Hum Genet (2001) 109 : 452–462
DOI 10.1007/s004390100590

ORIGINAL INVESTIGATION

K. M. Anwarul Hassan · Thomas Norwood
Giorgio Gimelli · Stanley M. Gartler · R. Scott Hansen

Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication

Received: 14 June 2001 / Accepted: 24 July 2001 / Published online: 14 September 2001

Abstract Mutation in the *DNMT3B* DNA methyltransferase gene is a common cause of ICF (immunodeficiency, centromeric heterochromatin, facial anomalies) immunodeficiency syndrome and leads to hypomethylation of satellites 2 and 3 in pericentric heterochromatin. This hypomethylation is associated with centromeric decondensation and chromosomal rearrangements, suggesting that these satellite repeats have an important structural role. In addition, the satellite regions may have functional roles in modifying gene expression. The extent of satellite hypomethylation in ICF cells is unknown because methylation status has only been determined with restriction enzymes that cut infrequently at these loci. We have therefore developed a bisulfite conversion-based method to determine the detailed cytosine methylation patterns at satellite 2 sequences in a quantitative manner for normal and ICF samples. From our sequence analysis of unmodified DNA, the internal repeat region analyzed for methylation contains an average of 17 CpG sites. The average level of methylation in normal lymphoblasts and fibroblasts is 69% compared with 20% in such cells from ICF patients with *DNMT3B* mutations and 29% in normal sperm. Although the mean satellite 2 methylation values for these groups do not overlap, there is considerable overlap at the level of individual DNA strands. Our analysis has also revealed a pattern of methylation specificity, suggesting that some CpGs in the repeat are more prone to methylation than other sites. Variation in satellite 2 methylation among lymphoblasts from different ICF patients has prompted us to determine the frequency of cytogenetic abnormalities in these cells. Although our data suggest that some degree of hypomethylation is necessary for pericentromeric decondensation, factors other than DNA methylation appear to play a major role in this phenomenon. Another such factor may be altered replication timing because we have discovered that the hypomethylation of satellite 2 in ICF cultures is associated with advanced replication.

DDBJ/EMBL/GenBank accession nos. AF361703-AF361743

K.M.A. Hassan · S.M. Gartler · R.S. Hansen (✉)
Department of Medicine, University of Washington,
Seattle, WA 98195, USA

T. Norwood
Department of Pathology, University of Washington,
Seattle, WA 98195, USA

G. Gimelli
Laboratorio de Citogenetica, Istituto G. Gaslini, Genova, Italy

S.M. Gartler
Department of Genetics, University of Washington,
Seattle, WA 98195, USA

Present address:
R. Scott Hansen
Department of Genetics, Box 357360, University of Washington,
Seattle, WA 98195–7360, USA,
e-mail: supreme@u.washington.edu,
Tel.: +1-206-543-4184 or +1-206-543-1619, Fax: +1-206-543-0754

Introduction

MT3β DNA methyltransferase deficiencies lead to a selective loss of methylation in certain heterochromatic regions in both humans (Hansen et al. 1999; Okano et al. 1999; Xu et al. 1999) and mice (Okano et al. 1999). The mutant forms of the enzyme that underlie human ICF (immunodeficiency, centromeric heterochromatin, facial anomalies) immunodeficiency syndrome are associated with hypomethylation of satellites 2 and 3 at pericentromeric heterochromatin (Hansen et al. 1999; Okano et al. 1999; Wijmenga et al. 2000; Xu et al. 1999) and in heterochromatic regions of the Y and inactive X chromosomes (Hansen et al. 2000).

Pericentromeric satellite hypomethylation in ICF syndrome appears to promote the decondensation and other cytogenetic abnormalities that are characteristic of the disorder. Satellite hypomethylation may also play a role in higher order gene regulation (Hansen et al. 2000; Xu et al. 1999). The only satellite methylation data currently available are those based on nonquantitative restriction enzyme analyses of a small number of sites. Satellite 2 sequences are mainly concentrated in megabase-sized re-

gions near the centromeres of chromosomes 1 and 16. The higher order satellite 2 repeat unit is larger than 1 kb and is comprised of basic repeats of 23–26 bp containing frequent GGAAT and ATC sequences, and a less abundant sequence that is methylateable, viz., CGAAT (Jeanpierre 1994). In a typical 1.3-kb repeat unit of satellite 2, for example, there are about 69 CpGs, but only one or two *Bst*BI sites, the usual restriction enzyme chosen to analyze satellite 2 methylation. Less than 3% of the CpGs, therefore, can be analyzed for methylation with this method.

