DNA METHYLATION AND HUMAN DISEASE

Keith D. Robertson

Abstract | DNA methylation is a crucial epigenetic modification of the genome that is involved in regulating many cellular processes. These include embryonic development, transcription, chromatin structure, X chromosome inactivation, genomic imprinting and chromosome stability. Consistent with these important roles, a growing number of human diseases have been found to be associated with aberrant DNA methylation. The study of these diseases has provided new and fundamental insights into the roles that DNA methylation and other epigenetic modifications have in development and normal cellular homeostasis.

CPG ISLAND
A genomic region of ~1 kb that has a high G–C content, is rich in CpG dinucleotides and is usually hypomethylated.

Department of Biochemistry and Molecular Biology, Shands Cancer Center, University of Florida, Box 100245, 1600 S.W. Archer Road, Gainesville, Florida 32610, USA.
e-mail: keithr@ufl.edu
doi:10.1038/nrg1655

With the completion of the Human Genome Project, we have a nearly complete list of the genes needed to produce a human. However, the situation is far more complex than a simple catalogue of genes. Of equal importance is a second system that cells use to determine when and where a particular gene will be expressed during development. This system is overlaid on DNA in the form of epigenetic marks that are heritable during cell division but do not alter the DNA sequence.

The only known epigenetic modification of DNA in mammals is methylation of cytosine at position C5 in CpG dinucleotides[1]. By contrast, the other main group of epigenetic modifications — the post-translational modification of histones — shows a high level of diversity and complexity[2]. The mammalian DNA methylation machinery is composed of two components, the DNA methyltransferases (DNMTs), which establish and maintain DNA methylation patterns, and the methyl-CpG binding proteins (MBDs), which are involved in 'reading' methylation marks. An in-depth discussion of these proteins is not provided here, but their main features are given in online supplementary information S1 (table) (and were recently reviewed by REFS 3,4). There is also clear evidence that a DNA demethylase contributes to regulating DNA methylation patterns during embryonic development, although the activity responsible for this has not been identified[5].

In normal cells, DNA methylation occurs predominantly in repetitive genomic regions, including satellite DNA and parasitic elements (such as long interspersed transposable elements (LINES), short interspersed transposable elements (SINES) and endogenous retroviruses)[6]. CPG ISLANDS, particularly those associated with promoters, are generally unmethylated, although an increasing number of exceptions are being identified[7,8]. DNA methylation represses transcription directly, by inhibiting the binding of specific transcription factors, and indirectly, by recruiting methyl-CpG-binding proteins and their associated repressive chromatin remodelling activities (see online supplementary information S1 (table)). Little is known about how DNA methylation is targeted to specific regions; however, this probably involves interactions between DNMTs and one or more chromatin-associated proteins[3].

Properly established and maintained DNA methylation patterns are essential for mammalian development and for the normal functioning of the adult organism. DNA methylation is a potent mechanism for silencing gene expression and maintaining genome stability in the face of a vast quantity of repetitive DNA, which can otherwise mediate illegitimate recombination events and cause transcriptional deregulation of nearby genes. Embryonic stem cells that are deficient for DNMTs are viable, but die when they are induced to differentiate[9]. Mouse knockout studies have shown that *Dnmt1* and *Dnmt3b* are essential for embryonic development and that mice that lack *Dnmt3a* die within a few weeks of birth[10,11]. In addition, loss of normal DNA methylation patterns in somatic cells results in loss of growth control.

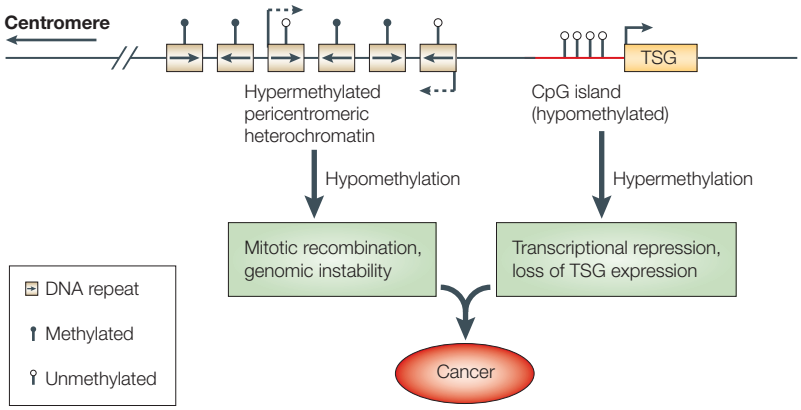

Figure 1 | **DNA methylation and cancer.** The diagram shows a representative region of genomic DNA in a normal cell. The region shown contains repeat-rich, hypermethylated pericentromeric heterochromatin and an actively transcribed tumour suppressor gene (TSG) associated with a hypomethylated CpG island (indicated in red). In tumour cells, repeat-rich heterochromatin becomes hypomethylated and this contributes to genomic instability, a hallmark of tumour cells, through increased mitotic recombination events. *De novo* methylation of CpG islands also occurs in cancer cells, and can result in the transcriptional silencing of growth-regulatory genes. These changes in methylation are early events in tumorigenesis.

The importance of DNA methylation is emphasized by the growing number of human diseases that are known to occur when this epigenetic information is not properly established and/or maintained, and there is increasing interest in developing ways of pharmacologically reversing epigenetic abnormalities[12]. Further interest in this area comes from important new evidence that concerns the regulation of DNA methylation, indicating interactions between DNA methylation and the histone modification machinery, and showing a potential role of small RNAs[13]. Here, I discuss a diverse group of diseases associated with abnormalities of DNA methylation, studies of which have provided insights into how DNA methylation patterns are regulated and the pathological consequences of their disruption. I also propose two related models for how a single common defect might be able to give rise to the diverse array of defective DNA methylation patterns in human diseases. It should be noted that I focus only on those diseases for which defective DNA methylation patterns have been demonstrated. For this reason, I do not discuss Rett syndrome, which involves a defect in the machinery that reads methylation marks[14].

DNA methylation and cancer

A link between DNA methylation and cancer was first demonstrated in 1983, when it was shown that the genomes of cancer cells are hypomethylated relative to their normal counterparts[15]. Hypomethylation in tumour cells is primarily due to the loss of methylation from repetitive regions of the genome[6] (FIG. 1), and the resulting genomic instability is a hallmark of tumour cells. Rearrangements that involve the large block of pericentromeric heterochromatin on chromosome 1, for example, are among the most frequent genomic instabilities in many tumour types[16]. Reactivation of transposon promoters following demethylation might also contribute to aberrant gene regulation in cancer by TRANSCRIPTIONAL INTERFERENCE or the generation of antisense transcripts[17]. Loss of genomic methylation is a frequent and early event in cancer, and correlates with disease severity and metastatic potential in many tumour types[18].

Gene-specific effects of hypomethylation also occur. For example, the melanoma antigen (MAGE) family of cancer–testis genes, which encode tumour antigens of unknown function, are frequently demethylated and re-expressed in cancer[19]. Demethylation accompanied by increased expression has been reported for the S100 calcium binding protein A4 (*S100A4*) gene in colon cancer[20], the serine protease inhibitor gene *SERPINB5* (also known as maspin) in gastric cancer[21], and the putative oncogene γ-synuclein (*SNCG*) in breast and ovarian cancers[22]. Global demethylation early in tumorigenesis might predispose cells to genomic instability and further genetic changes, whereas gene-specific demethylation could be a later event that allows tumour cells to adapt to their local environment and promotes metastasis.

Research on genome-wide demethylation in cancer cells has been largely overshadowed by studies of gene-specific hypermethylation events, which occur concomitantly with the hypomethylation events discussed above. Aberrant hypermethylation in cancer usually occurs at CpG islands (FIG. 1), most of which are unmethylated in normal somatic cells[8,23], and the resulting changes in chromatin structure (such as histone hypoacetylation) effectively silence transcription. Indeed, a distinct subset of many tumour types has a CpG-island-methylator phenotype, which has been defined as a 3–5 fold increase in the frequency of aberrant hypermethylation events[24]. Genes involved in cell-cycle regulation, tumour cell invasion, DNA repair, chromatin remodelling, cell signalling, transcription and apoptosis are known to become aberrantly hypermethylated and silenced in nearly every tumour type (see online supplementary information S2 and S3 (tables)). This provides tumour cells with a growth advantage, increases their genetic instability (allowing them to acquire further advantageous genetic changes), and allows them to metastasize. In tumours with a well-defined progression, such as colon cancer, aberrant hypermethylation is detectable in the earliest precursor lesions, indicating that it directly contributes to transformation and is not a late event that arises from genetic alterations[25].

Use of RESTRICTION LANDMARK GENOMIC SCANNING to analyze the methylation status of 1,184 CpG islands from 98 tumour samples showed that *de novo* methylation of CpG islands is widespread in tumour cells. In this study, the extent of methylation varied between individual tumours and tumour types, and an average of 608 CpG islands were aberrantly hypermethylated[26]. Because the hypermethylation of CpG islands is relatively rare in normal cells, is an early event in transformation, and robust assays can detect methylated DNA in bodily fluids, it represents a good potential biomarker for early cancer detection[27]. The underlying cause of methylation defects in cancer remains unknown, but

TRANSCRIPTIONAL INTERFERENCE
Repression of one transcriptional unit by another such unit that is linked in *cis*.

RESTRICTION LANDMARK GENOMIC SCANNING (RLGS). A genome-wide method for analyzing the DNA methylation status of CpG islands. Radiolabelled fragments obtained by digestion with *Not*I (a methylation-sensitive restriction enzyme) are separated by two-dimensional gel electrophoresis, allowing differentiation between methylated and unmethylated regions.

Box 1 | **A comparison of the CTCF and BORIS proteins**

CTCF (CCCTC-binding factor)

- Contains an 11 zinc finger (ZF) DNA-binding region[30].
- Binds diverse DNA sequences, including most ICRs and many CpG islands[30,137].
- Binding is methylation sensitive (if binding sites contain CpGs)[31].
- Has insulator function (regulates access of enhancers to promoters)[30,70].
- Has boundary element function (blocks the spread of heterochromatin)[30,70].
- Protects regions from DNA methylation[138,139].
- Is ubiquitously expressed, except in spermatocytes[126].
- Overexpression inhibits cell proliferation[140].
- Gene is located at 16q22, a region of frequent loss of heterozygosity in cancer, and is often mutated in cancer[141].

BORIS

- Contains an 11 ZF DNA-binding region that has high similarity to the CTCF ZF region[30].
- Presumed to show significant overlap with CTCF binding sites.
- Methylation sensitivity of binding unknown.
- Insulator function unknown.
- Boundary element function unknown
- Methylation protection function unknown
- Expression is normally restricted to testis (BORIS is a member of the cancer–testis gene family), and is reactivated in cancers[126].
- Overexpression promotes cell proliferation[30].
- Gene is located at 20q13, a region that is frequently amplified in cancer.

possible mechanisms are discussed at the end of this review. DNA methylation and cancer are also related through the loss of imprinted methylation patterns in many tumours, as discussed below.

DNA methylation and imprinting disorders

Genomic imprinting — a brief overview. Genomic imprinting is defined as an epigenetic modification of a specific parental chromosome in the gamete or zygote that leads to differential expression of the two alleles of a gene in the somatic cells of the offspring[28]. Differential expression can occur in all cells, or in specific tissues or developmental stages. About 80 genes are known to be imprinted, and this number is continually increasing.

