, unmethylated cytosine. Monozygotic twins originate from a single zygote, and their initial epigenetic status is more similar than that of dizygotic twins, who develop from two separate zygotes with different epigenetic profiles. The epigenetic modifications in both monozygotic twins and dizygotic twins are subject to stochastic and, to a lesser extent, environmental factors, which induce similar amounts of somatic epigenetic variation in tissues. Owing to epigenetic differences in the original zygotes, however, dizygotic twins have more epigenetic variation in their somatic cells than do monozygotic twins. This could account for the large phenotypic differences (green stars) observed between dizygotic twins compared with monozygotic twins.

of DNA methylation profiles in inbred and outbred animals[45] indicate that the hypothesis that some somatic epigenetic differences detected in dizygotic twins originate from epigenetic differences in the zygotes cannot be rejected. Furthermore, we propose that monozygotic co-twins are more phenotypically similar than dizygotic co-twins, not only because the monozygotic twins have identical genomes but also because they have an epigenetically similar starting point at the zygote stage (Fig. 1). Our results[45] are consistent with those of targeted epigenetic studies, in which the epigenetic modifications of individual genes were assessed[47]. Partial somatic epigenetic stability, together with germline (zygotic) epigenetic differences, can also explain the paradox reported by Klaus Gärtner: that, in mice, monozygotic co-twins show a greater degree of phenotypic similarity than dizygotic co-twins, despite both groups being isogenic and housed in controlled environments[14].

Another facet of epigenetic heritability is transgenerational heritability. Transgenerational epigenetic heritability is the concept that epigenetic factors, to some degree, can survive not only epigenetic changes after fertilization[42,43] but also the process of epigenetic reprogramming during gametogenesis and are therefore transmitted to the next generation[48,49]. This type of epigenetic heritability is best documented in plants, including the recent findings that the stable segregation of parental epialleles (alleles that differ in their epigenetic modification) is involved in the variation of flowering time and height of *Arabidopsis thaliana*[50]. In addition, DNA methylation levels are not reset in early development in zebrafish (*Danio rerio*)[51]. And, in mammals, methylated DNA marks are not always completely erased during gametogenesis[52,53]. The list of genetic loci that are resistant to major epigenetic reprogramming events in mice includes some transgenes and retrotransposons[54]. More subtle effects in transgenerational epigenetic dynamics may also be possible. In the study of CpG islands in six genes in human sperm (discussed above), there was much less intra-individual epigenetic variation than inter-individual variation[41]. Several interpretations of this finding are possible, including that the demethylation of DNA in primordial cells (that is, in precursor germ cells) might not be complete or that other mechanisms can partially restore parental epigenetic patterns.

In the DNA-sequence-based paradigm, twin-based heritability and transgenerational heritability, in principle, mean the same thing, although the measured values of heritabilities can differ somewhat for several reasons. For example, epistasis might result in overestimates of heritability in twin studies, whereas shared environment (although it is assumed to be unimportant) might inflate heritability estimates in family studies. Twin-based heritability and transgenerational heritability are not the same, however, from the perspective of epigenetic heritability, creating a new, already controversial, semantic issue[55–57]. Twin-based epigenetic heritability is limited to a single generation and originates from partial epigenetic stability in somatic cells, whereas for transgenerational heritability there also needs to be at least some degree of epigenetic stability during gametogenesis. Differences in heritability values between twin studies and family studies[58,59] could therefore reflect epigenetic events during gametogenesis.