Given the importance of satellite methylation, we have developed a method to examine satellite 2 methylation in more detail in normal and ICF DNAs. The assay is based on the bisulfite conversion method that allows every CpG and other cytosines to be analyzed for methylation (Frommer et al. 1992). We examined methylation in an internal satellite 2 repeat of about 330 bp in which over 75% of satellite 2 CpGs should be represented. We have found that 63%–87% of the CpGs are methylated in normal cultured fibroblasts and lymphoblasts, whereas only 12%–31% are methylated in ICF cells with mutations in the *DNMT3B* gene. The ICF methylation levels are comparable with those observed in normal sperm cells (~28%). Although the mean methylation values for ICF versus normal somatic cells do not overlap, there is considerable overlap at the level of individual DNA strands in that 5% of ICF DNA strands have methylation levels around the mean of normals and vice versa. This compositional heterogeneity in methylation density could reflect a pattern of variation along the repeat array or heterogeneity between arrays on different chromosomes. We have also found methylation heterogeneity within individual clones in that certain conserved CpG sites are methylated at higher frequencies than other sites.

Our examination of the relationship between the characteristic centromeric decondensation in ICF and the extent of satellite 2 hypomethylation suggests that some degree of satellite 2 hypomethylation appears to be necessary for centromeric decondensation, but other factors play a significant role. Altered replication timing may be such a factor, for we have found that satellite 2 sequences in ICF cells replicate earlier than normal, consistent with other observations of a correlation between hypomethylation and advance replication. Advanced satellite replication may thus have a role in promoting the abnormal chromatin structure of these regions in ICF cells.

Materials and methods

Cells and culture conditions

Cells from six patients with classical features of ICF syndrome were examined in this study. These patients have been described previously (Gimelli et al. 1993; Hansen et al. 1999, 2000; Wijmenga et al. 2000) and include four with known mutations in *DNMT3B* (PT1 male, PT3 female, PT4 female, and PT5 male), and two that apparently lack *DNMT3B* mutations (PT10 female and PT12 male). Both lymphoblasts (LB) and fibroblasts (FB) were included in the study. Lymphoblastoid cells included those from PT1 (PT1 LB), PT4 (PT4 LB or GM08714), PT5 (PT5 LB), PT10 (PT10 LB), and PT12 (PT12 LB); fibroblasts included those from PT3 (PT3 FB) and PT4 (PT4 FB or GM08747). Lymphoblasts from PT4's father (F4 LB or GM08729) and mother (M4 LB or GM08728) were also used in this study. ICF family cell lines GM08714, GM08747, GM08728, and GM08729 were obtained from Coriell Cell Repositories. Other normal lymphoblasts and fibroblasts were as described previously (Hansen et al. 2000). Normal sperm cells were obtained as discarded material from an in vitro fertilization laboratory. Clonal human-hamster somatic cell hybrids HD14A1 and 14IC3, derived from fusion of PT4 LB and CHO-YH21 cells, were as described previously (Hansen et al. 1999, 2000). Cells were cultured as described previously (Hansen et al. 1988, 1998).

Genomic DNA isolation and bisulfite modification

Genomic DNA from fibroblasts, lymphoblasts, and hybrid cells were purified by the proteinase K/phenol method as described previously (Hansen et al. 1998). DNA from sperm was isolated by the Puregene DNA isolation kit (Gentra Systems, Minneapolis, Minn.) according to the manufacturer's instructions. Conversion of unmethylated cytosine to uracil by sodium bisulfite (Frommer et al. 1992) was essentially as described previously (Stöger et al. 1997). Duplicate samples of approximately 50–500 ng genomic DNA were used for bisulfite conversion.

Polymerase chain reaction amplification of satellite 2 sequences

To amplify a ~400 bp region of satellite DNA from the unmodified DNA, primers sat2–384U (5′-ATGGAAATGAAAGGGGTCATCATCT-3′) and sat2–781L (5′-ATTCGAGTCCATTCGATGATTCCAT-3′) were designed from the satellite 2 sequence with accession no. gi1220362. Approximately 200 ng genomic DNA was used in a 100-µl reaction with 200 µM dNTPs, 2.5 mM $MgCl_2$, 0.5 µM each primer, and 1.25 U *Taq* polymerase; the polymerase chain reaction (PCR) was performed under the following conditions in a PTC-200, MJ Research thermal cycler (Waltham, Mass.): 95°C for 5 min for one cycle, and then 95°C for 1 min, 60°C for 1 min, 72°C for 2 min, for 30 cycles, followed by 72°C for 7 min for one cycle. To amplify the same region from the upper strand of bisulfite-converted DNA, primer sequences were designed such that they contained several guanines and few cytosines, and CpG dinucleotides were synthesized as degenerate (C/T) in order to amplify both methylated and unmethylated sequences. The first round of PCR was performed with primers st2biup-1f, 5′-TTGAATGGAAATGAAAGGGGTTATTA-3′ (nt 380–405, gi1220362), and st2biup-2r, 5′-C[G/A]AATCCATTC[G/A]ATAATTCCATTCC-3′ (nt 754–778, gi1220362), and the reaction mixture contained 50 ng bisulfite-converted DNA (10 µl) in a 100-µl reaction, dNTPs, $MgCl_2$; primers and *Taq* polymerase were added as described above, and the PCR was performed in a PTC-200, MJ Research thermal cycler at 95°C for 5 min for one cycle, then 95°C for 1 min, 60°C for 1 min, 72°C for 2 min, for 7 cycles, followed by 95°C for 1 min, 58°C for 1 min, 72°C for 2 min for 13 cycles, and finally, 72°C for 7 min for one cycle. The PCR product was purified by using the Qiaquick PCR purification Kit (Qiagen, Valencia, Calif.) to remove primers and then eluted in a 100-µl volume.