Differential allele-specific DNA methylation is one of the hallmarks of imprinted regions and is usually localized to regions termed differentially methylated regions (DMRs). DMRs include imprinting control regions (ICRs), which control gene expression within imprinted domains, often over large distances. Whereas DMRs can be reprogrammed during development, differential methylation in ICRs is generally established in germ cells and maintained throughout development[29].

CCCTC-binding factor (CTCF), an 11-zinc-finger protein that binds to highly divergent sequences using different combinations of zinc fingers, is an important regulator of imprinted gene expression (BOX 1). CTCF is a chromatin insulator that separates the genome into independent functional domains or, in the case of imprinted loci, regulates the ability of distant ENHANCERS to access promoters[30]. In several ICRs, *in vitro* and *in vivo* studies have shown that CTCF binds only to the unmethylated parental allele[31,32], providing an elegant means for ICR-regulated genes to be expressed in an allele-specific manner. CTCF binding might also be essential to protect DMRs from *de novo* methylation[33].

Loss of imprinting (LOI) is the disruption of imprinted epigenetic marks through gain or loss of DNA methylation, or simply the loss of normal allele-specific gene expression[28]. Studies of LOI have provided important insights into the role of methylation in imprinting, and have highlighted the importance of CTCF as a regulator of both imprinted gene expression and DNA methylation patterns. Below, I discuss several human diseases for which the role of aberrant genomic imprinting has been particularly well established. Features of these diseases are summarized in the online supplementary information S3 (table). Importantly, imprinted DNA methylation patterns are frequently disrupted by *in vitro* manipulation of embryos in both animals and humans[34–36], and the use of assisted reproductive technologies has been associated with an increased risk for several imprinting diseases[37] (BOX 2).

LOI in cancer. LOI of a growth-promoting imprinted gene, leading to activation of the normally silent allele, results in abnormally high expression of the gene product and gives cells a growth advantage. The imprinted *IGF2/H19* locus (FIG. 2a) encodes both IGF2, an autocrine growth factor with an important role in many types of cancer, and *H19*, a non-coding RNA of unknown function with growth suppressive properties[38]. *IGF2* and *H19* are normally expressed from the paternal and maternal alleles, respectively, and relaxed silencing of the maternal *IGF2* allele results in increased *IGF2* expression and reduced *H19* expression. LOI of *IGF2* is the most common LOI event across the widest range of tumour types, including colon, liver, lung, and ovarian cancer, as well as Wilms' tumour — the embryonic kidney cancer in which LOI was first discovered[39,40].

Additional evidence for the role of LOI at *IGF2/H19* in cancer came from a recent study that used a mouse model. Animals with LOI showed a twofold increase in intestinal tumours, and normal intestinal epithelium was shown to be less well differentiated than in wild-type mice[41]. LOI defects in *IGF2/H19* vary with tumour type. For example, in Wilms' tumour[42] and colorectal cancer[43], the regulatory region ICR1 on the maternal allele becomes methylated *de novo*, which inhibits CTCF binding and thereby allows enhancers upstream of *H19* to access the *IGF2* promoter on both alleles (FIG. 2a). An alternative mechanism has also been proposed, in which DMRs upstream of *IGF2* become demethylated[44].

LOI of growth-inhibitory imprinted genes, through silencing of the one normally active allele, might also result in deregulated cell growth. An example of this involves the maternally expressed cyclin-dependent

ENHANCER
A regulatory DNA element that usually binds several transcription factors and can activate transcription from a promoter at great distance and in an orientation-independent manner.

kinase inhibitor 1C (*CDKN1C*) (also known as $p57^{KIP2}$) gene, which encodes a cyclin-dependent kinase inhibitor that mediates G1/S-phase arrest. In Wilms' tumour, LOI of *CDKN1C* occurs in ~10% of tumours (LOI of *IGF2/H19* occurs in 70% of these tumours)[28,45]. Other imprinted genes that undergo LOI include the RAS-related gene *DIRAS3* (also known as *ARHI*) in breast and ovarian cancer[46] and the mesoderm specific transcript homologue (*MEST*) gene in breast, lung, and colon cancer[47]. Although the number of genes that show LOI in cancer seems small from this discussion, the frequency of such alterations is probably similar in extent to the DNA methylation abnormalities of non-imprinted genes discussed in the previous section. Examples of LOI seem limited only because relatively few of the estimated total number of imprinted genes have been identified and even fewer have been well characterized.

Beckwith–Wiedemann syndrome (BWS). BWS is predominantly a maternally transmitted disorder and involves fetal and postnatal overgrowth and a predisposition to embryonic tumours, such as Wilms' tumour (see online supplementary information S3 (table)). The BWS locus, at 11p15.5, spans ~1 Mb and includes several imprinted genes (FIG. 2a). This large region consists of two independently regulated imprinted domains[39,40,48]. The more telomeric domain contains *IGF2* and *H19*, followed by an intervening region with several genes that do not seem to be imprinted. The second, more centromeric, domain contains the maternally expressed gene *KCNQ1* (potassium voltage-gated channel, KQT-like subfamily, member 1; also known as K_vLQT1), the paternally expressed *KCNQ1* antisense transcript *KCNQ1OT1* (KCNQ1 overlapping transcript 1; also known as *LIT1*), and several other maternally expressed genes, including *CDKN1C*. ICR1, which is located upstream of *H19*, controls differential expression of the *IGF2/H19* domain, and there are also several DMRs upstream of *IGF2*; ICR2 lies at the 5′-end of *KCNQ1OT1* (REFS 43,48,49). In normal cells, the paternal and maternal alleles of ICR1 and ICR2, respectively, are methylated, and both contain binding sites for CTCF[31,50].

Several perturbations of this complex genomic region are associated with BWS. These include paternal UNIPARENTAL DISOMY (UPD) (resulting in increased *IGF2* and reduced *CDKN1C* expression), LOI of *KCNQ1OT1* in ~50% of cases that do not involve UPD, and LOI at

Box 2 | DNA-methylation abnormalities and assisted reproductive technology

A link between assisted reproductive technology and birth defects

It has been estimated that one in ten individuals of reproductive age are infertile. The use of assisted reproductive technologies (ARTs), such as intracytoplasmic sperm injection and *in vitro* fertilization, has therefore been increasing, and children conceived by ART now account for 1–3% of all births in some Western countries. Although the vast majority of these children develop normally, recent studies have raised concerns that these procedures might lead to epigenetic defects. ART is associated with an increase in multiple births and low birth weight, and this risk has been known for some time[142]. Although the numbers are small and larger studies are needed, ARTs have been linked to a 3–6-fold increase in the occurrence of Beckwith–Wiedemann syndrome (BWS) and Angelman syndrome (AS)[37,143,144] (see online supplementary information S3 (table)).

An epigenetic basis

Germ cell and embryonic development are two periods that are characterized by profound changes in DNA methylation patterns, also known as epigenetic reprogramming. Primordial germ cells are globally demethylated as they mature (including at imprinted regions), and then become methylated *de novo* during gametogenesis, which is also the time at which most DNA methylation imprints are established. The second period of pronounced change occurs after fertilization, where the paternal genome is rapidly and actively demethylated, followed by passive replication-dependent demethylation of the maternal genome (here, imprinting marks seem to resist demethylation). A wave of *de novo* methylation once again establishes the somatic-cell pattern of DNA methylation following implantation[145,146].

The increased incidence of BWS and AS associated with ART all arise from epigenetic, rather than genetic, defects. In most BWS cases studied, loss of imprinting (LOI) at *KCNQ1OT1* was observed, and in AS cases the defect was found to be LOI at *SNRPN* (loss of maternal methylation, see FIG. 2a,b)[37,143]. Epigenetic mechanisms and imprinted genes also seem to be involved in the regulation of birth weight in mouse models; therefore, the disruption of imprints at currently unknown loci following ART in humans might also contribute to the increased incidence of this problem[147]. Low birth weight is a serious issue in its own right as it contributes to long-term adverse health effects such as cardiovascular disease[148].

Causes of epigenetic defects

There has been a trend in fertility clinics towards extended culture time of embryos and transfer of blastocysts, rather than earlier cleavage-stage embryos. This allows selection of the most robust embryos, proper temporal synchronization of the embryo and the uterus, and the possibility of maintaining a high pregnancy rate while reducing the risk of multiple pregnancy (because fewer embryos need to be transferred)[35]. However, several studies have demonstrated that *in vitro* culture conditions have notable effects on imprinting marks, gene expression and developmental potential in mouse models[34,149]. Whether the culture conditions used for human embryos provide all of the necessary factors for proper epigenetic control is unknown. It is also possible that the epigenetic defects might be the cause of the infertility itself and are present in the gametes of the mother and/or father. Therefore, rather than ARTs causing epigenetic defects directly, they might simply be unmasking an existing defect.

UNIPARENTAL DISOMY
Inheritance of a chromosome or chromosome region from a single parent.