The incomplete erasure of epigenetic marks in the germ line provides a new explanation for sporadic (that is, when there is no family history) and familial cases of disease. Complete or substantial erasure of epimutations (or pathological epigenetic marks) in the germ line of an affected individual will halt the propagation of a disease (Fig. 2a). If a germline epimutation(s) is resistant to erasure, however, it will be transmitted to the next generation, and the disease will appear in a familial manner (Fig. 2b). Such intergenerational transmission of DNA methylation patterns has been observed for the transgene *TKZ751* in mice[60]. Whether these patterns were erased in the germ line was determined by the genetic background of the non-transgenic parent. When the epigenetic signal was not erased, the wave of methylation spread by 6–10 kilobases with each subsequent generation, correlating with decreasing steady-state levels of transgene messenger RNA levels in each generation. This pattern is consistent with genetic anticipation: that is, with increasing disease severity and/or decreasing age of onset in younger generations (Fig. 2b). Traditionally, genetic anticipation is associated with the expansion of DNA repeats[61]. Clinical patterns that are consistent with genetic anticipation have been detected for numerous diseases[62,63]. The search for unstable DNA in individuals affected with complex diseases has not been successful, however, so interest in this approach has been declining. The example of the *TKZ751* transgene suggests that anticipation can instead have an epigenetic origin[64]. Another possible example of epigenetic anticipation has been found

