Context matters: DNA methylation in human disease

Vukić, M.


CDCA7

CHAPTER 1

General introduction

Adapted from

DNA methylation in disease: Immunodeficiency, Centromeric instability, Facial anomalies syndrome

Maja Vukić[1] and Lucia Daxinger[1]

[1] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

Essays Biochem. 2019 Dec 20;63(6):773-783.

EPIGENETICS

Every living cell, at its essence, has a 'simple', yet elegant DNA double helix. Since 1944, DNA is known as the carrier of genetic information [1]. However, DNA sequence by itself cannot explain an exquisite phenomenon - a single fertilized egg can develop into a human body, consisting of more than 400 different cell types. While every cell in the body carries the same DNA sequence, each cell type expresses a unique set of genes and thus develops a particular phenotype. So how does a cell know, which genes should be expressed and when?

In the nucleus, DNA is coated with approximately twice as much protein as DNA, known as histones. Together, DNA and histone proteins form a structure called chromatin. There are two different forms of chromatin – euchromatin and heterochromatin. Euchromatin is the more open form, allowing the transcriptional machinery access to DNA, commonly referred to as an 'active' state. Heterochromatin, on the other hand, is organized into a dense structure, keeping DNA inaccessible, thereby 'compressing' the transcriptionally inactive part of the genome. Different chemical modifications or 'tags' can be placed directly onto the DNA or its associated histone proteins. These 'tags', known as epigenetic marks, can influence accessibility of different protein machineries to chromatin and DNA [2].

In the early 1940s, already before the discovery of the double helix, the British developmental biologist Conrad Hal Waddington coined the term epigenetics (prefix epi- in Greek—over, on), and referred to it as 'the processes involved in the mechanism by which the genes of the genotype bring about phenotypic effects' [3]. During the past eight decades, the term has evolved into the current working definition of epigenetics: 'the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence' [4]. In other words, a layer of epigenetic information, placed on top of the DNA, allows for a fine-tuned control of how genetic information, in a spatiotemporal manner is translated to gene expression patterns and the establishment and maintenance of cell identity.

DNA METHYLATION AS A REGULATORY EPIGENETIC MARK

One of the most studied and best characterized epigenetic marks is DNA methylation, which generally refers to the direct chemical modification of DNA. In living cells, DNA methylation was first described more than 80 years ago [5,6], almost simultaneously with DNA being identified as the genetic material [1]. In the late 1970s, hypotheses on a role for DNA methylation in the regulation of gene expression were proposed for the first time [7,8]. Today its important function in diverse biological processes in mammalian cells, such as the silencing of germline genes [9] and transposable elements [10], genomic imprinting [11] and X-chromosome inactivation [12,13], is well established. DNA methylation is critical for normal development [14]. Not surprisingly, aberrant DNA methylation is a feature of different types of human diseases including cancer [15], as well as developmental disorders such as Sotos and Kabuki syndrome [16]. Furthermore, genetic defects in several components of the DNA methylation machinery have been linked to human congenital disease such as Tatton-Brown-Rahman syndrome and Immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome [17,18].

In mammals, DNA methylation predominantly occurs on cytosine bases in the symmetrical CpG dinucleotide context, which enables the post-replicative maintenance of DNA methylation patterns. Specifically, the fifth carbon of the cytosine pyrimidine

ring is covalently modified by the addition of a methyl group [17]. DNA methylation is widespread in mammals, and, in somatic cells, approximately 70–80% of all CpGs are methylated [19]. Depending on the genomic context in which it is placed, DNA methylation can be interpreted differently and fulfil diverse regulatory roles (**Figure 1A**). For example, transposable elements, which are dispersed throughout the genome and account for approximately 40% of the mammalian genome [20], are highly methylated sequences [21], with about 95% of them becoming methylated (**Figure 1B**). In somatic cells, DNA methylation is required to suppress transposon expression and mobility [21], thereby protecting the genome from deleterious effects of these elements which when mobilized, can cause genome instability [22]. In contrast, CpG islands (CGI), which are CpG-rich sequences often found in the vicinity of gene promoters, are usually devoid of DNA methylation [23] (Figure 1). Promoter DNA methylation correlates with gene silencing and typically occurs in a tissue- or developmental stage-specific manner [9,24,25]. Historically, DNA methylation has been associated with gene repression, but it is now known that DNA methylation can be associated with active transcription, when found across gene bodies [18]. Gene body methylation has been suggested to prevent transcription from intragenic/cryptic promoters [26], to play a role in the regulation of splicing [27-29] and to function in transcript elongation [30].

THE DNA METHYLATION MACHINERY

DNA methylation patterns are established by the *de novo* methyltransferases DNA methyltransferase 3 alpha (DNMT3A) and DNA methyltransferase 3 beta (DNMT3B) during early embryonic and germline development (**Figure 2**) [31,32]. DNMT3s are generally known as enzymes that lack DNA sequence specificity. However, recent *in vitro* studies suggest that their catalytic activity can be influenced by nucleotides flanking CpG sites [33,34]. Furthermore, these enzymes can be targeted to or excluded from selected genomic regions by different mechanisms including post-translational modifications and via interaction with instructive partner proteins [17]. For example, DNMT3A can form a complex with its catalytically inactive cofactor, DNA methyltransferase 3 like (DNMT3L), which stimulates DNMT3A activity during germline development [18,35-38], and in mouse embryonic stem cells (mESCs) [39]. At their N-terminus, the DNMT3 enzymes contain a Pro-Trp-Trp-Pro (PWWP) domain, which can specifically recognize histone H3 molecules trimethylated at lysine 36, a histone modification enriched at actively transcribed gene bodies [40-42]. Another important domain for chromatin interaction is the ATRX-DNMT3-DNMT3L (ADD) domain, which enables DNMT3s to specifically recognize unmodified histone H3 lysine 4 [43-45], while methylation of histone H3 lysine 4 is known to inhibit this recognition [43,45,46]. Consequently, histone H3 lysine 4 tri-methylated (H3K4me3), a mark associated with active transcription, is thought to protect promoter CpG dinucleotides from gaining a methyl group and thereby preventing gene repression [44].

Once established, DNA methylation patterns have to be maintained after every mitotic cell division. Cells can achieve this through at least two different DNA methylation maintenance pathways. One pathway relies on proteins that are present on the replication fork, which is referred to as DNA replication coupled maintenance, and takes place in early S-phase [47,48]. The second pathway, which is termed DNA replication uncoupled maintenance, occurs in late S-phase independently of the replication fork, and is thought to ensure DNA methylation maintenance at CpGs found in heterochromatic regions [48]. The maintenance DNA methyltransferase 1 (DNMT1) is the only known enzyme responsible for the conversion of hemi-methylated DNA in newly generated daughter cells back

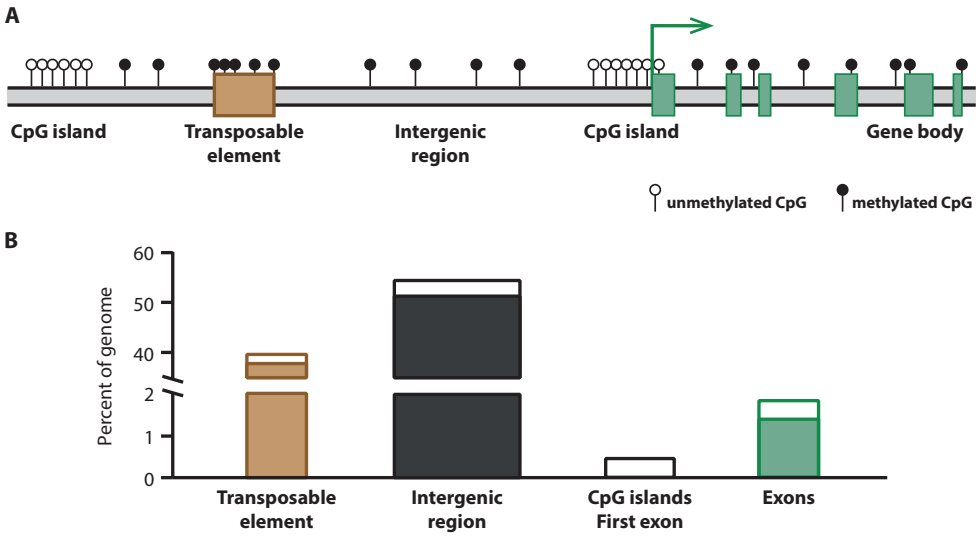

Figure 1. DNA methylation landscape. A Illustration of DNA methylation distribution across different genomic annotations, such as transposable elements, intergenic region, CGI and gene bodies in mammalian genome (adapted from: https://commons.wikimedia.org/wiki/File:DNAme_landscape.png). **B** Bar plot representing the percentage of indicated genomic annotations in the mammalian genome. The coloured part of each bar represents the methylated proportion of the different genomic annotations (adapted from [23]).

to symmetrically methylated molecules (**Figure 2**) [44]. Recruitment of DNMT1 to hemi-methylated sites is facilitated by ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1) [49,50]. UHRF1 is a multi-domain E3 ubiquitin ligase which contains a ubiquitin-like (UBL) domain, a tandem Tudor domain (TTD), a plant homeodomain (PHD) finger, a SET and RING-associated (SRA) domain, and a really interesting new gene (RING) domain [51]. Through its SRA domain, UHRF1 can recognise hemi-methylated DNA [52,53] and utilise its RING domain to ubiquitinate proliferating cell nuclear antigen (PCNA)-associated factor 15 (PAF15) in the DNA replication coupled, or the tail of histone H3 in the DNA replication uncoupled maintenance pathway [47]. Subsequently, DNMT1 recognises ubiquitinylated proteins and methylates CpGs on the newly synthetised strand in dividing cells.