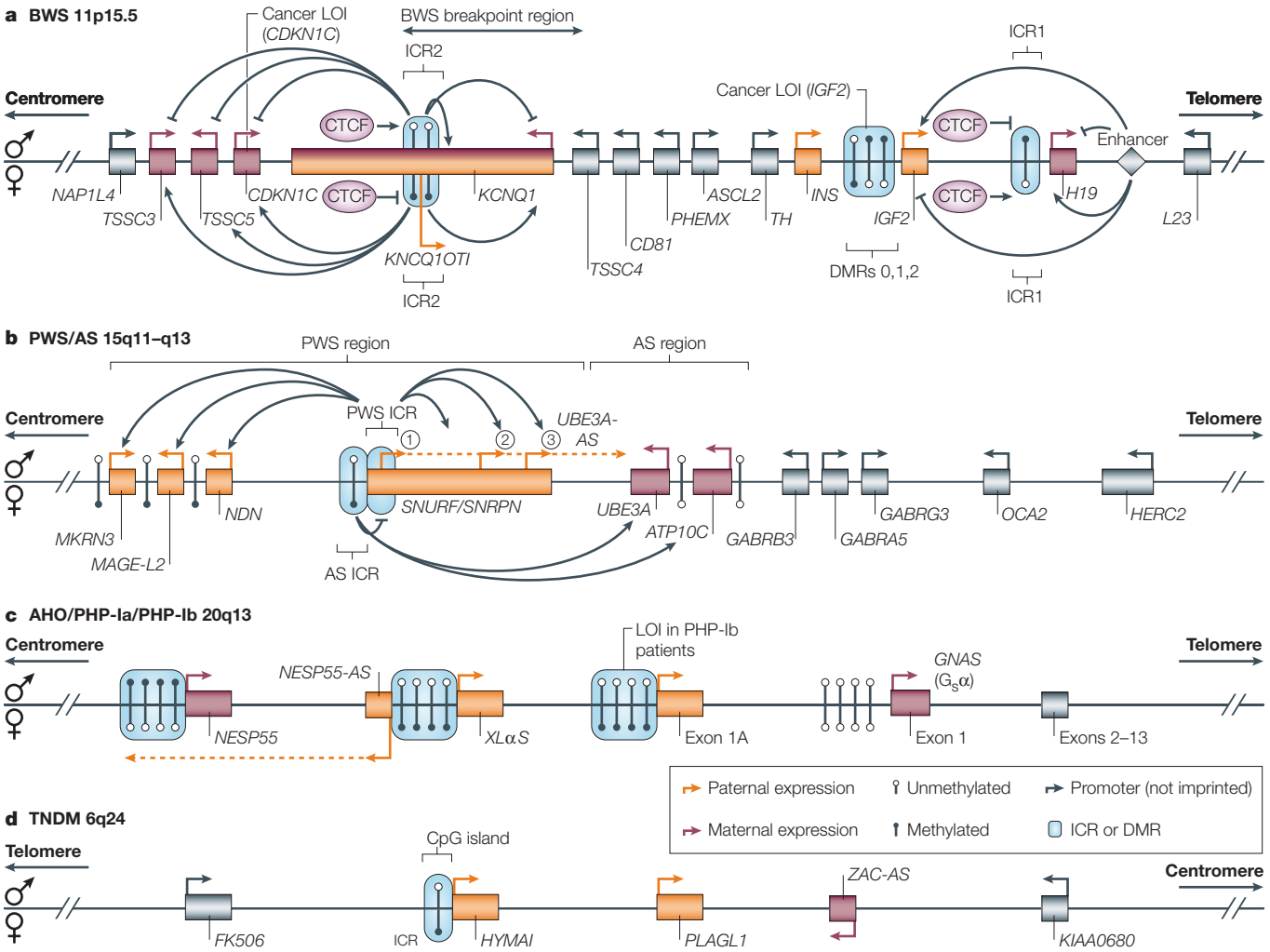

Figure 2 | **Regions involved in disease-associated genomic imprinting defects.** Sense and antisense transcripts are indicated by arrows above and below the genes, respectively. In each panel, the methylation status of the paternal allele is indicated on the top of the chromosome, and that of the maternal allele on the bottom of the chromosome. **a** | The Beckwith–Wiedemann syndrome (BWS) locus and cancer-associated loss of imprinting (LOI). The region can be divided into two imprinted domains (*TSSC3* to *KCNQ1*, and *INS* to *H19*) that are regulated independently. CTCF binds to the unmethylated alleles of both imprinting control regions, ICR1 and ICR2. Cancer-associated LOI commonly occurs within the *IGF2/H19* region (resulting in increased *IGF2* and decreased *H19* expression) and less commonly at *CDKN1C*. BWS arises from several defects in this region, including LOI at ICR1, ICR2 and the differentially methylated regions (DMRs 0, 1 and 2) upstream of *IGF2*; paternal UPD; and mutations and translocations on the maternal allele. **b** | The Prader–Willi syndrome (PWS)/Angelman syndrome (AS) locus. The ICR for this locus has a bipartite structure, with the more centromeric component acting as the AS ICR and the more telomeric region, which contains the promoter of the *SNURF/SNRPN* gene, acting as the PWS ICR. *SNURF/SNRPN* produces an extremely long and complicated transcript that encodes not only the SNURF/SNRPN (1), and *IPW* transcripts (2), but also several small nucleolar RNAs (snoRNAs) (3), which actually flank *IPW* (not shown), that are processed from its introns. This transcript is also thought to inhibit expression of *UBE3A* and *ATP10C* on the paternal allele through an antisense mechanism. PWS is a multi-gene disorder that arises from the loss of expression of genes in this region from the paternal allele. AS arises from loss of maternally expressed *UBE3A*. **c** | The GNAS locus, which is the site of the molecular defects that underlie Albright hereditary osteodystrophy (AHO), pseudohypoparathyroidism Ia (PHP-Ia) and PHP-Ib. Four alternative first exons (those of *NESP55*, *XLαs*, exon 1A and exon 1), each driven by its own promoter, splice onto a common set of downstream exons (2–13). The promoters of *NESP55*, *XLαs/NESP55-AS* and exon 1A are differentially methylated and imprinted. Exon 1 ($G_s\alpha$) demonstrates preferential maternal expression in only certain tissues (such as the renal proximal tubules); however, it is not differentially methylated. The location of the LOI defect at exon 1A in PHP-Ib patients is shown. AHO and PHP-Ia arise from paternal and maternal transmission of $G_s\alpha$ mutations, respectively, while PHP-Ib is due to LOI at exon 1A and, in some cases, also at the upstream DMRs. **d** | The transient neonatal diabetes mellitus (TNDM) locus. A differentially methylated CpG island upstream of *HYMAI* acts as an ICR for this region, which contains three imprinted genes. In TNDM patients, DNA methylation is lost from the maternal allele and expression of *PLAGL1* is upregulated. *CDKN1C*, cyclin-dependent kinase inhibitor 1C; *HYMAI*, hydatidiform mole associated and imprinted; *IGF2*, insulin-like growth factor 2; *INS*, insulin; *IPW*, imprinted in Prader–Willi syndrome; *KCNQ1*, potassium voltage-gated channel, KQT-like subfamily, member 1; *NESP55*, neuroendocrine secretory protein 55; *PLAGL1*, pleiomorphic adenoma gene-like 1; *SNRPN*, small nuclear ribonucleoprotein polypeptide N; *SNURF*, *SNRPN* upstream reading frame; *TSSC3*, tumour-suppressing STF cDNA 3 (also known as *PHLDA2*); *UBE3A*, ubiquitin protein ligase E3A.

IGF2 in ~20% of these cases (both resulting in biallelic expression of the respective genes). Mutations and translocations that involve the maternal allele account for the remaining cases. LOI at ICR2 involves loss of maternal-allele-specific DNA methylation, whereas at ICR1 the defect is usually hypermethylation of the maternal allele, as seen in most Wilms' tumours[49,51]. Defects in the centromeric imprinted domain (silencing of maternally expressed genes) are thought to give rise predominantly to the BWS phenotype (anatomical malformation) whereas defects in the telomeric imprinted domain (activation of maternally repressed *IGF2*) could be the main driving force for tumorigenesis [52]. Hypomethylation of the maternal ICR2 and activation of *KCNQ1OT1* causes a marked downregulation of genes that are centromeric to this, such as *CDKN1C*; however, the molecular mechanism of this regulation is not understood[53]. One model is that *ICR2* functions in a similar way to ICR1 — that is, it might regulate, through binding of CTCF, the ability of an enhancer to access the *CDKN1C* and *KCNQ1OT1* promoters differentially on the two alleles[49].

Prader–Willi syndrome (PWS). PWS occurs in ~1 in 20,000 births and is characterized by a failure to thrive during infancy, hyperphagia and obesity during early childhood, mental retardation, and behavioural problems (see online supplementary information S3 (table))[54]. The molecular defect is complex and involves a ~2 Mb imprinted domain at 15q11–q13 that contains both paternally and maternally expressed genes (FIG. 2b). One maternally expressed gene, ubiquitin protein ligase E3A (*UBE3A*), is imprinted only in the brain and is discussed in the next section.

PWS arises from loss of paternally expressed genes in this region, although no single gene has been shown to cause PWS when its expression is lost[55]. The analysis of samples from PWS patients with microdeletions allowed the PWS ICR to be mapped to a segment of ~4 kb that spans the first exon and promoter of the small nuclear ribonucleoprotein polypeptide N (*SNRPN*)/SNRPN upstream reading frame (*SNURF*) locus[56]. In normal cells, the 5′-end of the ICR, which is needed for maternal gene expression and is involved in Angelman syndrome (discussed below), is methylated on the maternal allele[55,57]. The 3′ end of the ICR is required for expression of the paternally expressed genes (the PWS region shown in FIG. 2b) and is also the origin of the extremely long *SNURF/SNRPN* transcript. The maternally expressed genes are not differentially methylated, and their silencing on the paternal allele is probably regulated by an antisense RNA generated from *SNURF/SNRPN*[58].

The most common molecular defects in PWS are paternal deletions spanning the 15q11–q13 region (65–70% of cases) and maternal UPD (~20–30% of cases). Paternally derived microdeletions of the ICR, LOI at the ICR (*de novo* methylation of the paternal allele), and BALANCED TRANSLOCATIONS that disrupt *SNURF/SNRPN*, account for the remainder[51]. Mouse knockout studies indicate that *Ndn*[59] and the region from *Snurf/Snrpn* to *Ube3a*[60] contribute to aspects of PWS, but further models are needed to define the contribution of all paternally expressed genes in this region. The finding that a DNA methylation inhibitor restores expression of the silent maternal gene copy of *SNURF/SNRPN* is intriguing, and indicates that such inhibitors might be useful for PWS treatment[61].

Angelman syndrome (AS). AS occurs in ~1 in 15,000 births and its main characteristics include mental retardation, speech impairment and behavioural abnormalities (see online supplementary information S3 (table))[55,62]. The AS defect lies within the imprinted domain at 15q11–q13 and is due to the loss of maternally expressed genes (FIG. 2b). Unlike PWS, however, AS arises from the loss of a single gene — maternally expressed *UBE3A* — which is imprinted only in the brain and encodes an E3 ubiquitin ligase involved in the UBIQUITIN–PROTEASOME DEGRADATION PATHWAY .

The most common molecular defects that give rise to AS are maternally-derived deletions of 15q11–q13 (~65–70% of cases), paternal UPD (~5% of cases), maternal *UBE3A* mutations (~10% of cases, and up to ~40% in one study), and imprinting defects (~5% of cases). About 10% of cases arise from an unknown defect[62]. Analysis of samples from AS patients with microdeletions allowed the AS ICR to be mapped to an 880 bp region ~35 kb upstream of *SNURF/SNRPN*[63]. Imprinting defects involved in AS include loss of maternal DNA methylation or maternal ICR deletion, and patients with these defects tend to have the mildest AS phenotypes[62]. Important unanswered questions related to AS biology include the nature of the molecular defects in the ~10% of cases that have no identifiable *UBE3A* mutation and the role of the nearby gene *ATP10C* (ATPase, class V, type 10A), which is often co-deleted with *UBE3A*, in modulating the AS phenotype.

Albright hereditary osteodystrophy (AHO), pseudohypoparathyroidism type Ia (PHP-Ia) and PHP-Ib. These three diseases are related because they all arise from defects at a complex imprinted locus on chromosome 20q13 called *GNAS*, which encodes the α subunit ($G_s\alpha$) of the heterotrimeric GTP-binding protein G_s (FIG. 2c; and see online supplementary information S3(table)). AHO is a paternally inherited disorder that occurs within PHP-Ia pedigrees, and involves mental retardation and subcutaneous ossification.

PHP-Ia is maternally transmitted in a dominant manner and patients show all the AHO symptoms, plus resistance to the peripheral action of several hormones involved in activating pathways coupled to G_s. Individuals with PHP-Ib, which is also maternally inherited, show normal $G_s\alpha$ activity and only renal parathyroid hormone resistance, with no other endocrine deficiencies[64].

The *GNAS* locus encodes three proteins and an untranslated RNA from four alternative first exons that splice onto common exons 2–13 (FIG. 2c). *NESP55*, generated from the most centromeric of the first exons, encodes a CHROMOGRANIN-like protein, *XLαs* encodes a

BALANCED TRANSLOCATION
A condition in which two pieces of chromosomal material have switched places, but the correct number of chromosomes has been maintained.

UBIQUITIN–PROTEASOME DEGRADATION PATHWAY
Degradation pathway in which a protein that has been post-translationally modified with several ubiquitin polypeptides is targeted for destruction to the proteasome, a large cytosolic protein complex with several proteolytic activities.

CHROMOGRANINS
A group of acidic, soluble, secretory proteins that are produced by neurons and neuroendocrine cells.