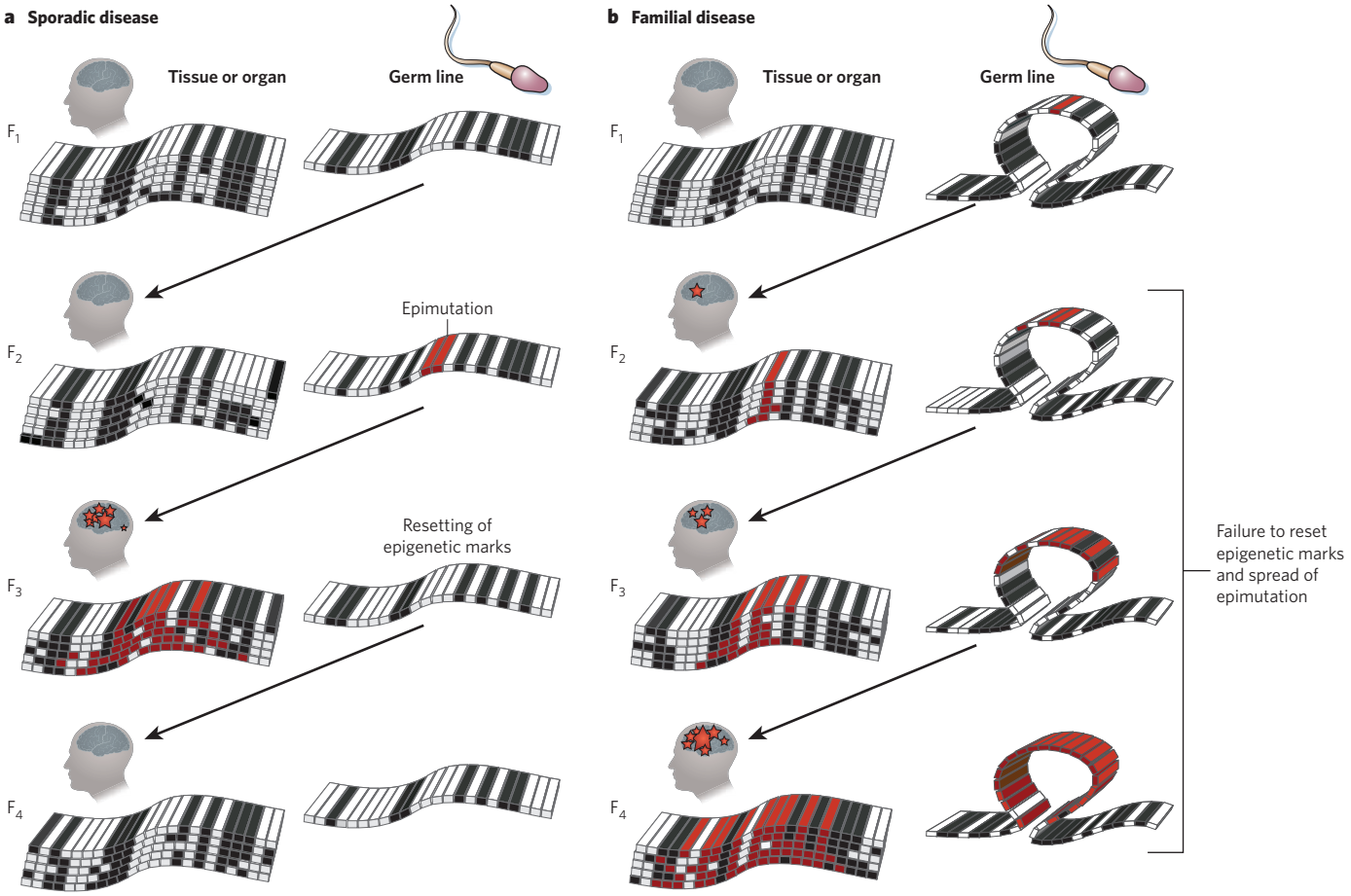

Figure 2 | Epigenetic interpretation of cases of sporadic disease and familial disease. DNA methylation profiles are presented as black and white keys in the germ line or zygote (one layer) and somatic cells (multiple layers). Black denotes, for example, methylated cytosine, and white denotes, for example, unmethylated cytosine. Red denotes pathological epigenetic marks (or epimutations). **a**, Sporadic disease. An epimutation occurs in the germ line of the second generation (F_2). It is transmitted to F_3, spreads in the somatic tissues and induces disease (red stars). The epimutation is, however, corrected in the germ line of F_3 and is not transmitted to F_4. **b**, Familial disease. A germline epimutation occurs in F_1, is transmitted to F_2 and induces disease. The epimutation fails to be corrected in the germ line of F_2, and it is transmitted to F_3. It is not known why correction fails, but the failure might be caused by an aberrant configuration of local chromatin (shown as a DNA loop). Another correction attempt fails, and epimutations are transmitted from F_3 to F_4. In each cycle of gametogenesis, the germline epimutation becomes more severe, resulting in epigenetic anticipation: that is, disease is more severe and occurs earlier in younger generations (depicted as an increasing number of red stars). In this way, sporadic and familial patterns of disease may have a similar molecular epigenetic origin but differ because of the differential efficacy of epigenetic reprogramming during gametogenesis and/or after fertilization.

in the fruitfly *Drosophila melanogaster*, in which the penetrance of ectopic outgrowth induced by a single exposure to the heat-shock-protein inhibitor geldanamycin increased in subsequent generations, even without further exposure to the drug[65].

In the domain of epigenetics, the line between 'inherited' and 'acquired' is fuzzy. Stable epigenetic 'nature' merges fluidly with plastic epigenetic 'nurture'. The ratio between inherited and acquired epigenetic influences can vary considerably depending on species, tissue, age, sex, environmental exposure and stochastic epigenetic events, all of which are consistent with empirical observations that heritability is dynamic and not static. Another close link between heritable factors and environmental factors in epigenetic regulation is the observation that exposure to certain environments has effects that, in some cases, are transmitted epigenetically for several generations[31].

All of the ideas that I have discussed here are highly relevant to the understanding of the fundamental principles of evolution. 'Soft', epigenetic, inheritance can have a key role in adaptation to environmental changes and can endure for more than a generation[66]. Phenotypic plasticity might stem mainly from the ability of epigenetic genotype (or epigenotype) — rather than genotype[67] — to produce different pheno-types in different environments. Heritable epigenetic variation could explain the faster-than-expected adaptation to environmental change that is often observed in natural populations[68]. In addition, the large intra-individual epigenetic variation in the germ line may shed new light on the problem presented by one of the first geneticists, Hugo De Vries, more than a century ago, in his book *Species and Varieties: Their Origin by Mutation*[69], when he wrote "Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest."