DNA methylation can also be removed, by two distinct mechanisms (**Figure 2**). Passive demethylation is the first mechanism. It refers to the dilution of DNA methylation through cell divisions and can be caused by the absence or inhibition of DNMT1 [54,55], absence of UHRF1 [56] or delocalization of UHRF1 and DNMT1 to the cytoplasm [57,58]. For instance, STELLA (developmental pluripotency-associated protein - DPPA3 or PGC7), a small protein highly expressed in primordial germ cells (PGCs), oocytes and preimplantation embryos and essential for female fertility, is involved in the delocalization of UHRF1 in oocytes via binding to its PHA domain [59,60]. Active demethylation is the second process. It involves the removal of the methyl group from 5-methylcytosine (5mC) and is carried out by the ten-eleven translocation (TET) family of proteins. The three family members, TET1, TET2 and TET3, exhibit oxidizing activity and can catalyse the conversion of 5mC to

5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [61,62]. In addition, thymine-DNA-glycosylase (TDG)-catalysed base excision and the DNA base excision repair pathway can remove 5fC and 5caC so that unmodified cytosines can be incorporated[63].

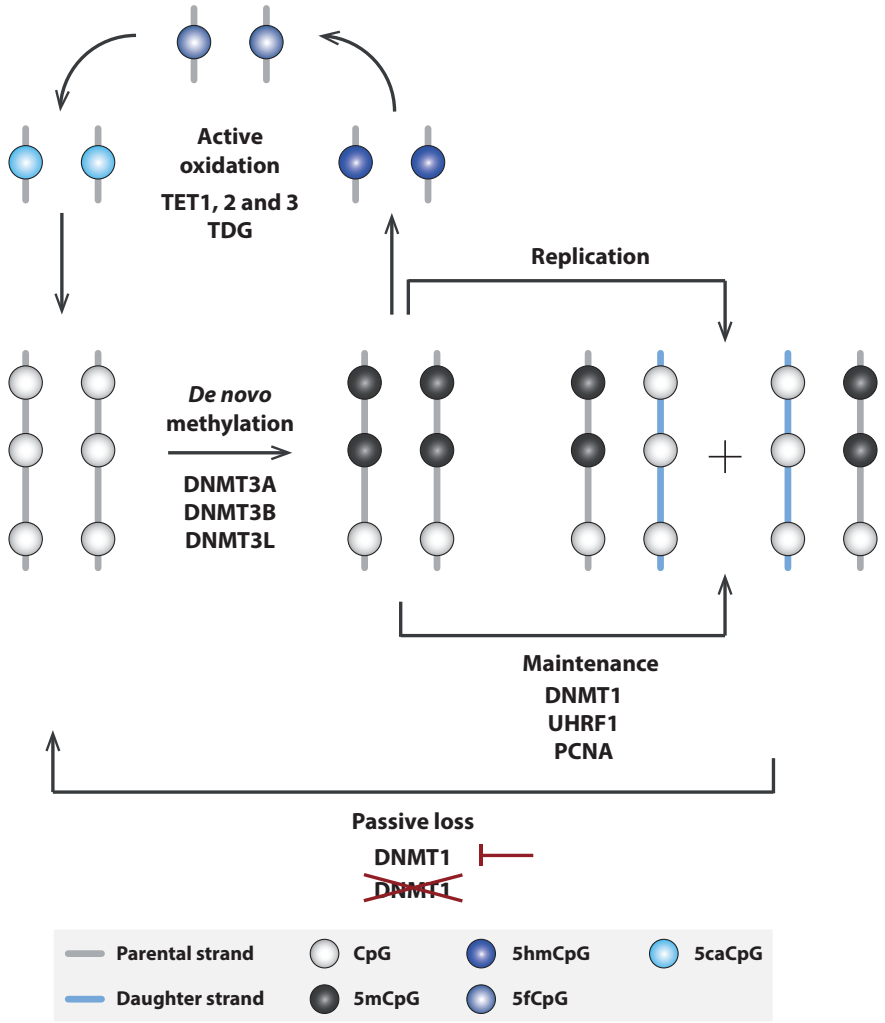

Figure 2. Model of functions of the DNA methylation machinery in establishment and maintenance of DNA methylation patterns. During early embryonic development and gametogenesis DNA methylation is established by the de novo DNA methyltransferases DNMT3A and DNMT3B, with the help of co-factors such as DNMT3L. After every cell cycle, methylation patterns are maintained in daughter cells by DNMT1, based on the methylation state of the parental strand. DNMT1 recognizes replication foci and hemi-methylated DNA with the help of PCNA and UHRF1. DNA methylation can also be removed; either through absence or inhibition of DNMT1 (passive loss), or through the oxidizing activity of TET enzymes (active demethylation).

EPIGENETIC REPROGRAMMING

For the correct placement of DNA methylation patterns and normal development, DNA methylation first needs to be erased during early preimplantation development and germline formation, a process often referred to as epigenetic reprogramming [64]. This mechanism of erasure ensures the restoration of the developmental potency of the zygote.

Epigenetic reprogramming includes two rounds of genome-wide erasure of epigenetic marks. More specifically, it occurs in the developing primordial germ cells (PGCs) during early to mid-gestation in utero, and in the preimplantation embryo. During PGC reprogramming, global erasure of DNA methylation and histone modifications takes place [65,66]. In females, this process continues even after birth in the growing oocytes of juvenile mice [67,68]. After the epigenetic marks are stripped, DNA methylation re-establishment in PGCs is achieved through the combined action of the *de novo* methyltransferases DNMT3B, DNMT3A and the DNMT3A cofactor DNMT3L [54]. The second wave of reprogramming takes place during pre-implantation development and involves the clearing of the epigenetic marks that defined the gene expression patterns of the mature parental gametes i.e., the sperm or the oocyte [69,70]. This process enables the zygote to return to a state of totipotency, and allows the cells of the early embryo to differentiate into any of the different cell types in our body. Post-zygotic DNA methylation re-establishment starts in the inner cell mass (ICM) of the blastocyst, with the major wave of *de novo* methylation by DNMT3A and DNMT3B taking place between E3.5 and E6.5 during mouse development [54,64].

Notably, it has been suggested that in both rodents and humans, epigenetic reprogramming is not complete, and regions that can 'escape' reprograming have been reported [71-74]. For example, retrotransposons that belong to the LINE/LTR family, and also imprinted regions, were found to be protected from the global demethylation process during PGC reprogramming in the mouse [74]. Similar results have been reported for humans where evolutionarily young repetitive elements but also some single-copy genes can escape from reprogramming during early human [73] and PGC development [72], reviewed in [75].

HISTONE PROTEINS AND THEIR POST-TRANSLATIONAL MODIFICATIONS

To fit DNA, which is approximately 10,000 times longer than an average diameter of a nucleus, within a nucleus, and at the same time 'expose' genes that need to be transcribed or 'conceal' genes that should be silenced, cells have developed an elegant and efficient DNA packaging strategy. Specifically, ~147 base pairs (bp) of DNA are wrapped around an octamer of histone proteins (two copies of four different core histon es, most often H2A, H2B, H3 and H4), that together form the nucleosome, the basic unit of chromatin [76] Chromatin is often described as 'beads-on-a-string', where the nucleosomes represent the beads, which are held together by the DNA 'string' [77].

The core histone proteins are built of a globular domain and free amino-terminal ends, called histone tails, that extrude from the surface of the nucleosome [78]. Both globular domains and histone tails, especially those of histones H3 and H4, can acquire a diverse array of post-translational modifications (PTMs), also known as covalent modifications or histone marks. They include, but are not limited to, methylation, acetylation, phosphorylation and ubiquitination [79]. Histone modifications can be found throughout the mammalian genome. Depending on the genomic context and across different

genomic elements, histone modification composition differs [80]. The correct genome-wide distribution of each histone mark, is tightly controlled by 'writer' and 'eraser', enzymes that deposit or remove these chemical tags from specific amino acid residues of histone proteins, respectively [81]. For instance, histone methyl transferases (HMTs) and histone acetyl transferases (HATs) place methyl and acetyl groups, respectively. On the other hand, histone demethylases and histone deacetylases (HDACs) have the opposite effect and are required to remove them. Furthermore, a large number of histone marks can serve as docking sites for nuclear regulatory proteins i.e., readers [82,83]. Together, the 'writers', 'erasers' and 'readers' cooperate, to shape the chromatin landscape throughout development and in response to environmental cues.

Initially defined by Emil Heitz as densely stained parts of the chromosomes [84], condensed heterochromatin is decorated with repressive histone marks and stays mainly transcriptionally silent during interphase [85]. Heterochromatin can be further subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin is tightly packed and attenuates transcription of more than half of the mammalian genome. This includes the highly repetitive (peri)centric and subtelomeric regions, non-coding and gene-poor regions and transposable elements [86]. In general, constitutive heterochromatin is maintained throughout development and can be considered as non-cell-type specific [86]. In mammals, methylation of H3K9, the main constitutive heterochromatin mark, is catalysed by six different HMTs that belong to three different families, namely Suppressor of variegation 3–9 homologue 1 and 2 (SUV39H1/SUV39H2) [87], SET domain bifurcated 1 and 2 (SETDB1/SETDB2), and G9A (also known as euchromatic histone-lysine N-methyltransferase 2 – EHMT2) [88]/ G9A-like protein (GLP, also known as EHMT1). Recent work in mouse embryonic fibroblasts (MEFs) showed that only the combined knock out of all six enzymes results in a collapse of heterochromatin [89], a phenotype that has been attributed to their partial redundancy. However, all six HMTs also have unique genomic targets, many of which appear to be different types of repetitive DNA. For instance, SUV39H1/2, initially discovered in a genetic screen for modifiers of position effect variegation (PEV) in *Drosophila* (as Su(var)3–9) [90] as a heterochromatin associated protein [91], are required for H3K9me3 of satellite repeats located in pericentric and telomeric heterochromatin in MEFs [92,93]. In addition, SUV39H1-mediated H3K9me3 is required for silencing of LINE1 retrotransposons in mESCs [94]. On the other hand, SETDB1 is required to catalyse H3K9me3 at intracisternal A-particle (IAP) retrotransposons in mESCs [95]. IAP elements are evolutionarily young and retain the ability to undergo transposition, therefore they pose a threat to the genome if activated [96,97]. This makes them high priority silencing targets. In mESCs, SETDB1 targets these elements for H3K9 methylation through its interaction with Krüppel-associated box (KRAB)-interacting protein 1 (KAP1, also known as Trim28) and KAP1's KRAB domain-containing zinc-finger (KRAB-ZFPs) partner proteins [86,98]. In somatic cells, the combinatorial effect of repressive histone marks, mainly H3K9me3 and H3K9me2, together with DNA methylation, ensures the locked chromatin state at these sites [86], thereby protecting genome integrity.