Golgi-specific $G_s\alpha$ isoform, and exon 1 encodes *Gsα*. Transcripts derived from exon 1A do not seem to be translated. Each alternative first exon is driven from its own promoter, and the first three promoters are differentially methylated. *NESP55* is maternally expressed and its promoter region is methylated on the paternal allele. $G_s\alpha$ (exon 1) is preferentially expressed from the maternal allele, although its promoter region seems to be unmethylated on both alleles. $G_s\alpha$ is imprinted only in certain hormone-responsive tissues, such as the renal proximal tubules, which is the probable cause of the tissue-specific effects of these disorders. By contrast, *XLαs*, a *NESP55* antisense transcript, and exon 1A are all paternally expressed, and their promoters are methylated on the maternal allele[64].

The molecular defects that underlie AHO, PHP-Ia, and PHP-Ib, are well characterized. AHO and PHP-Ia arise from the paternal and maternal transmission of $G_s\alpha$ mutations, respectively[64]. By contrast, PHP-Ib involves LOI at exon 1A (loss of maternal-specific methylation), which occurs in all cases. Interestingly, it was recently demonstrated that sporadic PHP-Ib (the most common form) also involves imprinting defects at the upstream DMRs, in the form of biallelic methylation at *NESP55* and hypomethylation at the *NESP* antisense (*NESPAS*)/*XLαs* promoter region**.** By contrast, familial PHP-Ib is associated with maternally transmitted deletions within the nearby syntaxin 16 (*STX16*) gene and LOI at the exon 1A DMR only[65]. *NESP55*, *XLαs* and exon 1A are probably regulated by antisense transcription and/or differential promoter methylation. However it is still unclear how $G_s\alpha$ exon 1 is regulated, as it is not differentially methylated and no antisense transcript has been identified. The deleted region in *STX16* might contain a *cis*-element that acts over a large distance to establish imprinting at the exon 1A DMR[65]. LOI at exon 1A (and biallelic $G_s\alpha$ expression) also occurs in a subset of growth-hormone-secreting pituitary adenomas, supporting the idea that imprinting of each *GNAS* promoter is independently regulated[66].

Transient neonatal diabetes mellitus (TNDM). TNDM is a rare form of diabetes (occuring in ~1 in 400,000 births) that presents during the first few weeks after birth. Remission usually occurs after 3–6 months, but a significant proportion of patients develop diabetes later in life[67]. The TNDM locus on chromosome 6q24 spans ~400 kb and contains three imprinted genes — the paternally expressed genes hydatidiform mole associated and imprinted (*HYMAI*) and pleiomorphic adenoma gene-like 1 (*PLAGL1*; also known as *ZAC* or *LOT1*), and the maternally expressed PLAGL1 antisense transcript (*ZAC-AS*) (FIG. 2d). *HYMAI* gives rise to an apparently untranslated RNA of unknown function, whereas *PLAGL1* encodes a zinc finger protein that is a putative tumour suppressor and transcriptional regulator involved in modulating the cell cycle and apoptosis. A putative ICR within the CpG island at the 5′-end of *HYMAI* is methylated on the maternal allele and acts as a transcriptional repressor when methylated[68].

Three mechanisms give rise to TNDM: paternal UPD of chromosome 6, paternal duplications of 6q24, and LOI at the TNDM CpG-island-associated ICR, which leads to increased *PLAGL1* expression[67]. TNDM patients who lack chromosomal abnormalities demonstrate LOI of the ICR, specifically loss of DNA methylation from the maternal allele[68]. Because the methylated ICR acts as a repressor of *PLAGL1* transcription, LOI has the same consequences as paternal UPD and 6q24 duplication.

The TNDM locus has a much simpler organization than those that are involved in BWS, PWS and AHO/PHP. However, this further emphasizes that most imprinted genes are found in clusters under the control of regulatory regions that are often distant from the genes themselves, and that this control is dependent on allele-specific DNA methylation.

A wider role of imprinting in disease. These examples of diseases associated with aberrant imprinting are probably only the tip of the iceberg. Unlike the global genomic hypomethylation and widespread CpG island hypermethylation seen in cancer and cancer-associated LOI, which probably affects many imprinted regions, the DNA methylation defects in imprinting disorders are subtle. Numerous other pathologies demonstrate parent-of-origin effects or result from UPD, indicating that imprinted genes might be involved[69]. Interestingly, accumulating evidence from the diseases described above indicates that loss of normal CTCF function has an important role in imprinting disorders. Studies of the mechanism by which CTCF regulates DNA methylation patterns in imprinted regions also indicate that it could have a much broader role in regulating DNA methylation patterns throughout the genome than is currently appreciated[70].

Methylation and repeat-instability diseases

DNA methylation defects have been linked to members of a large group of human diseases that are associated with repeat instability. Expansion of trinucleotide repeats (TNRs) during gametogenesis leads to mutation or silencing of associated genes, with pathological consequences[71]. The causes of TNR instability remain largely unknown, although DNA polymerase slippage, DNA repair mechanisms, transcription, nucleosome positioning and aberrant OKAZAKI-FRAGMENT processing within the repeat have been proposed[72]. Some of these diseases involve expansions of non-methylatable repeats, such as CAG in Huntington's disease (HD) and several forms of spinocerebellar ataxia (SCA1, 2, 3, 6, 7, and 17), CTG in myotonic dystrophy type 1 (DM1) and SCA8, and GAA in Friedreich's ataxia (FRDA)[71]. Expansion of methylatable CGG repeats occurs at several sites that are known or likely to be involved in human disease, and increased DNA methylation is thought to have a role in pathogenesis. In contrast to the TNR expansion diseases, facioscapulohumeral muscular dystrophy (FSHD) is due to the contraction of a much larger repeat accompanied by hypomethylation. These diseases offer insights into possible mechanisms for the targeting of *de novo* DNA methylation.

OKAZAKI FRAGMENTS
Short pieces of DNA that are synthesized on the lagging strand at the replication fork.

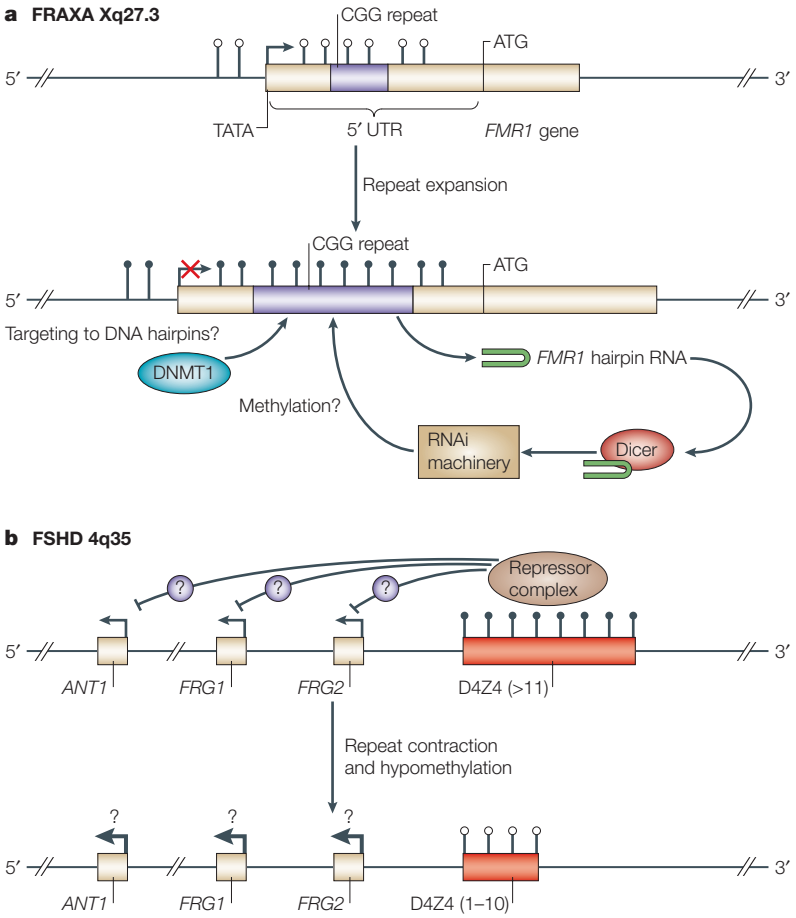

Figure 3 | **DNA methylation and diseases associated with repeat instability.** **a** | Trinucleotide repeat instability associated with aberrant DNA methylation in FRAXA. A schematic representation of the 5′-end of the *FMR1* gene is shown. The CGG repeat in *FMR1* lies in the 5′ untranslated region (UTR), which is embedded in a hypomethylated CpG island in the normal state. In fully affected individuals with large repeat expansions, the repeat and CpG island become *de novo* methylated, resulting in gene silencing. Models for how the repeat is targeted for methylation include the formation of unusual DNA structures that attract DNA methyltransferase 1 (DNMT1), and the formation of hairpin RNA from the transcribed repeat, which might be cleaved by Dicer and subsequently recruit the RNAi silencing machinery, including DNMTs. **b** | Facioscapulohumeral muscular dystrophy (FSHD) is caused by contraction of the D4Z4 repeat on chromosome 4q35 and loss of DNA methylation. Whereas repeat contraction is not always present in FSHD, repeat hypomethylation is an invariant feature. D4Z4 repeat contraction and hypomethylation are believed to cause the abnormal expression of adjacent genes (*FRG2* is considered a good candidate) or genes in other regions of the genome. One model has proposed that a repressor complex binds to the repeats and regulates the expression of nearby genes, although the role of DNA methylation was not addressed in this study. *ANT1*, solute carrier family 25, also known as *SLC25A4*; *FRG2*, FSHD region gene 2 protein.

Fragile X syndrome (FRAXA). FRAXA, an X-linked disorder, is a common cause of inherited mental retardation and occurs in ~1 in 4000 males (see online supplementary information S3 (table))[73,74]. The disease maps to the fragile X mental retardation 1 (*FMR1*) gene at Xq27.3, which is also a FOLATE-SENSITIVE FRAGILE SITE. The FMR1 protein (FMRP) is an RNA-binding protein expressed in many fetal and adult tissues, with particularly high levels in brain and testis. It seems to be involved in regulating translation and mRNA transport, and is believed to regulate synaptic plasticity by controlling translation at synapses[75].

FOLATE-SENSITIVE FRAGILE SITE
A region of chromatin that fails to compact normally during mitosis and that can be observed after culturing cells in media that is deficient in folic acid and thymidine.

The molecular basis of FRAXA lies in the highly polymorphic CGG repeat within the 5′-untranslated region of *FMR1* (FIG. 3a). Normally, 6–52 copies of the repeat are present, which increases to 52–200 copies in the premutation state (now recognized as a separate disease called fragile X tremor/ataxia syndrome[76]), and to more than 200 repeats in individuals with FRAXA[73,77]. Affected individuals show *de novo* methylation of the expanded CGG repeat and silencing of *FMR1* transcription[77]. DNA methylation is clearly an important mediator of silencing as DNMT inhibitors reactivate transcription of the full-mutation allele, although FMRP is not produced[78].

Other folate-sensitive fragile sites. The FRAXA locus is one of more than 20 folate-sensitive fragile sites in the genome. CGG repeat expansion and hypermethylation have been documented for several other such sites associated with genes (FRAXE, FRAXF, FRA10A, FRA11B and FRA16A) although only FRAXE has been firmly linked to human disease (nonspecific X-linked mental retardation)[74,79]. It is reasonable to assume that all folate-sensitive fragile sites similarly occur at CGG repeat expansions that are accompanied by DNA methylation. All fragile sites characterized so far that have a genomic structure similar to the *FMR1* gene (a CGG repeat in the untranslated region and a nearby CpG island) show hypermethylation of both regions in the repeat-expanded state, and transcription of the associated gene is silenced[74].