Outlook

In aetiological studies of complex diseases and traits, taking an epigenetic perspective allows "handling the same bundle of data as before, but placing them in a new system of relations with one another by giving them a different framework" — the process involved in a paradigm shift, as described by Thomas Kuhn[70]. Together, stable and plastic epigenetic regulation might help researchers to understand the molecular basis of heritable and non-heritable factors. In addition to the inherited and acquired dimensions of phenotype, epigenetic misregulation is also consistent with various epidemiological, clinical and molecular features of complex diseases. These features include sexual dimorphism (for example in autism, systemic lupus erythematosus and mood disorders), parental origin effects (as in psoriasis and asthma), remissions

Box 2 | Technological and methodological issues in epigenomic studies

There are myriad techniques for identifying epigenetic modifications of DNA[74]. The first approach to mapping locus-specific differences in DNA modification involved the use of a pair of restriction enzymes known as isoschizomers, which recognized the same target sequence in the DNA but were either sensitive or insensitive to its modification. Over the past decade, the gold standard for the fine mapping of modified cytosine bases has become bisulphite conversion coupled with DNA sequencing. When DNA is exposed to bisulphite, cytosines are deaminated to uracils (and subsequently, during amplification by PCR, become thymidines), but methylated cytosines remain unchanged, thereby allowing methylated sites to be identified. However, neither the restriction-enzyme-based method nor the bisulphite conversion method can distinguish between methylated cytosines and hydroxymethylated cytosines. To differentiate between these, certain chemical approaches or antibodies specific for each subtype of cytosine modification can be used. Antibody-based approaches are also widely applied to identify the various types of histone protein modification in chromatin. All of these techniques can be scaled up by using DNA microarrays or next-generation (high-throughput) sequencing techniques so that the entire epigenome can be analysed[75].

The laboratory techniques are, however, straightforward compared with the principles of experimental design for such studies, which still need development. There are several issues that need to be considered, three of which are discussed below.

First, most genetic studies can be carried out on DNA samples from any tissue, but epigenetic profiles can differ significantly across cell types. Therefore, for an epigenetic study, researchers must assess tissues and organs that contribute to the phenotype of interest, for example the brain when studying psychiatric diseases, the gut for inflammatory bowel disease and the skin for psoriasis. An additional level of complexity is that the organs of interest often consist of a variety of cell populations; for example, brain tissue contains neurons and glia. So, when a complete tissue is analysed, there is the risk that important epigenetic signals will be missed as a result of cellular heterogeneity. Alternatively, isolating the numerous cell populations that make up a tissue and carrying out epigenomic studies on these could be technically demanding and highly complex.

Second, the cause-and-effect relationship between epigenetics and phenotype is not simple. If differences in epigenetic modifications are detected between individuals who have a disease and those who do not, these differences do not necessarily indicate that epigenetic factors cause the disease or predispose an individual to developing the disease. The disease itself, the treatment regimen or other disease-associated factors could also induce epigenetic changes. Carrying out a parallel study of tissues that are not altered by the disease (for example peripheral blood cells or cells from the buccal mucosa in individuals with schizophrenia or colon cancer) in both affected and unaffected individuals might help to distinguish causal associations from non-causal associations. It should be noted that unaffected tissues are expected to contain vestiges of primary epigenetic defects that were inherited or that occurred before tissue differentiation during embryogenesis. Carrying out prospective studies of high-risk populations and designing animal models for epigenetic studies may also help to address the cause-and-effect relationship further.

Third, the optimal sample size in epigenomic studies is not known because there are no precedents for epigenome-wide analyses of complex diseases. Moreover, power analysis, which can be used to compute the minimum sample size for a study to return statistically significant results, is almost meaningless without realistic estimates of the degree of possible epigenetic difference between cases and controls.

and relapses (as in multiple sclerosis and inflammatory bowel disease), decline of clinical symptoms with age (as in major psychiatric diseases), non-decreasing incidence of disease despite the significantly reduced reproductive fitness of the affected individuals (as in autism, schizophrenia and type 1 diabetes before the discovery of insulin) and other non-Mendelian features[71]. This versatile epigenetic theory warrants a comprehensive molecular research programme dedicated to gaining a complete understanding of epigenetic stability and plasticity in normal biology and disease, as well as of the role of epigenetic control over the genome. Experimental techniques for studying DNA modifications and histone modifications can be adapted from those originally developed for DNA sequence analysis (for example microarrays and next-generation sequencing); however, some of the methodological principles in epigenomic studies will differ from those of DNA-sequence-based studies (Box 2).