Facultative heterochromatin was originally described as 'developmentally regulated heterochromatinization of only one allele of a homologous chromosome pair' [99]. Today, facultative heterochromatin refers to the parts of the genome where chromatin is generally in a condensed state, but can adopt an active state at a specific developmental stage or in a specific cell type [100,101]. In mammals, facultative heterochromatin is mainly found over developmental genes that are expressed in a cell type specific manner [102]. They

include the well-known example of the clustered *homeotic* (*Hox*) genes [103,104]. Furthermore, X-chromosome inactivation serves as a paradigm for facultative heterochromatin formation [101]. Mechanistically, the formation of facultative heterochromatin is primarily orchestrated by Polycomb Group proteins (PcG) [102]. In addition, the histone methyltransferases G9a and Glp have been described to take part [86]. PcG are multi protein complexes that can be divided in two main groups, Polycomb repressive complex 1 (PRC1) and PRC2 [105]. The PRC2 complex is required for placing methylation on lysine 27 of histone H3 (H3K27me1/2/3 marks). Its catalytic subunits, namely enhancer of zest 1 (EZH1) (expressed in all the cells) and EZH2 (more active in actively dividing cells) place methylation on H3K27 [102]. The PRC1 complex, on the other hand, establishes mono-ubiquitination of histone H2AK (H2AK119ub1) via its RING1 (A or B) catalytic subunit [105]. To propagate and maintain a repressive chromatin environment and gene silencing, the two PcG complexes usually function in cooperation [102]. Dysregulation of either of the two complexes can lead to disruption of normal development and has been associated with neurodevelopmental and neurodegenerative diseases, type II diabetes and cancer [106-109].

Euchromatin was originally defined by its sparse staining pattern [84]. This definition has evolved and now, it represents decondensed and transcriptionally accessible chromatin. More specifically, euchromatin encompasses the gene-rich and actively transcribed regions of the genome and is marked by active histone modifications [80,85]. The well-defined active histone marks that decorate euchromatin include different methylation states of lysine 4 on histone H3 (H3K4) as well as histone acetylation [80]. The Complex Proteins Associated with Set1 (COMPASS) family, is responsible for catalysing methylation at H3K4 [110]. It is a multiprotein complex and in mammals, it can contain one of the six methyltransferase proteins. They include SET domain containing 1A (SETD1A) and SETD1B, Mixed lineage leukaemia 1 (MLL1, also known as KMT2A), MLL2 (also known as KMT2B), MLL3 (also known as KMT2C) and MLL4 (also known as KMT2D) [111]. H3K4me3 is typically found around promoters of actively transcribed genes. Additionally, H3K4 can be mono and di methylated. For example, H3K4me1 is found at active and poised enhancer elements while H3K4me2, similarly to H3K4me3, is found at transcriptionally active genes [112,113]. Acetylation of 27th lysine of histone H3 (H3K27ac) can be deposited by various acetyltransferases including p300 and CBP [114]. When found together with H3K4me1, the two modifications are a characteristic of active enhancers (**Figure 3**). Histones H3 that are positioned throughout transcribed gene bodies are marked by tri- methylation at lysine 36 (H3K36me3, **Figure 3**), which is deposited by SET domain containing 2 (SETD2). In addition, H3K36 exists in a di- methylated state (H3K36me2). In this form, it is found at intergenic euchromatin. Several HMTases, including NSD1, NSD2, NSD3 and ASH1L, can place the H3K36me2 modification [16]. The fact that mutations in histone methyltransferases responsible for labelling euchromatin cause neurodevelopmental growth syndromes, such as Sotos syndrome, Luscan–Lumish syndrome and Kabuki syndrome [16,115-117], emphasizes the importance of the correct placing of epigenetic marks for normal development.

MODIFIERS OF MURINE METASTABLE EPIALLELES DOMINANT SCREEN

With the aim to broaden our knowledge of epigenetic control and to discover and characterize all epigenetic players, forward mutagenesis screens have been carried out in different species, including *Drosophila*, *Arabidopsis* and mouse [118-120]. Some of the earliest screens were done in *Drosophila* and made use of position-effect variegation (PEV) [118,121-123], while the later screens mainly used transgenes and, in some cases, endogenous alleles susceptible to epigenetic silencing to reveal the identities of the genes important

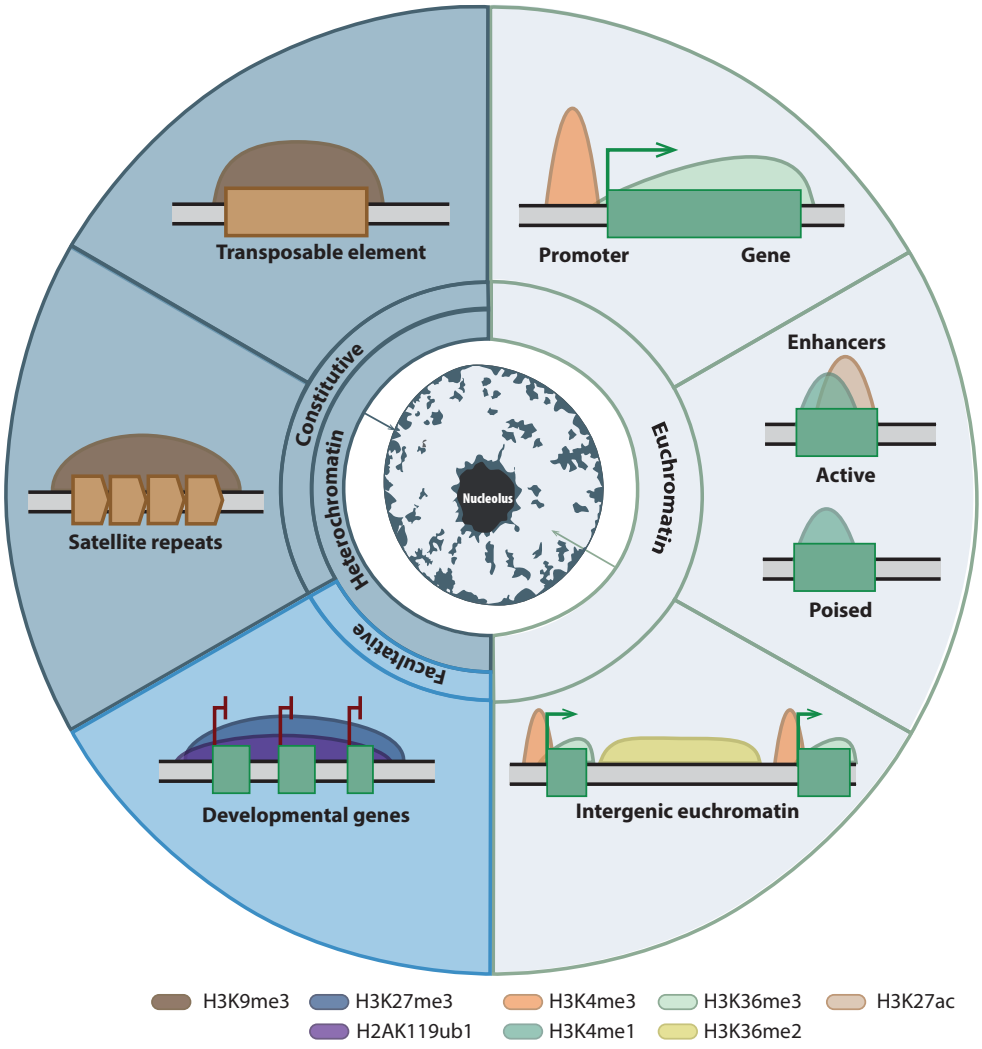

Figure 3. Characteristics of euchromatin and heterochromatin./Histone marks of euchromatin and heterochromatin. In the centre, a drawing of a nucleus as seen using electron microscopy. Darker patches represent heterochromatin, which stains more densely than euchromatin. Heterochromatin is primarily located at the periphery, while the euchromatin is spread throughout the rest of the nucleus. The right half of the circle illustrates genomic contexts where euchromatic marks prevail, while the left half depicts genomic regions where heterochromatin (either constitutive or facultative) is found, along with the corresponding histone marks and their representative distribution.

for epigenetic regulation [119,124-130]. Most relevant for this thesis is the *Modifiers of murine metastable epialleles Dominant* (*MommeD*) screen that was established and carried out in the laboratory of Prof. Emma Whitelaw [124,131]. The screen relies on a mouse line (*Line3*), which carries a green fluorescent protein (GFP) transgene array, tandemly integrated 5 times into a gene-poor region on chromosome 1 (**Figure 4**). The transgene is under the control of the