Facioscapulohumeral muscular dystrophy. FSHD is an autosomal dominant disease that affects ~1 in 20,000 individuals (see online supplementary information S3 (table)). The FSHD locus maps to 4q35 and the underlying molecular defect is well established, involving contraction of the polymorphic 3.3 kb D4Z4 repeat. Unaffected individuals have 11–150 copies of this repeat, whereas in most FSHD cases the repeat is reduced to 1–10 copies, although some patients (~5%) do not show repeat contraction. The abnormality that is common to all FSHD patients is loss of DNA methylation from D4Z4, indicating that aberrant hypomethylation is the main causal factor[80]. However, the genes affected by this hypomethylation are unknown.

Most studies have focused on the idea that contraction and hypomethylation of D4Z4 alters the transcription of one or more genes that flank the repeat. Genes in the vicinity of D4Z4 (FIG. 3b) include the adenine nucleotide translocator SLC25A4 (also known as *ANT1*), the overexpression of which induces apoptosis (characteristic of dystrophic muscle); FSHD region gene 1 (*FRG1*) which is ubiquitously expressed and might be involved in RNA processing; and *FRG2* (which is closest to D4Z4), the function of which is unknown but which has been shown to induce morphological changes in transfected myoblasts[81]. Results from different laboratories have yielded conflicting data about the degree of overexpression of these genes in FSHD muscle, although *FRG2* does seem to be a promising candidate for the disease gene[81–83].

There are also diverse models for how repeat contraction and hypomethylation alter gene expression and give rise to FSHD. One possible mechanism involves the binding of a repressor complex to the D4Z4 repeat, which is reduced as the repeat size decreases, causing elevated expression of nearby genes (although this study did not address the role of methylation)[82]. A looping model has also been proposed, whereby the D4Z4 repeat binds transcriptional regulatory factors and loops over to distant genes, which might be well outside the immediate vicinity of D4Z4. Repeat contraction and hypomethylation might disrupt the binding of crucial regulatory proteins or negatively affect the ability of the repeat to interact with other regulatory regions[83].

Is there a wider role for DNA methylation in diseases of repeat instability? Studies of repeat-instability diseases demonstrate that repetitive DNA, in the form of both naturally occurring repeats and pathologically expanded TNRs, is efficiently targeted by the methylation machinery. Why only some expanded TNRs are targeted for *de novo* methylation remains unknown. It might simply be the availability of a methylatable site within or near the repeat expansion that makes CGG rather than CTG or CAG TNRs attractive to the methylation machinery. However, several lines of evidence indicate that the silencing machinery might be recruited to expanded TNRs, regardless of their sequence. For example, a CpG island adjacent to the CTG repeat involved in DM1 (which also contains binding sites for CTCF) becomes hypermethylated in the severe congenital form of the disease[84,85]. In both DM1 and FRDA, the associated genes (*DMPK* and *FRDA*, respectively) are downregulated, and the expanded repeat adopts a repressive chromatin structure[86]. This indicates that TNR expansions might commonly be targeted for silencing at the chromatin level, and that additional factors might influence whether DNA methylation also occurs. In addition, DNA methylation stabilizes expanded repeats in model systems regardless of whether they contain a CpG site[87,88].

Other aspects that might influence methylation targeting include whether the repeats involved are transcribed, as they are in FRAXA and DM1, or transcribed and translated, as in HD, where there is no evidence of disease-associated hypermethylation[89]. Another factor might be the propensity of a repeat to form DNA hairpins, which have been shown to be preferred substrates for DNMT1[90]. The finding that FMR1 mRNA also forms hairpin structures, and the discovery that they are substrates for the RNAI -associated DICER ribonuclease provides a possible alternative explanation. Components of the histone-modification machinery that cause transcriptional repression are recruited to repeats by the RNAi pathway, which probably also recruits the DNA methylation machinery, although this has not been demonstrated directly (see later for a further discussion)[13,91]. In support of this model, the affected repeats and/or adjacent sequences adopt tightly packed chromatin structures in FRAXA, DM1 and FRDA[78,86]. Conversely, reduction in D4Z4 repeat number in FSHD could reduce the efficiency of RNAi targeting, eventually resulting in hypomethylation. The molecular defects that underlie these diseases are therefore consistent with the growing body of evidence that the insulator factor CTCF and/or RNAi are involved in regulating DNA methylation patterns.

Specific defects of the methylation machinery

Human diseases that arise from mutations in the DNA methylation machinery have provided important insights into the functional specialization of DNMTs. They have also shown that improper regulation of DNA methylation can occur in specific cell types, with pathological consequences.

Systemic lupus erythematosus (SLE). SLE is an autoimmune disease, with diverse clinical manifestations, that is 8–10 times more frequent in females than in males (~1 in 2000 in the overall population) (see online supplementary information S3 (table)). Common features in all cases are autoantibody production and the presence of antibodies against nuclear components, such as DNA, chromatin factors and small nuclear ribonuclear protein particles[92]. There is also evidence for defective T-cell function, which provides the main connection with DNA methylation[93].

T cells are believed to drive the autoantibody response in SLE, and this has been proposed to result from loss of DNA methylation in these cells. The genomes of SLE T cells are globally hypomethylated (15–20% reduction) and DNMT1 levels are reduced[94]. Treatment of normal CD4+ T CELLS with the DNA methylation inhibitor 5-aza-2′-deoxycytidine renders them autoreactive, and ADOPTIVE TRANSFER of these cells causes an SLE-like disease in mice[95]. Hypomethylated T cells are also able to promiscuously kill SYNGENEIC MACROPHAGES; this increased apoptosis, combined with reduced clearance of apoptotic material, is likely to have a role in inducing the anti-DNA antibodies that characterise SLE. In addition, several genes relevant to the SLE phenotype are methylated in normal T cells but are demethylated by both 5-aza-2′-deoxycytidine and in SLE patients[96,97]. An alternative hypothesis for the role of DNA methylation in SLE involves the hypomethylation of endogenous retroviruses (HERV) and the re-expression of viral proteins. Antibodies to HERV-encoded proteins have been detected in SLE patients, and peptides derived from these proteins cause aberrant CD4+ T-cell responses[98].

Exposure to several DNA methylation inhibitors also induces a lupus-like disease in humans, providing further support for a role of aberrant DNA methylation in SLE[93,99]. DNMT1 is upregulated in normal T cells following stimulation by signals transmitted through the extracellular signal-regulated kinase (ERK) pathway, which is impaired in lupus T cells. The drug hydralazine is thought to cause hypomethylation in T cells and an SLE-like disease in humans by inhibiting this pathway. Impaired ERK signalling probably contributes to reduced DNMT1 levels and

RNAI
The process whereby double-stranded RNAs are cleaved into 21-23 nucleotide duplexes termed small interfering RNAs, leading to inhibition of expression of genes that contain a complementary sequence.

DICER
A ribonuclease that processes dsRNAs to ~21 nucleotide siRNAs (for RNAi) or excises microRNAs from their hairpin precursors.

CD4+ T CELL
Also known as a helper T cell. Initiates both antibody production by B cells and stimulates the activation of other immune cells, such as macrophages, after recognizing a portion of a protein antigen on the surface of an antigen presenting cell.

ADOPTIVE TRANSFER
The process of conferring immunity to an individual by transferring cells or serum from another individual that has been immunized with a specific antigen.

SYNGENEIC MACROPHAGES
macrophages (immune cells that engulf foreign particles) that are transferred between genetically identical mice.

hypomethylation of SLE T cells[100]. SLE therefore demonstrates not only that DNA methylation defects can be restricted to only one or a few cell types, but also that DNA methylation is regulated differently, or responds to different stimuli, depending on cell type.

Immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome. ICF syndrome is an extremely rare autosomal recessive disease that is characterized by profound immunodeficiency, and which results from an absence or significant reduction of at least two immunoglobulin isotypes. Other more variable features include impaired cellular immunity and unusual facial features[101,102]. All ICF patients show marked hypomethylation and decondensation of pericentromeric heterochromatin on chromosomes 1 and 16, and, to a lesser extent, chromosome 9 (see online supplementary information S3 (table)). This abnormality is observed almost exclusively in ICF B cells or in LYMPHOBLASTOID CELL LINES that have been stimulated with mitogen[101,103,104].

The molecular defect in 60–70% of ICF patients is mutation of the *DNMT3B* gene at 20q11.2, leading to defective methylation[11,105,106]. Patients without detectable *DNMT3B* mutations might have abnormalities in other regions of the gene that have not been examined (such as the promoter and 5′ and 3′ UTRs), or might arise from mutations in a gene that regulates DNMT3B. Mutations in *DNMT3B* are usually heterozygous and target the catalytic domain. Complete loss-of-function of the encoded enzyme is probably embryonic lethal in humans, as it is in mice[11], and several of the catalytic domain mutations examined so far do not result in complete loss of enzyme activity[105,107]. The clinical variability of the disease might therefore be related to the level of residual DNMT3B activity. It was recently suggested that different methylation defects are present in ICF patients with and without *DNMT3B* mutations, but the molecular basis behind these differences is not known[108].

The instability of repeat-rich pericentromeric heterochromatin in this disease is probably due to its hypomethylation in ICF cells. Loss of DNA methylation in ICF syndrome is region specific and affects mainly the pericentromeric satellite 2 and 3 repeats and non-pericentromeric regions that include the inactive X chromosome, cancer–testis genes, and other sporadic repeats (such as NBL2 and the target in FSHD, D4Z4)[109–111]. This indicates that these regions are bona-fide targets for DNMT3B. Satellite 2 and 3 hypomethylation is common in cancer cells, and although no elevated cancer incidence has been reported in ICF patients, the small number of these patients and their short lifespan make it possible that a slightly increased incidence has gone undetected[112].

The role of *DNMT3B* mutations in the invariant immune defects in ICF syndrome is far less clear. Microarray analyses have revealed that many genes involved in immune function and B-cell immunoglobulin production are deregulated in ICF lymphoblastoid cell lines, indicating a defect in differentiation or activation. However, these experiments did not provide obvious candidate genes to explain the developmental and neurological aspects of ICF syndrome. Interestingly, genes with altered expression in ICF cells are not hypomethylated. This implies either that DNMT3B regulates gene expression independently of its ability to methylate DNA, or that the affected genes show altered expression owing to downstream effects of other genes, such as transcription factors, which might be activated by hypomethylation[113]. The former idea is supported by recent findings that DNMT3B interacts with histone deacetylases (HDACs), histone methylases and chromatin-remodelling enzymes[114,115]. In addition, DNMT3B can profoundly influence neuronal differentiation of multipotent cells in a manner that depends on its ability to interact with HDACs, but not on its ability to methylate DNA[116]. Both methylation-related and unrelated defects might therefore be important in ICF syndrome.

ATRX: a connection with chromatin remodelling

Alpha-thalassemia/mental retardation syndrome, X-linked (ATRX) is a very rare X-linked disease that occurs in <1 in 100,000 males. It is characterized by severe mental retardation and genital abnormalities, with alpha-thalassemia in many, but not all, patients. The affected gene, *ATRX*, is located at Xq13 (see online supplementary information S3 (table)) and encodes an ATP-dependent chromatin-remodelling protein of the chromodomain and helicase-like domains (CHD) subclass of SNF2-LIKE PROTEINS[117].