The considerable theoretical and experimental potential of an epigenetic perspective makes it a strong alternative to the existing research into complex, non-Mendelian, genetics and biology. Although the existence of competing theories may create some discomfort, it can also catalyse discoveries and is indicative of a mature scientific field. In their classic textbook, Vogel and Motulsky admitted that "human genetics, as all other branches of science, is by no way a completed and closed complex of theory and results that only need to be supplemented in a straightforward way and without major changes in conceptualization"[29]. ■

1. Hemminki, K., Lorenzo Bermejo, J. & Forsti, A. The balance between heritable and environmental aetiology of human disease. *Nature Rev. Genet.* **7,** 958–965 (2006).
2. Taubes, G. Epidemiology faces its limits. *Science* **269,** 164–169 (1995).
3. Austin, J. Schizophrenia: an update and review. *J. Genet. Couns.* **14,** 329–340 (2005).
4. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37,** 615–626 (2007).
5. Alberg, A. J. & Samet, J. M. Epidemiology of lung cancer. *Chest* **123,** 21S–49S (2003).
6. Kendler, K. S., Thornton, L. M. & Pedersen, N. L. Tobacco consumption in Swedish twins reared apart and reared together. *Arch. Gen. Psychiatry* **57,** 886–892 (2000).
7. Kendler, K. S. & Karkowski-Shuman, L. Stressful life events and genetic liability to major depression: genetic control of exposure to the environment? *Psychol. Med.* **27,** 539–547 (1997).
8. Plomin, R. & Daniels, D. Why are children in the same family so different from one another? *Behav. Brain Sci.* **14,** 373–427 (1987).
9. Plomin, R. Environment and genes. Determinants of behavior. *Am. Psychol.* **44,** 105–111 (1989).
10. Bouchard, T. J. Jr & McGue, M. Genetic and environmental influences on human psychological differences. *J. Neurobiol.* **54,** 4–45 (2003).
11. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343,** 78–85 (2000).
12. Harris, J. R. *The Nurture Assumption: Why Children Turn Out the Way They Do* 462 (Touchstone, 1999).
13. Turkheimer, E. & Waldron, M. Nonshared environment: a theoretical, methodological, and quantitative review. *Psychol. Bull.* **126,** 78–108 (2000).
14. Gartner, K. & Baunack, E. Is the similarity of monozygotic twins due to genetic factors alone? *Nature* **292,** 646–647 (1981).
In this study, the phenotypic variation among isogenic monozygotic twins in mice was compared with that of dizygotic twins, and the authors concluded that a significant proportion of phenotypic variation cannot be explained by DNA sequences and environmental factors.
15. Edwards, J. L. *et al.* Cloning adult farm animals: a review of the possibilities and problems associated with somatic cell nuclear transfer. *Am. J. Reprod. Immunol.* **50,** 113–123 (2003).
16. Rhind, S. M. *et al.* Cloned lambs — lessons from pathology. *Nature Biotechnol.* **21,** 744–745 (2003).
17. Yanagimachi, R. Cloning: experience from the mouse and other animals. *Mol. Cell. Endocrinol.* **187,** 241–248 (2002).
18. Martin, N., Boomsma, D. & Machin, G. A twin-pronged attack on complex traits. *Nature Genet.* **17,** 387–392 (1997).
19. Boomsma, D., Busjahn, A. & Peltonen, L. Classical twin studies and beyond. *Nature Rev. Genet.* **3,** 872–882 (2002).
20. Jockin, V., McGue, M. & Lykken, D. T. Personality and divorce: a genetic analysis. *J. Pers. Soc. Psychol.* **71,** 288–299 (1996).
21. Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B. & Gottesman, I. I. Socioeconomic status modifies heritability of IQ in young children. *Psychol. Sci.* **14,** 623–628 (2003).
22. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era — concepts and misconceptions. *Nature Rev. Genet.* **9,** 255–266 (2008).
This paper discusses the complexities in interpreting heritability.
23. Heath, A. C., Eaves, L. J. & Martin, N. G. Interaction of marital status and genetic risk for symptoms of depression. *Twin Res.* **1,** 119–122 (1998).
24. Stokstad, E. Biophysics: DNA on the big screen. *Science* **275,** 1882 (1997).
25. Gottesman, I. I. Twins: en route to QTLs for cognition. *Science* **276,** 1522–1523 (1997).
26. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456,** 18–21 (2008).
27. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7,** 85–97 (2006).
28. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10,** 241–251 (2009).
29. Vogel, F. & Motulsky, A. *Human Genetics: Problems and Approaches* 851 (Springer, 1997).
30. Richards, E. J. Inherited epigenetic variation — revisiting soft inheritance. *Nature Rev. Genet.* **7,** 395–401 (2006).
This review provides an informative and balanced summary of epigenetic heritability across generations and its possible role in evolution.