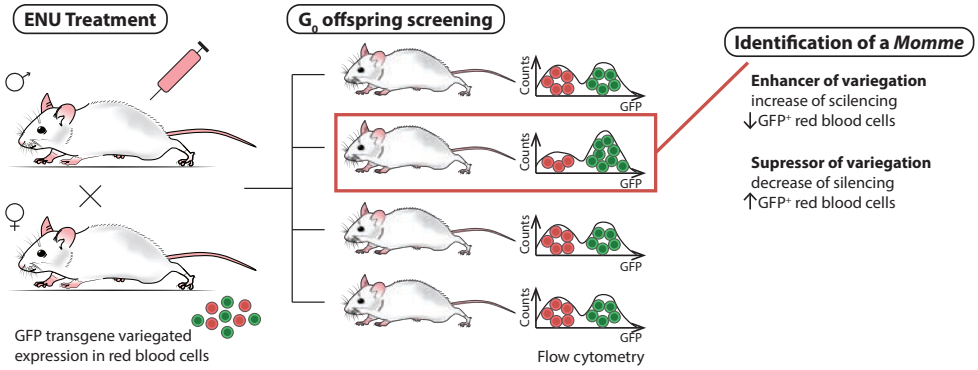

Figure 4. *Modifiers of murine metastable epialleles Dominant* screen. *Line3* males are treated with ENU mutagen, and after recovery bred with *Line3* females. G_0 offspring is screen for shift in percentage of GFP expressing red blood cells using drop of peripheral blood and flow cytometry analysis. Factors which when mutated increase percentage of GFP positive cells are referred to as Suppressors of variegation *Su(var)*, while factors that when mutated decrease percentage of GFP positive red blood cells are termed Enhancers of variegation *E(var)* (adapted from [119]).

alpha globin promoter and enhancer sequences, thus directed to be expressed specifically in red blood cells. In *Line3*, GFP is expressed in a variegated manner, approximately 55% of erythrocytes are GFP positive [132]. Based on the hypothesis that epigenetic modifiers would alter the extent of variegation, a *N-ethyl-N-nitrosourea* (ENU) screen was carried out and a shift in the percentage of GFP expressing red blood cells in the offspring of ENU-treated males (≥10% in either direction) was used as a readout to identify mutants. According to *Drosophila* nomenclature, factors which when mutated increase the percentage of GFP positive cells are referred to as *Suppressors of variegation Su(var)*, while factors that when mutated decrease the percentage of GFP positive red blood cells are termed *Enhancers of variegation E(var)*. This screen is for dominant effects. In total, more than 40 mutant alleles (termed *MommeD*) with mutations in over 30 different genes have been described, utilizing this screen [119,131]. Among the proteins discovered through the *MommeD* screen are factors involved in DNA methylation, histone modification and chromatin remodelling pathways. They include well-known epigenetic modifiers such as DNMT1 [131,133], DNMT3B [134], SETDB1 [131], UHRF1 [131], SMARCA4 [131] and HDAC1 [124,135], but also lesser-known ones, including SMCHD1 [124,136], NSD1 (Vukic and Daxinger, unpublished), Widely-interspaced zinc finger (WIZ) [131] and Rearranged L-myc fusion (RLF) [131,135,137], Rap-interacting factor 1 (RIF1) [131]. MommeD's have been implicated in X-chromosome inactivation [136], transposon silencing [138-140], the regulation of complex genomic structures such as large gene clusters [141-143] and (satellite) repeats [134,140]. Furthermore, MommeD's are emerging as key players in human diseases. Somatic mutations and copy number variations for *MommeD*'s have been reported for many cancers (COSMIC, TCGA). The most notable congenital disorders with direct links to mutations in *MommeD* genes are Facioscapulohumeral muscular dystrophy (FSHD), Bosma arhinia microphthalmia syndrome (BAMS) [144], Coffin-Siris, Sotos and ICF syndrome [31,115,145-148].

IMMUNODEFICIENCY, CENTROMERIC INSTABILITY, FACIAL ANOMALIES SYNDROME

Genetic defects in all three catalytically active DNA methyltransferases, *DNMT1*, *DNMT3A*, and *DNMT3B*, have been associated with human congenital disorders [17,149,150], emphasizing the importance of DNA methylation for normal mammalian development. One of the earliest reports of abnormal DNA methylation patterns in disease was in patients with ICF syndrome (OMIM 602900) approximately 40 years ago [151-153]. ICF syndrome is a rare, autosomal recessive disorder and less than 150 patients have been reported worldwide [154-158]. The main characteristics of the disease are unusual facial features, reduced levels or absence of serum immunoglobulins in the presence of B cells, and chromosome instability which is reflected in aberrant configurations of chromosomes 1, 9, and 16 in mitogen-stimulated lymphocytes [159]. The mild facial dysmorphisms frequently include hypertelorism, epicanthic folds, a flat nasal bridge, and low-set ears [160]. Most ICF patients are diagnosed at a young age, suffer from recurrent gastrointestinal and respiratory tract infections, sepsis and a failure to thrive, which often results in early childhood death [159]. An early study on ICF patient-derived peripheral blood found a lack of memory B and plasma cells that could explain the hypo- and agammaglobulinemia phenotype [161], and hematopoietic stem cell transplantation has been used to treat ICF patients [159,162]. DNA hypomethylation of pericentromeric satellite 2 and 3 repeats [163] is the molecular hallmark of the disease and makes ICF syndrome the first example of a human disorder linked to a constitutive defect in DNA methylation. The finding that the centromeric α-satellite repeats are heterogeneously affected by the loss of DNA methylation among ICF patients has led to the early recognition of genetic heterogeneity of the disease [163].

The genetics of ICF syndrome

In 1999, the first mutation underlying the ICF syndrome was described and found to be located in the gene that encodes *DNMT3B* [31,147,148]. As mentioned above, DNMT3B is a *de novo* methyltransferase involved in the establishment of DNA methylation patterns early in life and during cell differentiation. The majority of ICF1 patients carry missense mutations in the C-terminal part of *DNMT3B*, where the catalytic domain is located. In addition, *DNMT3B* nonsense and splice-site mutations have been reported, and all patients described to date appear to have impaired methyltransferase activity [164,165]. *DNMT3B* nonsense mutations only occur together with missense mutations, and it has been suggested that complete loss of DNMT3B catalytic activity is not compatible with life [166]. Consistent with this hypothesis, deletion of *Dnmt3b* in mice results in embryonic lethality [164], while mice carrying an ICF patient mutation in *Dnmt3b* are viable and resemble some of the phenotypic features of ICF syndrome patients [164,167]. Patients with *DNMT3B* mutations are designated ICF type 1 (ICF1) and account for approximately half of the known ICF patients. Three additional ICF syndrome-associated genes have been identified. Nonsense mutations in *zinc-finger and BTB domain-containing 24* (*ZBTB24*) are found in approximately 30% of ICF patients, referred to as ICF2 [168]. Twelve ICF cases with mutations in *cell division cycle associated 7* (*CDCA7*) or in *helicase, lymphoid specific* (*HELLS*) have been reported [169,170], and are referred to as ICF3 and ICF4, respectively. A few patients remain who showed molecular and phenotypic features characteristic of ICF syndrome but do not have mutations in the four known genes, and were referred to as ICFX [159]. Recently, a compound heterozygous mutation in the DNA methylation maintenance factor *UHRF1* was reported in an ICF patient. While this patient displays the molecular hallmarks, such as (peri)centromeric hypomethylation, when compared to ICF1-4 patients, the *UHRF1* patient also showed atypical symptoms

(including macroglossia and navel hernia) and a mild immune phenotype [171].

Functions for ICF genes in DNA methylation pathways and beyond

While aberrant hypomethylation of pericentromeric satellite 2 and 3 repeats is a feature of all ICF patients [171,172], and is used for the diagnosis of ICF syndrome, the previously reported differences in repeat hypomethylation in ICF patients [163,173], can now be associated with the different genetic defects. More specifically, the α-satellite repeat hypomethylation phenotype is shared between ICF2, ICF3, and ICF4 types [172], while DNA methylation levels of subtelomeric repeats are not affected in these patients [174]. In contrast, heavily methylated α-satellites [173] and the hypomethylation of subtelomeric repeats [174] are unique features of patients carrying *DNMT3B* mutations, and can be used to distinguish ICF1 from other ICF patients. On the other hand, the newly reported UHRF1 patient shows aberrant hypomethylation of both α-satellites as well as subtelomeric repeats [171]. Consistent with a function for DNMT3B in the establishment of DNA methylation in early development [9,175], ICF1 patients show promoter hypomethylation of germline genes [176] and loss of methylation at X-linked genes [172]; these sequences are not affected in ICF2, ICF3, and ICF4 patients [172]. Indeed, in both mice and humans, it has been shown that ZBTB24, CDCA7, and HELLS could be involved in the DNMT1-dependent DNA methylation maintenance pathway. In mouse embryonic fibroblasts (MEFs), siRNA-mediated depletion of ZBTB24, CDCA7, and HELLS resulted in loss of DNA methylation at minor satellite repeats [169]. Similarly, a progressive loss of DNA methylation at satellite repeats has been reported in human embryonic kidney (HEK) 293 cells knocked out for ZBTB24, CDCA7, or HELLS [177]. Interestingly, transient depletion of each of the four ICF factors in normal human primary fibroblasts did not result in hypomethylation of satellite 2 or subtelomeric sequences [174].

Biochemical studies have demonstrated that HELLS can interact with DNMT1 *in vitro* [178], and interactions between CDCA7 and HELLS [172,177,179], and CDCA7 and UHRF1 have also been reported [177]. A recent study in *Xenopus* showed that Cdca7 is required for localization of Hells to chromatin and that Cdca7 can stimulate Hells nucleosome remodeling activity [179]. Furthermore, studies in human cell lines showed that HELLS/LSH enhances UHRF1 chromatin association [180] and suggest that replication-uncoupled maintenance relies on the chromatin remodeling activity of HELLS/LSH [48]. Therefore, the abnormal DNA methylation patterns detected in ICF3 patients could be the result of a defect in HELLS recruitment to chromatin [179]. Of note, some hypomethylated loci are shared between ICF1-4 patients [172], and there is evidence that HELLS participates in DNMT3A/B-mediated *de novo* methylation in MEFs and during the differentiation of mESCs [181,182]. Thus, roles for ZBTB24, CDCA7, and HELLS in *de novo* DNA methylation cannot be excluded, but are as yet undetermined. Importantly, a recent evolutionary study could trace back the presence of CDCA7 and HELLS to the last eukaryotic common ancestor (LECA), and further showed that the two proteins are generally absent in species that do not have DNA methylation such as *C.elegans* and *Drosophila* [183].