Functional domains of ATRX include a PHD zinc finger-like motif at its N-terminus (homologous to the PHD motifs in DNMT3A and DNMT3B), and a helicase domain at its C-terminus. Mutations in *ATRX* are clustered in these two domains and are thought to impair its nuclear localization, protein–protein interactions, or chromatin-remodelling functions[118–120]. ATRX localizes to pericentromeric heterochromatin and PROMYELOCYTIC LEUKEMIA (PML) NUCLEAR BODIES in a cell-cycle-dependent manner, and exists in a complex with the transcriptional regulatory death-associated protein 6 (DAXX)[120–122]. It is therefore thought to regulate transcription, although its target genes and the nature of the regulation (positive or negative) are unknown.

Interestingly, ATRX patients also have DNA-methylation defects. These are diverse, and include both aberrant hypermethylation (at the DYZ2 Y-chromosome repeat) and hypomethylation events (at ribosomal DNA repeats), but no overall change in 5-methylcytosine levels[123]. It remains unclear, however, if methylation abnormalities directly contribute to the ATRX phenotype. There have been important advances in recent years in our understanding of the direct and indirect connections between the DNA methylation and chromatin-remodelling machineries, and the identification of DNA methylation defects in ATRX patients has reinforced this link in human cells[124].

LYMPHOBLASTOID CELL LINE
Immortalized B cell line created by infecting primary B cells with Epstein-Barr virus.

SNF2-LIKE PROTEINS
ATP-dependent chromatin remodelling enzymes that contain a region homologous to an extended family of proteins that include known RNA and DNA helicases.

PML BODIES
Dot-like structures in the nucleus of most mammalian cells. These were originally defined by the localization of the PML protein, which is involved in transcriptional regulation.

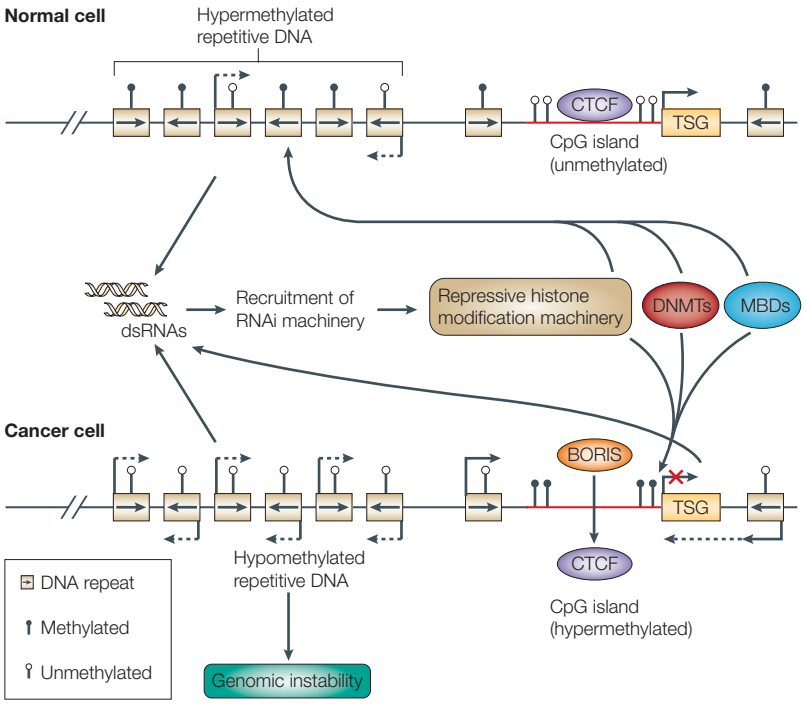

Figure 4 | **A model for the disruption of normal DNA methylation in cancer cells.** In normal cells (top), low-level transcription from hypermethylated repetitive DNA results in the generation of double-stranded RNAs that are cleaved by Dicer (not shown). This results in recruitment of the repressive RNAi machinery and, directly or indirectly, DNA methylation (involving DNA methyltransferases (DNMTs) and methyl-CpG binding proteins (MBDs)). CpG islands are largely hypomethylated and the associated genes (a tumour suppressor gene (TSG) in this case) are transcribed. Binding of CCCTC-binding factor (CTCF) might protect regions from DNA methylation. During the progression of a cell towards a cancerous state (bottom), repetitive DNA tends to lose methylation marks (possibly owing to dysregulation of DNMTs or the RNAi system) resulting in increased genomic instability. Repeat hypomethylation near genes or promoters could disrupt transcription of TSGs by several mechanisms, including transcriptional interference (not shown) or generation of antisense RNA and recruitment of the RNAi machinery. In addition, aberrant re-expression of BORIS (possibly also due to genomic hypomethylation) might displace CTCF from its binding sites and disrupt the methylation boundary functions of CTCF, allowing DNMTs access to CpG islands.

Future perspectives: towards a unified model

A fundamental question in the DNA-methylation field (and perhaps epigenetics as a whole) is whether a common mechanism underlies diseases that arise from DNA methylation defects. If so, can this mechanism also explain how DNA methylation patterns are established during normal development and accurately maintained in somatic cells? Exciting recent findings from the imprinting and RNAi fields have provided glimpses of such putative mechanisms. For example, there is growing evidence that CTCF not only reads DNA-methylation marks and ensures allele-specific gene expression, but also has a role in determining DNA methylation patterns[30,33,125] (BOX 1). If the default state of the genome is full methylation, CTCF binding could demarcate methylation-free zones, making it a sort of DNA methylation insulator, in addition to its other known properties.

A recent twist in the CTCF story is the discovery of a paralogous gene, *BORIS* (also called CCCTC-binding factor-like, *CTCFL*), which is highly homologous to CTCF within the zinc finger region (BOX 1)[30,126]. The exact role of BORIS remains unclear, but it is tempting to speculate that it might antagonize the function of CTCF and disrupt methylation boundaries, because the expression of the two seems to be mutually exclusive in normal cells, although BORIS is upregulated in cancer cells[126]. Competition between CTCF and BORIS for specific binding sites might explain the aberrant hypermethylation events that occur in cancer (FIG. 4) and the LOI defects in imprinting disorders (although, for LOI, aberrant BORIS expression might be only transiently dysregulated during a crucial developmental period). A breakdown of the barrier between unmethylated and methylated regions could also result in hypomethylation events by 'diluting out' the repression machinery and causing it to move away from heterochromatin. Given the functionally haploid nature of imprinted genes, they might be very sensitive to the levels of these two proteins.

The propensity of a particular region to undergo aberrant DNA methylation in disease could be governed by a combination of the CpG-richness of the CTCF binding sites (determining the degree to which CTCF binding can be inhibited by DNA methylation) and the affinity of BORIS for the site that is normally occupied by CTCF. CTCF would only need to be transiently displaced from its binding site, or sites, to allow access to DNMTs and subsequent methylation, which would then prevent CTCF from re-binding. Such ideas are experimentally testable, and important progress in this area is likely to emerge in the next few years.

Another interesting model relates to the RNAi pathway[13]. Recent evidence, primarily from fission yeast, indicates that the RNAi machinery (the RNA-induced initiation of transcriptional gene silencing (RITS) complex) directs elements of the repressive histone modification apparatus to regions destined for silencing[127]. This machinery is guided to heterochromatin through the transcription of repetitive DNA elements, which, owing to their heterogeneous nature, can be transcribed from either DNA strand and consequently give rise to double-stranded RNA (dsRNA)[128–130]. Much of the RNAi pathway is conserved in mammalian cells, and recent data indicate that dsRNA can silence genes through *de novo* methylation and that centromeric and pericentromeric heterochromatic repeats are transcribed in mammals[131–133]. Therefore, low-level transcription from repetitive elements might be the means for targeting repressive histone modifications and DNA methylation throughout the genome.

In cancer, the low-level hypomethylation of repetitive DNA near promoter regions early in transformation could generate antisense transcripts that result in the formation of dsRNA homologous to the gene and/or its promoter (FIG. 4). Recruitment of the RNAi machinery would lead to the establishment of repressive epigenetic marks, including DNA methylation, at these sites. A recent report that describes a human disease arising from a gene rearrangement that results in the generation of an aberrant antisense transcript to one of the α-globin genes illustrates how aberrant dsRNA production might lead to DNA methylation defects. The gene, although fully intact, was silenced by DNA

methylation at its CpG-island promoter[134]. Indeed, studies have shown that antisense transcription is increased overall in tumour cells, and this could be due to DNA hypomethylation and spurious transcription from repeats[135].

Similar mechanisms could be used as part of normal regulation of allele-specific transcription in imprinted regions, as antisense transcripts of differentially expressed genes are a common feature of imprinted regions[136]. Diseases associated with repeat instability might potentially result from an affinity of the RNAi machinery for the aberrantly expanded TNR-containing mRNAs that form hairpins. Conversely, repeat contraction could reduce the efficiency of targeting of the RNAi machinery, for example, resulting in the hypomethylation of D4Z4 observed in FSHD (assuming D4Z4 is transcribed at some point).