31. Jirtle, R. L. & Skinner, M. K. Environmental epigenomics and disease susceptibility. *Nature Rev. Genet.* **8,** 253–262 (2007).
This review summarizes the evidence that environmental factors can change the epigenetic regulation of genes, as well as that certain environmentally induced epigenetic modifications can be heritable.
32. Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nature Neurosci.* **7,** 847–854 (2004).
33. Collins, A. *et al.* Exercise improves cognitive responses to psychological stress through enhancement of epigenetic mechanisms and gene expression in the dentate gyrus. *PLoS ONE* **4,** e4330 (2009).
34. Fagiolini, M., Jensen, C. L. & Champagne, F. A. Epigenetic influences on brain development and plasticity. *Curr. Opin. Neurobiol.* **19,** 207–212 (2009).
35. Ushijima, T. *et al.* Fidelity of the methylation pattern and its variation in the genome. *Genome Res.* **13,** 868–874 (2003).
36. Wong, A. H., Gottesman, I. I. & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum. Mol. Genet.* **14,** R11–R18 (2005).
37. Bouchard, T. J. Jr, Lykken, D. T., McGue, M., Segal, N. L. & Tellegen, A. Sources of human psychological differences: the Minnesota Study of Twins Reared Apart. *Science* **250,** 223–228 (1990).
This landmark study in human research challenges ideas about the importance of environment on several physical and psychological traits, which were investigated in pairs of monozygotic and dizygotic twins who had been reared apart and together.
38. Finch, C. E. & Kirkwood, T. *Chance, Development, and Aging* (Oxford Univ. Press, 2000).
39. Gartner, K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim.* **24,** 71–77 (1990).
40. Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. & Whitelaw, E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* **2,** e49 (2006).
41. Flanagan, J. M. *et al.* Intra- and interindividual epigenetic variation in human germ cells. *Am. J. Hum. Genet.* **79,** 67–84 (2006).
42. Haaf, T. Methylation dynamics in the early mammalian embryo: implications of genome reprogramming defects for development. *Curr. Top. Microbiol. Immunol.* **310,** 13–22 (2006).
43. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403,** 501–502 (2000).
44. Rideout, W. M., Eggan, K. & Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* **293,** 1093–1098 (2001).
45. Kaminsky, Z. A. *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genet.* **41,** 240–245 (2009).
46. Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nature Genet.* **40,** 904–908 (2008).
47. Heijmans, B. T., Kremer, D., Tobi, E. W., Boomsma, D. I. & Slagboom, P. E. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human *IGF2/H19* locus. *Hum. Mol. Genet.* **16,** 547–554 (2007).
48. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877–881 (2008).
49. Surani, M. A., Durcova-Hills, G., Hajkova, P., Hayashi, K. & Tee, W. W. Germ line, stem cells, and epigenetic reprogramming. *Cold Spring Harb. Symp. Quant. Biol.* **73,** 9–15 (2008).
50. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5,** e1000530 (2009).
51. Macleod, D., Clark, V. H. & Bird, A. Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nature Genet.* **23,** 139–140 (1999).
52. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35,** 88–93 (2003).
53. Silva, A. J. & White, R. Inheritance of allelic blueprints for methylation patterns. *Cell* **54,** 145–152 (1988).
54. Youngson, N. A. & Whitelaw, E. Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* **9,** 233–257 (2008).
55. Chong, S., Youngson, N.A. & Whitelaw, E. Heritable germline epimutation is not the same as transgenerational epigenetic inheritance. *Nature Genet.* **39,** 574–575 (2007).
56. Suter, C. M. & Martin, D. I. K. Inherited epimutation or a haplotypic basis for the propensity to silence? *Nature Genet.* **39,** 573 (2007).
57. Suter, C. M. & Martin, D. I. K. Reply to "Heritable germline epimutation is not the same as transgenerational epigenetic inheritance". *Nature Genet.* **39,** 575–576 (2007).
58. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2,** e132 (2006).
59. Hong, L. E. *et al.* Sensory gating endophenotype based on its neural oscillatory pattern and heritability estimate. *Arch. Gen. Psychiatry* **65,** 1008–1016 (2008).
60. Allen, N. D., Norris, M. L. & Surani, M. A. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61,** 853–861 (1990).
61. Petronis, A. & Kennedy, J. L. Unstable genes — unstable mind? *Am. J. Psychiatry* **152,** 164–172 (1995).
62. Timshel, S., Therkildsen, C., Bendahl, P. O., Bernstein, I. & Nilbert, M. An effect from anticipation also in hereditary nonpolyposis colorectal cancer families without identified mutations. *Cancer Epidemiol.* **33,** 231–234 (2009).
63. McFaul, C. D. *et al.* Anticipation in familial pancreatic cancer. *Gut* **55,** 252–258 (2006).
64. Petronis, A., Kennedy, J. L. & Paterson, A. D. Genetic anticipation: fact or artifact, genetics or epigenetics? *Lancet* **350,** 1403–1404 (1997).
65. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33,** 70–74 (2003).
66. Jablonka, E. & Lamb, M.J. Precis of evolution in four dimensions. *Behav. Brain Sci.* **30,** 353–365; 365–389 (2007).
67. Pigliucci, M. Modelling phenotypic plasticity. II. Do genetic correlations matter? *Heredity* **77,** 453–460 (1996).
68. Pal, C. & Miklos, I. Epigenetic inheritance, genetic assimilation and speciation. *J. Theor. Biol.* **200,** 19–37 (1999).
69. De Vries, H. *Species and Varieties: Their Origin by Mutation* (Open Court, 1904).
70. Kuhn, T. S. *The Structure of Scientific Revolutions* 172 (Univ. Chicago Press, 1962).
71. Petronis, A. Human morbid genetics revisited: relevance of epigenetics. *Trends Genet.* **17,** 142–146 (2001).
72. Pauling, L. *et al.* Sickle cell anemia, a molecular disease. *Science* **109,** 543–548 (1949).
73. Risch, N. Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet. Epidemiol.* **7,** 3-16; 17–45 (1990).
74. Laird, P. W. Principles and challenges of genome-wide DNA methylation analysis. *Nature Rev. Genet.* **11,** 191–203 (2010).
75. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).

Acknowledgements I thank A. Wong, I. Gottesman, A. Paterson and C. Ptak for editorial suggestions, A. Schumacher for figure design, and the reviewers for their comments. Work in my laboratory has been supported by the Canadian Institutes of Health Research (grants 199170 and 186007) and the US National Institutes of Health (grants MH074127, MH088413, DP3DK085698 and HG004535). I am Tapscott Chair in Schizophrenia Studies at the University of Toronto and a senior fellow of the Ontario Mental Health Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (arturas_petronis@camh.net).