Depending on the developmental stage, targeted genomic location, protein complex composition, cell cycle, or post-translational modifications, ZBTB24, CDCA7, and HELLS could also have functions that differ from DNA methylation regulation but are relevant to ICF syndrome pathology. For instance, both CDCA7 and HELLS have recently been linked to the classical non-homologous end joining (c-NHEJ) pathway [177]. CDCA7 was shown to interact with members of the c-NHEJ pathway, and depletion of both CDCA7 and HELLS in HEK293T cells was associated with c-NHEJ defects and delayed accumulation of

Ku80 at sites of DNA damage [177]. Unresolved Holiday junctions have been suggested to underlie multiradial chromosome formation in ICF patients [184], and the findings by Unoki et al. provide supporting evidence for this hypothesis. In the case of ZBTB24, its function as a transcription factor is now beginning to be understood. The functional connection between ZBTB24 and CDCA7, whereby ZBTB24 positively controls CDCA7 expression, was initially demonstrated in a mESC-based Zbtb24 knockout model and in fibroblasts and T cells derived from ICF2 patients [185]. Subsequent studies have confirmed these results [172,177,186-188]. Recent genome-wide ChIP-seq studies reported that ZBTB24 is mainly enriched at CpG-rich promoters [186,187] and that its C2H2 zinc fingers direct sequence-specific binding [187]. Down-regulation of cdca7 has also been observed in a zbtb24 knockout zebrafish model [189], demonstrating that CDCA7 is probably a highly conserved ZBTB24 target.

ICF syndrome patients have an altered DNA methylation landscape

Pericentromeric satellites 2, 3 and centromeric α-satellites are classical targets of DNA hypomethylation in ICF syndrome patients. In addition, abnormal DNA methylation of subtelomeric repeats is a feature of ICF1 patients [174,190,191] and the UHRF1 patient [171]. Initially ICF-related DNA hypomethylation appeared to be confined to the repetitive compartment of the genome [173,174,192-196], and only a few single copy loci exhibiting DNA hypomethylation in ICF patients were known [176,197-201]. Rapid developments in next-generation sequencing technologies and the identification of *ZBTB24*, *CDCA7*, and *HELLS* as ICF disease genes have stimulated research in this area, allowing a more detailed view of the ICF-related methylome. The first genome-wide DNA methylation map, generated from ICF1 patient-derived cultured lymphoblastoid cell lines (LCLs), has revealed widespread hypomethylation and a 41% reduction in global methylation levels when compared with a control individual [202]. All autosomes were equally affected and loss of methylation was not restricted to repetitive regions but detected along all genomic features including promoters, gene bodies, and intergenic regions (**Figure 5**) [202]. A later study, using independent ICF1 patient-derived cultured LCLs and reduced-representation bisulfite sequencing (RRBS), reported similar findings with respect to global DNA methylation alterations [203]. In mESCs, it has been reported that DNMT3B is the main enzyme responsible for genic DNA methylation [40], and mESCs lacking DNMT3B lost DNA methylation across exons and introns [26]. Consistent with this, ICF1 patients show hypomethylation of gene bodies [203]. It has also been shown that most DNMT3B binding occurs at intragenic positions of genes and that DNMT3B binding is not impaired upon the introduction of an ICF1 mutation in the catalytic domain [203]. Intriguingly, *DNMT3B* patients with a homozygous missense mutation close to the PWWP domain have been reported [204], demonstrating that the ICF syndrome phenotype can be recapitulated by mutations in the regulatory N-terminal part of DNMT3B. Patients carrying the PWWP mutation show DNA hypomethylation of satellite 2 repeats [204], and it will be interesting to determine whether they have additional methylation defects that are similar or different from the ones that have been described for ICF1 patients with impaired DNMT3B catalytic activity. Of note, introduction of the PWWP mutation into mESCs resulted in loss of DNMT3B recruitment to histone H3K36me3-decorated gene bodies [40]. While ICF1 patients show prominent loss of DNA methylation in CpG-rich regions (CGI and CpG shores) and gene bodies, it has become evident that in addition to the satellite repeats, CpG-poor, heterochromatic regions that are often referred to as 'open sea', are predominantly affected in ICF2, 3, and 4 patients (Figure 5). This was only recently revealed in a comprehensive study comparing 'primary' whole blood methylomes generated from 15 patients comprising samples from ICF1-4

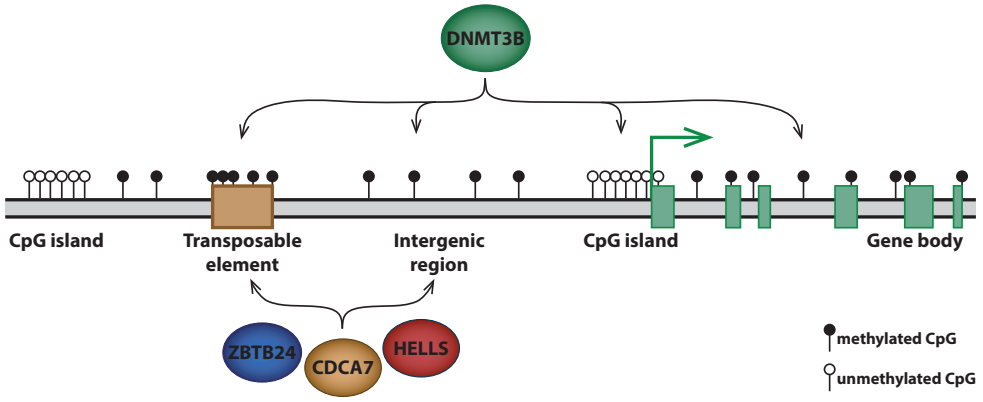

Figure 5. Schematic of the mammalian DNA methylation landscape and the putative influence of ICF genes on DNA methylation. The distribution of DNA methylation across different genomic elements, such as transposable elements, intergenic regions, CGI and gene bodies. DNMT3B can influence DNA methylation establishment genome-wide, with the exception of CGI that are usually protected from DNA methylation. ZBTB24, CDCA7, and HELLS contribute to DNA methylation maintenance and possibly establishment mainly at intergenic regions and repetitive elements. The empty circles represent unmethylated cytosines, while filled circles represent methylated cytosines (adapted from: https://commons.wikimedia.org/wiki/File:DNAme_landscape.png).

subtypes and ten gender- and age-matched controls [172]. Many of the hypomethylated CpGs in ICF2, 3, and 4 patients mapped to late-replicating regions and were found to be located in the open sea regions of the genome [172], which are characterized by low CpG density [205]. Heterochromatin is usually characterized by the presence of the repressive histone modifications H3K9me2 and H3K9me3, and binding of HP1 [206]. DNA methylation can contribute to heterochromatin formation through interaction with components of the chromatin remodeling and histone modification machinery, since DNMTs cannot bind and efficiently methylate nucleosomal DNA [207,208]. For instance, DNMTs can interact with the chromatin remodeler HELLS (also known as LSH) [178] or histone-modifying enzymes, such as SUV39H1 [209], and the H3K9me3 reader, UHRF1 [210]. In addition, DNMT3B can be directly recruited to centromeric heterochromatin by centromere protein C (CENPC) [211] or the linker histone H1 [212,213]. How ZBTB24, CDCA7, and HELLS contribute to DNA methylation and possibly heterochromatin formation at (peri)centromeric repeats and open sea regions, and why these sites are specifically affected in ICF2, ICF3 and ICF4 patients, remains an open question. Gene regulatory elements are also located in open sea regions, and it has recently been suggested that short-term depletion of HELLS in MEFs is associated with increased chromatin accessibility at selected genes and enhancers [214]. Hypomethylation of large chromosomal domains has been observed in MEFs derived from *Hells*$^{-/-}$ embryos, and a function for HELLS in controlling DNA methylation in a nuclear compartment partly defined by lamina-associated domains (LADs) has been proposed [215]. Given the functional connections between the three disease genes and the overlapping hypomethylation phenotype in ICF2–4 type patients it will be interesting to test whether ZBTB24, CDCA7, and HELLS may also have roles in LAD and/or enhancer regulation. Members of big gene cluster families including the *olfactory receptor* genes and *the clustered protocadherins* (*PCDH*) exhibit aberrant DNA methylation in ICF1-4 patient subtypes [172].

Epigenetic mechanisms play important roles in the regulation of gene clusters [216] and their dysregulation could be associated with ICF syndrome phenotypes. For example, the clustered *PCDH* genes, which are predominantly expressed in the nervous system, are key components of neuronal diversity [217]. In mice, it has been shown that DNMT3B is essential for the *de novo* methylation of *Pcdh* cluster genes in early embryonic development [218]. Defects in *Pcdh* cluster regulation have also been reported upon neuronal ablation of the H3K9me3 histone methyltransferase, SETDB1, in mice [219], suggesting cross-talk between DNA methylation and histone modification pathways in the regulation of this locus. In addition, SETDB1 is important for preserving a large topologically associated domain (TAD) encompassing the *Pcdh* locus [219]. Since intellectual disability is a common feature in ICF syndrome patients [159], it will be interesting to determine whether defects in *PCDH* gene cluster regulation can contribute to aspects of the disease. Finally, it is worth mentioning that abnormal DNA hypermethylation of different genomic features including promoters has been observed in ICF patients [172,202,203]. Although its relevance is not yet understood, given the well-established function of DNA methylation in gene silencing, it can be envisioned that DNA hypermethylation may also contribute to the phenotypes observed in ICF syndrome patients.

Linking aberrant DNA methylation to ICF syndrome phenotype

Indeed, the challenge now is to link DNMT3B-, ZBTB24-, CDCA7-, and HELLS-related genome-wide DNA methylation alterations to spatiotemporal gene expression changes that could explain ICF patient phenotypes. For a large part, progress in this area has been hampered by the scarcity of ICF patient-derived tissues or cell lines. Consequently, an integrated analysis of genome-wide DNA methylation and gene expression data has so far only been reported for cultured LCLs from ICF1 patients [203]. That study showed that aberrant gene body methylation in ICF1 patients was associated with alternative transcription start site (TSS) usage, the regulation of sense-antisense transcription and alternative exon splicing [203]. A similar function for gene body methylation in preventing aberrant transcript initiation has been shown in mESCs and could be ascribed to the catalytic activity of DNMT3B [26]. While it is challenging to directly translate observations from model systems or cultured cells to the ICF patient situation, it is feasible that spurious transcript initiation could critically affect phenotype, when associated with the production of aberrant proteins.