These models for regulating DNA methylation patterns — involving CTCF/BORIS activity or RNAi — are not mutually exclusive. The former can be viewed as a system for protecting regions from methylation, whereas the latter represents a means for directing it to specific loci. One mechanism might predominate in certain regions of the genome, or in certain cell types and developmental stages. In the next few years, intensive research efforts will be devoted to studying these and other systems involved in epigenetic gene regulation. These initiatives will undoubtedly advance our understanding of gene regulation, epigenetic reprogramming and the epigenetic aetiology of a wide range of diseases with a known or suspected epigenetic component. It should also fuel the development of new therapies aimed at reversing aberrant epigenetic alterations in disease and development.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes. Dev.* **16**, 6–21 (2002).
An excellent review of DNA methylation in mammalian cells.
2. Fischle, W., Wang, Y. & Allis, C. D. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
3. Robertson, K. D. DNA methylation and chromatin — unraveling the tangled web. *Oncogene* **21**, 5361–5379 (2002).
4. Hendrich, B. & Tweedie, S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **19**, 269–277 (2003).
5. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
This paper demonstrates that active demethylation occurs during embryonic development in the absence of cell division.
6. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
7. Bird, A. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).
8. Song, F. *et al.* Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc. Natl Acad. Sci. USA* **102**, 3336–3341 (2005).
9. Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes. Dev.* **10**, 1991–2002 (1996).
10. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
11. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
References 10 and 11 describe in detail the detrimental effects that loss of DNA methylation has on mouse embryonic development.
12. Brown, R. & Strathdee, G. Epigenomics and epigenetic therapy of cancer. *Trends Mol. Med.* **8** (Suppl.), S43–S48 (2002).
13. Novina, C. D. & Sharp, P. A. The RNAi revolution. *Nature* **430**, 161–164 (2004).
14. Shahbazian, M. D. & Zoghbi, H. Y. Molecular genetics of Rett syndrome and clinical spectrum of MECP2 mutations. *Curr. Opin. Neurol.* **14**, 171–176 (2001).
15. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature Rev. Cancer* **4**, 1–11 (2004).
16. Mertens, F., Johansson, B., Hoglund, M. & Mitelman, F. Chromosomal imbalance maps of malignant solid tumors: A cytogenetic survey of 3185 neoplasms. *Cancer Res.* **57**, 2765–2780 (1997).
17. Robertson, K. D. & Wolffe, A. P. DNA methylation in health and disease. *Nature Rev. Genet.* **1**, 11–19 (2000).
18. Widschwendter, M. *et al.* DNA hypomethylation and ovarian cancer biology. *Cancer Res.* **64**, 4472–4480 (2004).
19. De Smet, C., Lurquin, C., Lethe, B., Martelange, V. & Boon, T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol. Cell. Biol.* **19**, 7327–7335 (1999).
20. Nakamura, N. & Takenaga, K. Hypomethylation of the metastasis-associated *S100A4* gene correlates with gene activation in human colon adenocarcinoma cell lines. *Clin. Exp. Metastasis* **16**, 471–479 (1998).
21. Akiyama, Y., Maesawa, C., Ogasawara, S., Terashima, M. & Masuda, T. Cell-type specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells. *Am. J. Clin. Pathol.* **163**, 1911–1919 (2003).
22. Gupta, A., Godwin, A. K., Vanderveer, L., Lu, A. & Liu, J. Hypomethylation of the synuclein γ gene CpG island promotes its aberrant expression in breast and ovarian carcinoma. *Cancer Res.* **63**, 664–673 (2003).
23. Strichman-Almashanu, L. Z. *et al.* A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Gen. Res.* **12**, 543–554 (2002).
24. Issa, J.-P. CpG island methylator phenotype in cancer. *Nature Rev. Cancer* **4**, 988–993 (2004).
25. Chan, A. O.-O. *et al.* CpG island methylation in aberrant crypt foci of the colorectum. *Am. J. Pathol.* **160**, 1823–1830 (2002).
26. Costello, J. F. *et al.* Aberrant CpG-island methylation has a non-random and tumor-type-specific patterns. *Nature Genet.* **25**, 132–138 (2000).
This paper represents one of the most comprehensive genome-wide scans for aberrant methylation in cancer and demonstrates just how widespread the defects are.
27. Laird, P. W. The power and the promise of DNA methylation markers. *Nature Rev. Cancer* **3**, 253–266 (2003).
28. Feinberg, A. P., Cui, H. & Ohlsson, R. DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Sem. Canc. Biol.* **12**, 389–398 (2002).
29. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature Rev. Genet.* **2**, 21–32 (2001).
30. Klenova, E. M., Morse, H. C., Ohlsson, R. & Lobanenkov, V. V. The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Sem. Canc. Biol.* **449**, 1–16 (2002).
31. Kanduri, C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* **10**, 853–856 (2000).
This paper shows that CTCF binding is methylation sensitive and provides important mechanistic insights into how cells distinguish between maternal and paternal alleles.
32. Thakur, N. *et al.* An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol. Cell. Biol.* **24**, 7855–7862 (2004).
33. Schoenherr, C. J., Levorse, J. M. & Tilghman, S. M. CTCF maintains differential methylation at the *Igf2/H19* locus. *Nature Genet.* **33**, 66–69 (2003).
34. Khosla, S., Dean, W., Brown, D., Reik, W. & Feil, R. Culture of preimplantation mouse embryos affects fetal development and the expression of imprinted genes. *Biol. Reprod.* **64**, 918–926 (2001).
35. Behr, B. & Wang, H. Effects of culture conditions on IVF outcome. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **115** (Suppl. 1), S72–S76 (2004).
36. Dean, W. *et al.* Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc. Natl Acad. Sci. USA* **98**, 13734–13738 (2001).
37. De Baun, M. R., Niemitz, E. L. & Feinberg, A. P. Association of *in vitro* fertilization with Beckwith–Wiedemann syndrome and epigenetic alterations of *LIT1* and *H19*. *Am. J. Hum. Genet.* **72**, 156–160 (2003).
References 34–37 show that *in vitro* culturing of embryos can lead to epigenetic defects in animals and that this might also have a role in humans.
38. Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E. & Tycko, B. Tumour-suppressor activity of H19 RNA. *Nature* **365**, 764–767 (1993).
39. Moulton, T. *et al.* Epigenetic lesions at the H19 locus in Wilms' tumor patients. *Nature Genet.* **7**, 440–447 (1994).
40. Steenman, M. J. C. *et al.* Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumor. *Nature Genet.* **7**, 433–439 (1994).
41. Sakatani, T. *et al.* Loss of imprinting of *Igf2* alters intestinal maturation and tumorigenesis in mice. *Science* **307**, 1976–1978 (2005).
42. Cui, H. *et al.* Loss of imprinting of *insulin-like growth factor-II* in Wilms' tumor commonly involves altered methylation but not mutations of CTCF or its binding site. *Cancer Res.* **61**, 4947–4950 (2001).
43. Nakagawa, H. *et al.* Loss of imprinting of the *insulin-like growth factor II* gene occurs by biallelic methylation in a core region of *H19*-associated CTCF-binding sites in colorectal cancer. *Proc. Natl Acad. Sci. USA* **98**, 591–596 (2001).
44. Cui, H. *et al.* Loss of imprinting in colorectal cancer linked to hypomethylation of *H19* and *IGF2*. *Cancer Res.* **62**, 6442–6446 (2002).
45. Thompson, J. S., Reese, K. J., DeBaun, M. R., Perlman, E. J. & Feinberg, A. P. Reduced expression of the cyclin-dependent kinase inhibitor gene *p57KIP2* in Wilms' tumor. *Cancer Res.* **56**, 5723–5727 (1996).
46. Yu, Y. *et al. NOEY2* (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proc. Natl Acad. Sci. USA* **96**, 214–219 (1999).
47. Pedersen, I. S. *et al.* Frequent loss of imprinting of *PEG1/MEST* in invasive breast cancer. *Cancer Res.* **59**, 5449–5451 (1999).
48. Walter, J. & Paulsen, M. Imprinting and disease. *Semin. Cell Dev. Biol.* **14**, 101–110 (2003).
49. Lee, M. P. *et al.* Loss of imprinting of a paternally expressed transcript, with antisense orientation to K_VLQT1, occurs frequently in Beckwith–Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc. Natl Acad. Sci. USA* **96**, 5203–5208 (1999).
50. Kanduri, C. *et al.* A differentially methylated imprinting control region within the *Kcnq1* locus harbors a methylation-sensitive chromatin insulator. *J. Biol. Chem.* **277**, 18106–18110 (2002).
51. Butler, M. G. Imprinting disorders: non-Mendelian mechanisms affecting growth. *J. Pediatr. Endocrinol. Metab.* **15**, 1279–1288 (2002).