In addition to the ICF1-specific anomalous gene body methylation, ICF patient methylomes are characterized by hypomethylation of the heterochromatic compartment of the genome, as discussed above. By definition, such regions are transcriptionally inert and enriched for repressive epigenetic marks [220]. Genome-wide expression profiling in ICF1 patient-derived LCLs identified modest changes in gene expression outside gene bodies, although global DNA hypomethylation was observed [203]. Indeed, histone modification ChIP-seq revealed increased histone H3 lysine 27 three-methylated (H3K27me3) levels predominantly at hypomethylated CpGs and a compensatory function for this repressive mark has been suggested [203]. Re-distribution of H3K27me3 at gene cluster regions and surprisingly mild changes in gene expression have also been reported in *Hells*$^{-/-}$ MEFs, consistent with a redundancy in epigenetic silencing pathways upon loss of DNA methylation [215]. Whether these findings reflect convergent functions of DNMT3B and HELLS in *de novo* DNA methylation, or are a unifying feature in all ICF patients, remains to be determined.

SCOPE OF THE THESIS

In this thesis, we focused on CDCA7, which when homozygously mutated can cause Immunodeficiency Centromeric instability, Facial anomalies (ICF) syndrome in humans, with the aim to expand our understanding of how CDCA7 influences DNA methylation genome-wide. We generated and characterized two novel CDCA7 mouse models: one model carrying an ICF3-causing missense mutation (*Cdca7*^{G305V}) and one CDCA7 loss of function model (*Cdca7* KO). By leveraging these *in vivo* models, we investigated the epigenetic and transcriptomic consequences linked to changes in CDCA7-mediated DNA methylation across different developmental timepoints and in different tissues. In **chapter 2**, using our *Cdca7*^{G305V} model, we discovered that CDCA7 is important for maintaining DNA methylation within the B genomic compartment, which is mainly characterized by stretches of inactive chromatin and gene clusters with specialized functions and makes up about half of the mammalian genome. Notably, CDCA7-associated aberrant DNA hypomethylation translated to localized, tissue-specific transcriptional dysregulation, particularly affecting large gene clusters. Specifically, we found that, in cerebrum, hypomethylation of the clustered *protocadherins* was accompanied by their altered gene expression pattern – a finding potentially relevant to the ICF syndrome-associated neurodevelopmental phenotype. In **chapter 3**, we showed that a complete loss of CDCA7 leads to stochastic death of homozygous animals during embryogenesis. At the genomic loci tested, the DNA methylation phenotype and detected transcriptional dysregulation observed in *Cdca7* KO mice, mirrored the phenotype seen in *Cdca7*^{G305V} mice. In addition, we identify the first genic locus where loss of promoter DNA methylation correlates with transcriptional dysregulation in brain of two independent ICF-relevant mouse models carrying either a *Dnmt3b* (ICF1) or a *Cdca7* (ICF3) mutation. Lastly, in **chapter 4** we employed *CDCA7* (ICF3) and *HELLS* (ICF4) patient-derived T cells to characterize perturbations in the DNA methylation landscape. We found that DNA methylation loss mainly occurred within the B genomic compartment, in agreement with our observations in the mouse. Further analysis showed that while CpGs that are aberrantly hypomethylated in T cells highly overlap between ICF3 (*CDCA7*) and ICF4 (*HELLS*), CpGs that gain DNA methylation seem to be an ICF4 (*HELLS*)-specific feature. Interestingly, we found loss of DNA methylation in patient cells at the *T cell receptor* (*TCR*) loci, which combined with reports of opportunistic infections in these patients, underscores the need for more detailed studies on how the ICF syndrome causing genes function in T cell development. Combined, the work in this thesis reveals that, in addition to the well-studied satellite repeats, the B genomic compartment, which harbours large gene clusters with specialized, often cell type-specific functions, is a major target of the ICF syndrome-associated DNA methylation defect. While genomic sites with impaired DNA methylation patterns where overlapping between different developmental timepoints and across tissues, transcriptional dysregulation was cell-type specific. This emphasizes that an intricated, context-dependent relationship exists between cytosine methylation and transcription, offering new insights into the mechanisms underlying ICF syndrome and gene regulation in general.

REFERENCES

1. Avery, O.T., Macleod, C.M. & McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J Exp Med* **79**, 137-58 (1944).
2. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-95 (2011).
3. Waddington, C.H. The epigenotype. 1942. *Int J Epidemiol* **41**, 10-3 (2012).
4. Riggs, A.D. & Porter, T.N. Overview of epigenetic mechanisms. *Cold Spring Harbor monograph series* **32**, 29-46 (1996).
5. Hotchkiss, R.D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**, 315-32 (1948).
6. Mattei, A.L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends Genet* **38**, 676-707 (2022).
7. Razin, A. & Riggs, A.D. DNA methylation and gene function. *Science* **210**, 604-10 (1980).
8. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-32 (1975).
9. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**, 1093-100 (2010).
10. Walsh, C.P., Chaillet, J.R. & Bestor, T.H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics* **20**, 116-117 (1998).
11. Reik, W., Collick, A., Norris, M.L., Barton, S.C. & Surani, M.A. Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328**, 248-51 (1987).
12. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).
13. Mohandas, T., Sparkes, R.S. & Shapiro, L.J. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**, 393-6 (1981).
14. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
15. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321 (2015).
16. Janssen, S.M. & Lorincz, M.C. Interplay between chromatin marks in development and disease. *Nat Rev Genet* **23**, 137-153 (2022).
17. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet* **19**, 81-92 (2018).
18. Greenberg, M.V.C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* (2019).
19. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
20. Waterston, R.H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62 (2002).
21. Deniz, Ö., Frost, J.M. & Branco, M.R. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics* **20**, 417-431 (2019).
22. Hancks, D.C. & Kazazian, H.H., Jr. Roles for retrotransposon insertions in human disease. *Mob DNA* **7**, 9 (2016).
23. Edwards, J.R., Yarychkivska, O., Boulard, M. & Bestor, T.H. DNA methylation and DNA methyltransferases. *Epigenetics & Chromatin* **10**, 23 (2017).
24. Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat Genet* **47**, 469-78 (2015).
25. Straussman, R. *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* **16**, 564-71 (2009).
26. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77 (2017).
27. Maunakea, A.K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res* **23**, 1256-69 (2013).
28. Yearim, A. *et al.* HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep* **10**, 1122-34 (2015).
29. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74-9 (2011).

30. Jones, P.A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-92 (2012).
31. 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**, 247-57 (1999).
32. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* **15**, 545 (2014).
33. Dukatz, M. *et al.* Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. *Nucleic Acids Res* **48**, 11495-11509 (2020).
34. Jeltsch, A., Adam, S., Dukatz, M., Emperle, M. & Bashtrykov, P. Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. *J Mol Biol* **433**, 167186 (2021).
35. Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. & Bestor, T.H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536-9 (2001).
36. Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-3 (2004).
37. Dan, J. & Chen, T. Genetic Studies on Mammalian DNA Methyltransferases. *Adv Exp Med Biol* **945**, 123-150 (2016).
38. Stäubli, A. & Peters, A.H. Mechanisms of maternal intergenerational epigenetic inheritance. *Curr Opin Genet Dev* **67**, 151-162 (2021).
39. Veland, N. *et al.* DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Research* **47**, 152-167 (2018).
40. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-7 (2015).
41. Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife* **4**, e06205 (2015).
42. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* **285**, 26114-20 (2010).
43. Ooi, S.K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-7 (2007).
44. Du, J., Johnson, L.M., Jacobsen, S.E. & Patel, D.J. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* **16**, 519-32 (2015).
45. Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640-4 (2015).
46. Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* **10**, 1235-41 (2009).
47. Nishiyama, A. *et al.* Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. *Nat Commun* **11**, 1222 (2020).
48. Ming, X. *et al.* Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. *Cell Res* **30**, 980-996 (2020).
49. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-4 (2007).
50. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-12 (2007).
51. Unoki, M. & Sasaki, H. The UHRF protein family in epigenetics, development, and carcinogenesis. *Proc Jpn Acad Ser B Phys Biol Sci* **98**, 401-415 (2022).
52. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818-21 (2008).
53. Avvakumov, G.V. *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822-5 (2008).
54. Seisenberger, S. *et al.* Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci* **368**, 20110330 (2013).
55. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133 (2014).
56. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *Embo j* **32**, 340-53 (2013).
57. Maenohara, S. *et al.* Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS Genet* **13**, e1007042 (2017).
58. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev* **22**, 1607-16 (2008).