52. Weksberg, R. *et al.* Tumor development in the Beckwith–Wiedemann syndrome is associated with a variety of constitutional molecular 11p15 alterations including imprinting defects of *KCNQ1OT1*. *Hum. Mol. Genet.* **10**, 2989–3000 (2001).
53. Diaz-Meyer, N. *et al.* Silencing of *CDKN1C* (*p57^{KIP2}*) is associated with hypomethylation at KvDMR1 in Beckwith–Wiedemann syndrome. *J. Med. Genet.* **40**, 797–801 (2005).
54. Goldstone, A. P. Prader–Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol. Metab.* **15**, 12–20 (2004).
55. Nicholls, R. D. & Knepper, J. L. Genome organization, function, and imprinting in Prader–Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.* **2**, 153–175 (2001).
56. Sutcliffe, J. S. *et al.* Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region. *Nature Genet.* **8**, 52–58 (1994).
57. Runte, M. *et al.* Comprehensive methylation analysis in typical and atypical PWS and AS patients with normal biparental chromosomes 15. *Eur. J. Hum. Genet.* **9**, 519–526 (2001).
58. Runte, M. *et al.* The IC-*SNURF–SNRPN* transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for *UBE3A*. *Hum. Mol. Genet.* **10**, 2687–2700 (2001).
59. Gerard, G., Hernandez, L., Wevrick, R. & Stewart, C. L. Disruption of the mouse *necdin* gene results in early post-natal lethality. *Nature Genet.* **23**, 199–202 (1999).
60. Tsai, T. F., Jiang, Y. H., Bressler, J., Armstrong, D. & Beaudet, A. L. Paternal deletion from *Snrpn* to *Ube3a* in the mouse causes hypotonia, growth retardation and partial lethality and provides evidence for a gene contributing to Prader–Willi syndrome. *Nature Genet.* **8**, 1357–1364 (1999).
61. Saitoh, S. & Wada, T. Parent-of-origin specific histone acetylation and reactivation of a key imprinted gene locus in Prader–Willi syndrome. *Am. J. Hum. Genet.* **66**, 1958–1962 (2000).
62. Lossie, A. C. *et al.* Distinct phenotypes distinguish the molecular classes of Angelman syndrome. *J. Med. Genet.* **38**, 834–845 (2001).
63. Buiting, K., Lich, C., Cottrell, S., Barnicoat, A. & Horsthemke, B. A 5-kb imprinting center deletion in a family with Angelman syndrome reduces the shortest region of deletion overlap to 880 bp. *Hum. Genet.* **105**, 665–666 (1999).
64. Lalande, M. Imprints of disease at *GNAS1*. *J. Clin. Inves.* **107**, 793–794 (2001).
65. Liu, J., Nealon, J. G. & Weinstein, L. S. Distinct patterns of abnormal *GNAS* imprinting in familial and sporadic pseudohypoparathyroidism type IB. *Hum. Mol. Genet.* **14**, 95–102 (2005).
66. Hayward, B. E. *et al.* Imprinting of the $G_s\alpha$ gene *GNAS1* in the pathology of acromegaly. *J. Clin. Inves.* **107**, R31–R36 (2001).
67. Temple, I. K. & Shield, J. P. H. Transient neonatal diabetes, a disorder of imprinting. *J. Med. Genet.* **39**, 872–875 (2002).
68. Arima, T. *et al.* A conserved imprinting control region at the *HYMAI/ZAC* domain is implicated in transient neonatal diabetes mellitus. *Hum. Mol. Genet.* **10**, 1475–1483 (2001).
69. Morison, I. M. & Reeve, A. E. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum. Mol. Genet.* **7**, 1599–1609 (1998).
An excellent review detailing all known disorders in humans and animals that have a parent-of-origin bias.
70. Lewis, A. & Murrell, A. Genomic imprinting: CTCF protects the boundaries. *Curr. Biol.* **14**, R284–R286 (2004).
71. Everett, C. M. & Wood, N. W. Trinucleotide repeats and neurodegenerative disease. *Brain* **127**, 2385–2405 (2004).
72. Cleary, J. D. & Pearson, C. E. The contribution of *cis*-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.* **100**, 25–55 (2003).
73. Crawford, D. C., Acuna, J. M. & Sherman, S. L. *FMR1* and the fragile X syndrome: human genome epidemiology review. *Genet. Med.* **3**, 359–371 (2001).
74. Sutherland, G. R. Rare fragile sites. *Cytogenet. Genome Res.* **100**, 77–84 (2003).
75. Jin, P., Alisch, R. S. & Warren, S. T. RNA and microRNAs in fragile X mental retardation. *Nature Cell Biol.* **6**, 1048–1053 (2004).
76. Hagerman, R. J. *et al.* Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**, 127–130 (2001).
77. Oberle, I. *et al.* Instability of a 550-base pair segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102 (1991).
78. Coffee, B., Zhang, F., Ceman, S., Warren, S. T. & Reines, D. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile X syndrome. *Am. J. Hum. Genet.* **71**, 923–932 (2002).
79. Sarafidou, T. *et al.* Folate-sensitive fragile site *FRA10A* is due to an expansion of a CGG repeat in a novel gene, *FRA10AC1*, encoding a nuclear protein. *Genomics* **84**, 69–81 (2004).
80. van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nature Genet.* **34**, 315–317 (2003).
81. Rijkers, T. *et al. FRG2*, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J. Med. Genet.* **41**, 826–836 (2004).
82. Gabellini, D., Green, M. R. & Tupler, R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**, 339–348 (2002).
83. Jiang, G. *et al.* Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* **12**, 2909–2921 (2003).
84. Steinbach, P., Glaser, D., Vogel, W., Wolf, M. & Schwemmle, S. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am. J. Hum. Genet.* **62**, 278–285 (1998).
85. Filippova, G. N. *et al.* CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the *DM1* locus. *Nature Genet.* **28**, 335–343 (2001).
86. Saveliev, A., Everett, C., Sharpe, T., Webster, Z. & Festenstein, R. DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature* **422**, 909–913 (2003).
87. Edamura, K. N., Leonard, M. R. & Pearson, C. E. Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet.* **76**, 302–311 (2005).
88. Gorbunova, V., Seluanov, A., Mittelman, D. & Wilson, J. H. Genome-wide demethylation destabilizes CTG•CAG trinucleotide repeats in mammalian cells. *Hum. Mol. Genet.* **13**, 2979–2989 (2004).
89. Reik, W., Maher, E. R., Morrison, P. J., Narding, A. E. & Simpson, S. A. Age of onset in Huntington's disease and methylation at D4S95. *J. Med. Genet.* **30**, 185–188 (1993).
90. Smith, S. S., Laayoun, A., Lingeman, R. G., Baker, D. J. & Riley, J. Hypermethylation of telomere-like foldbacks at codon 12 of the human c-Ha-*ras* gene and the trinucleotide repeat of the *FMR-1* gene of fragile X. *J. Mol. Biol.* **243**, 143–151 (1994).
91. Handa, V., Saha, T. & Usdin, K. The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by dicer. *Nucleic Acids Res.* **31**, 6243–6248 (2003).
92. Klippel, J. H. Systemic lupus erythematosus: demographics, prognosis, and outcome. *J. Rheumatol.* **24**, 67–71 (1997).
93. Richardson, B. DNA methylation and autoimmune disease. *Clin. Immunol.* **109**, 72–79 (2003).
94. Richardson, B. *et al.* Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum.* **33**, 1665–1673 (1990).
95. Quddus, J. *et al.* Treating activated $CD4^+$ T cells with either of two distinct DNA methyltransferase inhibitors, 5-azacytidine or procainamide, is sufficient to cause a lupus-like disease in syngeneic mice. *J. Clin. Inves.* **92**, 38–53 (1993).
96. Lu, Q. *et al.* Demethylation of *ITGAL* (CD11a) regulatory sequences in systemic lupus erythematosus. *Arthritis Rheum.* **46**, 1282–1291 (2003).
97. Oelke, K. *et al.* Overexpression of CD70 and overstimulation of IgG synthesis by lupus T cells and T cells treated with DNA methylation inhibitors. *Arthritis Rheum.* **50**, 1850–1860 (2004).
98. Sekigawa, I. *et al.* DNA methylation in systemic erythematosus. *Lupus* **12**, 79–85 (2003).
99. Scheinbart, L. S., Johnson, M. A., Gross, L. A., Edelstein, S. R. & Richardson, B. C. Procainamide inhibits DNA methyltransferase in a human T cell line. *J. Rheum.* **18**, 530–540 (1991).
100. Deng, C. *et al.* Hydralazine may induce autoimmunity by inhibiting extracellular signal-regulated kinase pathway signalling. *Arthritis Rheum.* **48**, 746–756 (2003).
101. Smeets, D. F. *et al.* ICF syndrome: a new case and review of the literature. *Hum. Genet.* **94**, 240–246 (1994).
102. Ehrlich, M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clin. Immunol.* **109**, 17–28 (2003).
103. Franceschini, P. *et al.* Variability of clinical and immunological phenotype in immunodeficiency-centromeric instability-facial anomalies syndrome. *Eur. J. Pediatr.* **154**, 840–846 (1995).
104. Tuck-Muller, C. M. *et al.* DNA hypomethylation and unusual chromosome instability in cell lines from ICF patients. *Cytogenet. Cell Genet.* **89**, 121–128 (2000).
105. Xu, G.-L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
106. Hansen, R. S. *et al.* The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA* **96**, 14412–14417 (1999).
107. Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* **277**, 20409–20414 (2002).
108. Jiang, Y. L. *et al.* DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum. Mut.* **25**, 56–63 (2005).
109. Kondo, T. *et al.* Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats *D4Z4* and *NBL2*. *Hum. Mol. Genet.* **9**, 597–604 (2000).
110. Tao, Q. *et al.* Defective *de novo* methylation of viral and cellular DNA sequences in ICF syndrome cells. *Hum. Mol. Genet.* **11**, 2091–2102 (2002).
111. Hansen, R. S. *et al.* Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum. Mol. Genet.* **9**, 2575–2587 (2000).
112. Ehrlich, M. DNA hypomethylation, cancer, the immunodeficiency, centromeric region instability, facial anomalies syndrome and chromosomal rearrangements. *J. Nutr.* **132**, S2424–S2429 (2002).
113. Ehrlich, M. *et al.* DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphomagenesis genes. *Hum. Mol. Genet.* **10**, 2917–2931 (2001).
114. Geiman, T. M. *et al.* DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem. Biophys. Res. Commun.* **318**, 544–555 (2004).
115. Geiman, T. M. *et al.* Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.* **32**, 2716–2729 (2004).
116. Bai, S. *et al.* DNA methyltransferase 3b regulates nerve growth factor-induced differentiation of PC12 cells by recruiting histone deacetylase 2. *Mol. Cell. Biol.* **25**, 751–766 (2005).
117. Havas, K., Whitehouse, I. & Owen-Hughes, T. ATP-dependent chromatin remodeling activities. *Cell. Mol. Life Sci.* **58**, 673–682 (2001).
118. Gibbons, R. J. *et al.* Mutations in transcriptional regulator *ATRX* establish the functional significance of a PHD-like domain. *Nature Genet.* **17**, 146–148 (1997).
119. Cardoso, C. *et al. ATR-X* mutations cause impaired nuclear location and altered DNA binding properties of the XNP/ATR-X protein. *J. Med. Genet.* **37**, 746–751 (2000).
120. Tang, J. *et al.* A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *J. Biol. Chem.* **279**, 20369–20377 (2004).
121. Ishov, A. M., Vlafimirova, O. V. & Maul, G. G. Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. *J. Cell. Sci.* **117**, 3807–3820 (2004).
122. Xue, Y. *et al.* The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc. Natl Acad. Sci. USA* **100**, 10635–10640 (2003).
123. Gibbons, R. J. *et al.* Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the patterns of DNA methylation. *Nature Genet.* **24**, 368–371 (2000).
124. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.* **3**, 415–428 (2002).
125. Filippova, G. N. *et al.* Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev. Cell* **8**, 31–42 (2005).

126. Loukinov, D. I. *et al.* BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl Acad. Sci. USA* **99**, 6806–6811 (2002). **The first in-depth characterization of the CTCF paralogue BORIS and a demonstration of the mutually exclusive expression of these two proteins in normal tissues.**
127. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
128. Noma, K.-i. *et al.* RITS acts in *cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genet.* **36**, 1174–1180 (2004).
129. Schramke, V. & Allshire, R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301**, 1069–1073 (2003).
130. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004). **References 127–130 provided the first direct, mechanistic evidence for how the RNAi machinery can target transcriptional repression back to heterochromatic repeat regions of the genome.**
131. Kawasaki, H. & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**, 211–217 (2004).
132. Rudert, F., Bronner, S., Garnier, J. M. & Dolle, P. Transcripts from opposite strands of γ satellite DNA are differentially expressed during mouse development. *Mamm. Genome* **6**, 76–83 (1995).
133. Lehnertz, B. *et al. Suv39h*-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
134. Tufarelli, C. *et al.* Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature Genet.* **34**, 157–165 (2003).
135. Shendure, J. & Church, G. M. Computational discovery of sense–antisense transcription in the human and mouse genomes. *Genome Biol.* **3**, 1–14 (2003).
136. Lavorgna, G. *et al.* In seach of antisense. *Trends Biochem. Sci.* **29**, 88–94 (2004).
137. Mukhopadhyay, R. *et al.* The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Gen. Res.* **14**, 1594–1602 (2004).
138. Fedoriw, A. M., Stein, P., Svoboda, P., Schultz, R. M. & Bartolomei, M. S. Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting. *Science* **303**, 238–240 (2004).
139. Rand, E., Ben-Porath, I., Keshet, I. & Cedar, H. CTCF elements direct allele-specific undermethylation at the imprinted H19 locus. *Curr. Biol.* **14**, 1007–1012 (2004).
140. Rasko, J. E. J. *et al.* Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res.* **61**, 6002–6007 (2001).
141. Filippova, G. N. *et al.* Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res.* **62**, 48–52 (2002).
142. Wright, V. C., Schieve, L. A., Reynolds, M. A., Jeng, G. & Kissin, D. Assisted reproductive technology surveillance — United States, 2001. *MMWR Surveill Summ.* **53**, 1–20 (2004).
143. Cox, G. F. *et al.* Intracytoplasmic sperm injection may increase the risk of imprinting defects. *Am. J. Hum. Genet.* **71**, 162–164 (2002).
144. Gicquel, C. *et al. In vitro* fertilization may increase the risk of Beckwith–Wiedemann syndrome related to the abnormal imprinting of the *KCN1OT* gene. *Am. J. Hum. Genet.* **72**, 1338–1341 (2003).
145. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).
146. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
147. De Chiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849–859 (1991).
148. Barker, D. J. *et al.* Fetal nutrition and cardiovascular disease in adult life. *Lancet* **341**, 938–941 (1993).
149. Mann, M. R. W. *et al.* Selective loss of imprinting in the placenta following preimplantation development in culture. *Development* **131**, 3727–3735 (2004).

Acknowledgments

I apologize to those whose work could not be cited owing to space limitations. Work in the author's lab is supported by the National Institutes of Health.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ATP10C | *ATRX* | *BORIS* | *CDKN1C* | CTCF | DAXX | *DIRAS3* | *Dnmt1* | *Dnmt3a* | *Dnmt3b* | *FMR1* | FRAXE | FRAXF | FRA10A | FRA11B | FRA16A | *FRG1* | *FRG2* | *GNAS* | *HYMAI* | IGF2 | KCNQ1 | KCNQ1OT1 | *MEST* | *PLAGL1* | *S100A4* | *SERPINB5* | *SNCG* | *SNRPN* | *SNURF* | *UBE3A*
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Albright hereditary osteodystrophy | alpha-thalassemia/mental retardation syndrome, X-linked | Angelman syndrome | Beckwith–Wiedemann syndrome | facioscapulohumeral muscular dystrophy | fragile X syndrome | immunodeficiency, centromeric instability and facial anomalies syndrome | Prader–Willi syndrome | pseudohypoparathyroidism Ib | Rett syndrome | transient neonatal diabetes mellitus

FURTHER INFORMATION

The DNA methylation database: http://www.methdb.de
The DNA methylation society: http://www.dnamethsoc.com
Human Epigenome Project: http://www.epigenome.org
The M. D. Anderson Cancer Center DNA Methylation in Cancer web site: http://www.mdanderson.org/departments/methylation
The Genomic Imprinting Website: http://www.geneimprint.com/index.html
MRC Mammalian Genetics Unit Mouse Imprinting web site: http://www.mgu.har.mrc.ac.uk/research/imprinting
CITE: Candidate Imprinted Transcript from gene Expression database: http://fantom2.gsc.riken.go.jp/imprinting
The DNMT3Bbase mutation registry for ICF syndrome: http://bioinf.uta.fi/DNMT3Bbase

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (table) | S3 (table)
Access to this links box is available online.