59. Du, W. *et al.* Stella protein facilitates DNA demethylation by disrupting the chromatin association of the RING finger-type E3 ubiquitin ligase UHRF1. *J Biol Chem* **294**, 8907-8917 (2019).
60. Li, Y. *et al.* Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564**, 136-140 (2018).
61. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
62. Kriaucionis, S. & Tahiliani, M. Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. *Cold Spring Harb Perspect Biol* **6**, a018630 (2014).
63. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18**, 517-534 (2017).
64. Hackett, J.A. & Surani, M.A. DNA methylation dynamics during the mammalian life cycle. *Philos Trans R Soc Lond B Biol Sci* **368**, 20110328 (2013).
65. Sasaki, H. & Matsui, Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* **9**, 129-40 (2008).
66. Tang, W.W., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M.A. Specification and epigenetic programming of the human germ line. *Nat Rev Genet* **17**, 585-600 (2016).
67. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* **8**, e1002440 (2012).
68. Smallwood, S.A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* **43**, 811-4 (2011).
69. Wennekamp, S., Mesecke, S., Nedelec, F. & Hiiragi, T. A self-organization framework for symmetry breaking in the mammalian embryo. *Nat Rev Mol Cell Biol* **14**, 452-9 (2013).
70. Eckersley-Maslin, M.A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol* **19**, 436-450 (2018).
71. Gill, M.E., Erkek, S. & Peters, A.H. Parental epigenetic control of embryogenesis: a balance between inheritance and reprogramming? *Curr Opin Cell Biol* **24**, 387-96 (2012).
72. Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**, 1437-52 (2015).
73. Okae, H. *et al.* Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet* **10**, e1004868 (2014).
74. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res* **23**, 616-27 (2013).
75. Vukic, M., Wu, H. & Daxinger, L. Making headway towards understanding how epigenetic mechanisms contribute to early-life effects. *Philos Trans R Soc Lond B Biol Sci* **374**, 20180126 (2019).
76. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60 (1997).
77. Allis, C.D., Caparros, M.L., Jenuwein, T. & Reinberg, D. *Epigenetics*, (Cold Spring Harbor Laboratory Press, 2015).
78. Ghoneim, M., Fuchs, H.A. & Musselman, C.A. Histone Tail Conformations: A Fuzzy Affair with DNA. *Trends Biochem Sci* **46**, 564-578 (2021).
79. Zhao, Y. & Garcia, B.A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7**, a025064 (2015).
80. Millán-Zambrano, G., Burton, A., Bannister, A.J. & Schneider, R. Histone post-translational modifications - cause and consequence of genome function. *Nat Rev Genet* **23**, 563-580 (2022).
81. Villaseñor, R. & Baubec, T. Regulatory mechanisms governing chromatin organization and function. *Curr Opin Cell Biol* **70**, 10-17 (2021).
82. Musselman, C.A., Lalonde, M.E., Côté, J. & Kutateladze, T.G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218-27 (2012).
83. Patel, D.J. A Structural Perspective on Readout of Epigenetic Histone and DNA Methylation Marks. *Cold Spring Harb Perspect Biol* **8**, a018754 (2016).
84. Heitz, E. *"Das" Heterochromatin der Moose*, (Bornträger, 1928).
85. Allis, C.D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016).
86. Padeken, J., Methot, S.P. & Gasser, S.M. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol Cell Biol* **23**, 623-640 (2022).
87. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-9 (2000).
88. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to

lysines 9 and 27 of histone H3. *J Biol Chem* **276**, 25309-17 (2001).

89. Montavon, T. *et al.* Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun* **12**, 4359 (2021).
90. Tschiersch, B. *et al.* The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *Embo j* **13**, 3822-31 (1994).
91. Schotta, G. *et al.* Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *Embo j* **21**, 1121-31 (2002).
92. Peters, A.H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-37 (2001).
93. García-Cao, M., O'Sullivan, R., Peters, A.H., Jenuwein, T. & Blasco, M.A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet* **36**, 94-9 (2004).
94. Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* **55**, 277-90 (2014).
95. Karimi, M.M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676-87 (2011).
96. Sharif, J., Shinkai, Y. & Koseki, H. Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs? *Philos Trans R Soc Lond B Biol Sci* **368**, 20110340 (2013).
97. Elmer, J.L. *et al.* Genomic properties of variably methylated retrotransposons in mouse. *Mob DNA* **12**, 6 (2021).
98. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. & Rauscher, F.J., 3rd. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**, 919-32 (2002).
99. Brown, S.W. Heterochromatin: Heterochromatin provides a visible guide to suppression of gene action during development and evolution. *Science* **151**, 417-425 (1966).
100. Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**, 1-13 (2007).
101. Żylicz, J.J. & Heard, E. Molecular Mechanisms of Facultative Heterochromatin Formation: An X-Chromosome Perspective. *Annu Rev Biochem* **89**, 255-282 (2020).
102. Yu, J.R., Lee, C.H., Oksuz, O., Stafford, J.M. & Reinberg, D. PRC2 is high maintenance. *Genes Dev* **33**, 903-935 (2019).
103. Zylicz, J.J. *et al.* Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *elife* **4**, e09571 (2015).
104. Mallo, M. & Alonso, C.R. The regulation of Hox gene expression during animal development. *Development* **140**, 3951-3963 (2013).
105. Kim, J.J. & Kingston, R.E. Context-specific Polycomb mechanisms in development. *Nat Rev Genet* **23**, 680-695 (2022).
106. von Schimmelmann, M. *et al.* Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nat Neurosci* **19**, 1321-30 (2016).
107. Gibson, W.T. *et al.* Mutations in EZH2 cause Weaver syndrome. *Am J Hum Genet* **90**, 110-8 (2012).
108. Lu, T.T. *et al.* The Polycomb-Dependent Epigenome Controls β Cell Dysfunction, Dedifferentiation, and Diabetes. *Cell Metab* **27**, 1294-1308.e7 (2018).
109. Parreno, V., Martinez, A.M. & Cavalli, G. Mechanisms of Polycomb group protein function in cancer. *Cell Res* **32**, 231-253 (2022).
110. Schneider, J. *et al.* Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* **19**, 849-56 (2005).
111. Cenik, B.K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and disease. *Nat Rev Genet* **22**, 38-58 (2021).
112. Rickels, R. *et al.* An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. *Mol Cell* **63**, 318-328 (2016).
113. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* **29**, 6074-85 (2009).
114. Scacchetti, A. & Bonasio, R. Histone gene editing probes functions of H3K27 modifications in mammals. *Nat Genet* **54**, 746-747 (2022).
115. Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* **30**, 365-6 (2002).
116. Luscan, A. *et al.* Mutations in SETD2 cause a novel overgrowth condition. *J Med Genet* **51**, 512-7 (2014).

117. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-3 (2010).

118. Elgin, S.C. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **5**, a017780 (2013).

119. Blewitt, M. & Whitelaw, E. The use of mouse models to study epigenetics. *Cold Spring Harb Perspect Biol* **5**, a017939 (2013).

120. Greenberg, M.V. *et al.* Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics* **6**, 344-54 (2011).

121. Sinclair, D.A., Mottus, R.C. & Grigliatti, T.A. Genes which suppress position-effect variegation in Drosophila melanogaster are clustered. *Molecular and General Genetics MGG* **191**, 326-333 (1983).

122. Schultz, J. Interrelations of factors affecting heterochromatin-induced variegation in Drosophila. in *Genetics* Vol. 35 134-134 (428 EAST PRESTON ST, BALTIMORE, MD 21202, 1950).

123. Spofford, J.B. SINGLE-LOCUS MODIFICATION OF POSITION-EFFECT VARIEGATION IN DROSOPHILA MELANOGASTER. I. WHITE VARIEGATION. *Genetics* **57**, 751-766 (1967).

124. Blewitt, M.E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A* **102**, 7629-34 (2005).

125. Percec, I., Plenge, R.M., Nadeau, J.H., Bartolomei, M.S. & Willard, H.F. Autosomal dominant mutations affecting X inactivation choice in the mouse. *Science* **296**, 1136-9 (2002).

126. Percec, I. *et al.* An N-ethyl-N-nitrosourea mutagenesis screen for epigenetic mutations in the mouse. *Genetics* **164**, 1481-94 (2003).

127. Wilson, L. *et al.* Random mutagenesis of proximal mouse chromosome 5 uncovers predominantly embryonic lethal mutations. *Genome Res* **15**, 1095-105 (2005).

128. Shumacher, A., Faust, C. & Magnuson, T. Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature* **383**, 250-3 (1996).

129. Carpinelli, M.R. *et al.* Suppressor screen in Mpl-/- mice: c-Myb mutation causes supraphysiological production of platelets in the absence of thrombopoietin signaling. *Proc Natl Acad Sci U S A* **101**, 6553-8 (2004).

130. Majewski, I.J. *et al.* Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol* **6**, e93 (2008).

131. Daxinger, L. *et al.* An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biol* **14**, R96 (2013).

132. Preis, J.I., Downes, M., Oates, N.A., Rasko, J.E. & Whitelaw, E. Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. *Curr Biol* **13**, 955-9 (2003).

133. Chong, S. *et al.* Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat Genet* **39**, 614-22 (2007).

134. Youngson, N.A. *et al.* No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. *Mamm Genome* **24**, 206-17 (2013).

135. Ashe, A. *et al.* A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol* **9**, R182 (2008).

136. Blewitt, M.E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet* **40**, 663-9 (2008).

137. Harten, S.K. *et al.* The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers. *BMC Biol* **13**, 21 (2015).

138. Isbel, L. *et al.* Trim33 Binds and Silences a Class of Young Endogenous Retroviruses in the Mouse Testis; a Novel Component of the Arms Race between Retrotransposons and the Host Genome. *PLoS Genet* **11**, e1005693 (2015).

139. Li, S. *et al.* Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *Proc Natl Acad Sci U S A* **113**, E5108-16 (2016).

140. Desai, V.P. *et al.* The role of MORC3 in silencing transposable elements in mouse embryonic stem cells. *Epigenetics Chromatin* **14**, 49 (2021).

141. Chen, K. *et al.* Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc Natl Acad Sci U S A* **112**, E3535-44 (2015).

142. Isbel, L. *et al.* Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. *Elife* **5**(2016).

143. Vukic, M. *et al.* CDCA7-associated global aberrant DNA hypomethylation translates to localized, tissue-specific transcriptional responses. *Sci Adv* **10**, eadk3384 (2024).

144. Gordon, C.T. *et al.* De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat Genet* **49**, 249-255 (2017).

145. Lemmers, R.J. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet* **44**, 1370-4 (2012).

146. Tsurusaki, Y. *et al.* Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet* **44**, 376-8 (2012).

147. Xu, G.L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187-91 (1999).

148. Hansen, R.S. *et al.* The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* **96**, 14412-7 (1999).

149. Velasco, G. & Francastel, C. Genetics meets DNA methylation in rare diseases. *Clin Genet* **95**, 210-220 (2019).

150. Heyn, P. *et al.* Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat Genet* **51**, 96-105 (2019).

151. Maraschio, P., Zuffardi, O., Dalla Fior, T. & Tiepolo, L. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. *J Med Genet* **25**, 173-80 (1988).

152. Hulten, M. Selective Somatic Pairing and Fragility at 1q12 in a Boy with Common Variable Immuno Deficiency. *Clinical Genetics* **14**, 294-294 (1978).

153. Tiepolo, L. *et al.* Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum Genet* **51**, 127-37 (1979).

154. van den Boogaard, M.L. *et al.* Expanding the mutation spectrum in ICF syndrome: Evidence for a gender bias in ICF2. *Clin Genet* **92**, 380-387 (2017).

155. Kiaee, F. *et al.* Clinical, Immunologic and Molecular Spectrum of Patients with Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome: A Systematic Review. *Endocr Metab Immune Disord Drug Targets* **21**, 664-672 (2021).

156. Baris, S. & Boluk, S.O. A novel mutation in DNMT3B gene causing ICF1 syndrome in an infant with refractory thrombocytopenia. *Clin Immunol* **256**, 109779 (2023).

157. Bilgic Eltan, S. *et al.* Evaluation of Clinical and Immunological Alterations Associated with ICF Syndrome. *J Clin Immunol* **44**, 26 (2023).

158. Fang, M. *et al.* T Cell Repertoire Abnormality in Immunodeficiency Patients with DNA Repair and Methylation Defects. *J Clin Immunol* **42**, 375-393 (2022).

159. Weemaes, C.M. *et al.* Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur J Hum Genet* **21**, 1219-25 (2013).

160. Ehrlich, M. *et al.* ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity* **41**, 253-71 (2008).

161. Blanco-Betancourt, C.E. *et al.* Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. *Blood* **103**, 2683-90 (2004).

162. Gennery, A.R. *et al.* Hematopoietic stem cell transplantation corrects the immunologic abnormalities associated with immunodeficiency-centromeric instability-facial dysmorphism syndrome. *Pediatrics* **120**, e1341-4 (2007).

163. Jeanpierre, M. *et al.* An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum Mol Genet* **2**, 731-5 (1993).

164. Ueda, Y. *et al.* Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* **133**, 1183-92 (2006).

165. Gagliardi, M., Strazzullo, M. & Matarazzo, M.R. DNMT3B Functions: Novel Insights From Human Disease. *Front Cell Dev Biol* **6**, 140 (2018).

166. Walton, E.L., Francastel, C. & Velasco, G. Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics* **6**, 1373-7 (2011).

167. Velasco, G. *et al.* Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* **107**, 9281-6 (2010).

168. de Greef, J.C. *et al.* Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am J Hum Genet* **88**, 796-804 (2011).

169. Thijssen, P.E. *et al.* Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat Commun* **6**, 7870 (2015).

170. Alghamdi, H.A. *et al.* Three Types of Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome Identified by Whole-Exome Sequencing in Saudi Hypogammaglobulinemia Patients: Clinical, Molecular, and Cytogenetic Features. *J Clin Immunol* **38**, 847-853 (2018).

171. Unoki, M. *et al.* Novel compound heterozygous mutations in UHRF1 are associated with atypical immunodeficiency, centromeric instability and facial anomalies syndrome with distinctive genome-wide DNA hypomethylation. *Hum Mol Genet* **32**, 1439-1456 (2023).

172. Velasco, G. *et al.* Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. *Hum Mol Genet* **27**, 2409-2424 (2018).

173. Jiang, Y.L. *et al.* DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum Mutat* **25**, 56-63 (2005).

174. Toubiana, S. *et al.* Subtelomeric methylation distinguishes between subtypes of Immunodeficiency, Centromeric instability and Facial anomalies syndrome. *Hum Mol Genet* **27**, 3568-3581 (2018).

175. Oda, M. *et al.* DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev* **20**, 3382-94 (2006).

176. Velasco, G. *et al.* Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. *Orphanet J Rare Dis* **9**, 56 (2014).

177. Unoki, M., Funabiki, H., Velasco, G., Francastel, C. & Sasaki, H. CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. *J Clin Invest* **129**, 78-92 (2019).

178. Dunican, D.S., Pennings, S. & Meehan, R.R. Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with Dnmt1. *Biomed Res Int* **2015**, 740637 (2015).

179. Jenness, C. *et al.* HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. *Proc Natl Acad Sci U S A* **115**, E876-e885 (2018).

180. Han, M. *et al.* A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. *Nucleic Acids Res* **48**, 12116-12134 (2020).

181. Zhu, H. *et al.* Lsh is involved in de novo methylation of DNA. *Embo j* **25**, 335-45 (2006).

182. Myant, K. *et al.* LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res* **21**, 83-94 (2011).

183. Funabiki, H., Wassing, I.E., Jia, Q., Luo, J.-D. & Carroll, T. Coevolution of the CDCA7-HELLS ICF-related nucleosome remodeling complex and DNA methyltransferases. (eLife Sciences Publications, Ltd, 2023).

184. Tuck-Muller, C.M. *et al.* DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. *Cytogenet Cell Genet* **89**, 121-8 (2000).

185. Wu, H. *et al.* Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. *Hum Mol Genet* **25**, 4041-4051 (2016).

186. Thompson, J.J. *et al.* ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res* **46**, 10034-10051 (2018).

187. Ren, R. *et al.* Structural basis of specific DNA binding by the transcription factor ZBTB24. *Nucleic Acids Res* (2019).

188. Wu, H., Vonk, K.K.D., van der Maarel, S.M., Santen, G.W.E. & Daxinger, L. A functional assay to classify ZBTB24 missense variants of unknown significance. *Hum Mutat* **40**, 1077-1083 (2019).

189. Rajshekar, S. *et al.* Pericentromeric hypomethylation elicits an interferon response in an animal model of ICF syndrome. *Elife* **7**(2018).

190. Simo-Riudalbas, L. *et al.* Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. *PLoS One* **10**, e0132517 (2015).

191. Sagie, S. *et al.* Induced pluripotent stem cells as a model for telomeric abnormalities in ICF type I syndrome. *Hum Mol Genet* **23**, 3629-40 (2014).

192. Miniou, P. *et al.* alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. *Hum Genet* **99**, 738-45 (1997).

193. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K. & Selig, S. Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum Mol Genet* **17**, 2776-89 (2008).

194. Deng, Z., Campbell, A.E. & Lieberman, P.M. TERRA, CpG methylation and telomere heterochromatin: lessons from ICF syndrome cells. *Cell Cycle* **9**, 69-74 (2010).

195. Miniou, P., Bourc'his, D., Molina Gomes, D., Jeanpierre, M. & Viegas-Pequignot, E. Undermethylation of Alu sequences in ICF syndrome: molecular and in situ analysis. *Cytogenet Cell Genet* **77**, 308-13 (1997).

196. 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).

197. Miniou, P. *et al.* Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. *Hum Mol Genet* **3**, 2093-102 (1994).

198. Bourc'his, D. *et al.* Abnormal methylation does not prevent X inactivation in ICF patients. *Cytogenet Cell Genet* **84**, 245-52 (1999).

199. Schuffenhauer, S. *et al.* DNA, FISH and complementation studies in ICF syndrome: DNA

hypomethylation of repetitive and single copy loci and evidence for a trans acting factor. *Hum Genet* **96**, 562-71 (1995).

200. Tao, Q. *et al.* Defective de novo methylation of viral and cellular DNA sequences in ICF syndrome cells. *Hum Mol Genet* **11**, 2091-102 (2002).

201. Jin, B. *et al.* DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum Mol Genet* **17**, 690-709 (2008).

202. Heyn, H. *et al.* Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. *Epigenetics* **7**, 542-50 (2012).

203. Gatto, S. *et al.* ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing. *Nucleic Acids Res* **45**, 5739-5756 (2017).

204. Shirohzu, H. *et al.* Three novel DNMT3B mutations in Japanese patients with ICF syndrome. *Am J Med Genet* **112**, 31-7 (2002).

205. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692-702 (2011).

206. Janssen, A., Colmenares, S.U. & Karpen, G.H. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **34**, 265-288 (2018).

207. Felle, M. *et al.* Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res* **39**, 6956-69 (2011).

208. Okuwaki, M. & Verreault, A. Maintenance DNA methylation of nucleosome core particles. *J Biol Chem* **279**, 2904-12 (2004).

209. Fuks, F., Hurd, P.J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* **31**, 2305-12 (2003).

210. Karagianni, P., Amazit, L., Qin, J. & Wong, J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol Cell Biol* **28**, 705-17 (2008).

211. Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G. & Robertson, K.D. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. *Human Molecular Genetics* **18**, 3178-3193 (2009).

212. Kashiwagi, K., Nimura, K., Ura, K. & Kaneda, Y. DNA methyltransferase 3b preferentially associates with condensed chromatin. *Nucleic Acids Res* **39**, 874-88 (2011).

213. Yang, S.M., Kim, B.J., Norwood Toro, L. & Skoultchi, A.I. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci U S A* **110**, 1708-13 (2013).

214. Ren, J., Finney, R., Ni, K., Cam, M. & Muegge, K. The chromatin remodeling protein Lsh alters nucleosome occupancy at putative enhancers and modulates binding of lineage specific transcription factors. *Epigenetics* **14**, 277-293 (2019).

215. Yu, W. *et al.* Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res* **24**, 1613-23 (2014).

216. Khamlichi, A.A. & Feil, R. Parallels between Mammalian Mechanisms of Monoallelic Gene Expression. *Trends Genet* **34**, 954-971 (2018).

217. Hirayama, T. & Yagi, T. Regulation of clustered protocadherin genes in individual neurons. *Semin Cell Dev Biol* **69**, 122-130 (2017).

218. Toyoda, S. *et al.* Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* **82**, 94-108 (2014).

219. Jiang, Y. *et al.* The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet* **49**, 1239-1250 (2017).

220. Lemaitre, C. & Bickmore, W.A. Chromatin at the nuclear periphery and the regulation of genome functions. *Histochem Cell Biol* **144**, 111-22 (2015).

Review also available at:


1