Serial dilutions of the primary amplification product were used for the second round PCR with the primers st2biup-3f, 5′-TGGAAATGAAAGGGGTTATTATTTA-3′ (nt 385–409, gi1220362), and st2biup-4r, 5′-AATTCCATTCCATTCCATTC[G/A]ATA-3′ (nt 741–764, gi1220362), in a 100-µl reaction as described for the first round but with the following thermocycler parameters: 95°C for 5 min for one cycle, then 95°C for 1 min, 62°C for 1 min, 72°C for 2 min, for 25 cycles, followed by 72°C for 7 min for one cycle. It was found that 2 µl of a 200-fold dilution of the first round PCR product gave the best amplification in the second round PCR, and this dilution was used for all samples.

Cloning and sequencing of the PCR products

The PCR products were purified by the Qiaquick PCR purification kit and ligated in vector pCR2.1 (TA-cloning Kit, Invitrogen, Carlsbad, Calif.), and the ligation product was electroporated into DH10B *Escherichia coli* (GIBCO/BRL, Life Technologies, Grand Island, N.Y.) by standard protocols. Generally, 30%–70% of the randomly selected white colonies had satellite 2-derived inserts. The inserts were amplified by vector primers M13-F3 and M13-R4 (Hansen et al. 1999). For both modified and unmodified DNA, inserts of 100 bp to 1.3 kb were obtained, and sequences of the inserts matching the expected size of ~380 bp were determined directly from the PCR product as described previously (Hansen et al. 1999).

Use of a 50:50 mixture of methylated and unmethylated control DNAs revealed no bias in the conversion, amplification, or recovery of clones derived from methylated versus unmethylated sequences. Although non-CpG methylation was observed at much higher frequencies at satellite 2 (5% of conserved cytosine positions) than at single copy loci (Hansen et al. 2000; Stöger et al. 1997), it appeared to result from nonconversion in most cases, as it occurred in runs and with a similar frequency to clones derived from unmethylated control DNAs. Clones with 10 or more unconverted non-CpG cytosines were found in both normal and ICF samples (10 of 387 clones) and were removed from methylation analyses.

Determination of the frequency of methylated CpG dinucleotides at the conserved sites

Sequences were determined for 41 independent clones of satellite 2 amplimers from unmodified genomic DNA (DDBJ/EMBL/GenBank accession nos. AF361703–AF361743) and were aligned by using the Multiple Alignment Program (MAP; http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html). Boxshade (http://www.ch.embnet.org/software/BOX_form.html) was then used to create a consensus sequence for satellite 2 in unmodified genomic DNA. To determine the methylation patterns at conserved CpG sites for each DNA sample, bisulfite-modified sequences were aligned to the consensus sequence that had been converted by word processing (search and replace) to the methylated bisulfite-converted equivalent. The number of CpGs at conserved positions was compared with the total number of potentially methylateable sites at these positions (CpG + TpG) and this frequency was expressed as a percentage.

Karyotyping and replication timing

Lymphoblast cultures were routinely cultivated in RPMI plus 15% heat-inactivated fetal bovine serum. For the preparation of metaphases for analyses of centromeric decondensation, cultures were harvested, washed, and set up at cell concentrations of $1–2\times10^5$/ml in folate-deficient RPMI plus 15% fetal bovine serum. Some 36 h later, fluorodeoxyuridine (FudR) was added to a concentration of 2×10^{-4} mM, and at 48 h, cultures were harvested by a standard protocol widely used in cytogenetic laboratories for lymphocyte cultures (Brown and Lawce 1997). Some preparations were G-banded as described by Seabright (1971).

Replication time analysis

Replication time was determined by analysis of newly replicated DNA, labeled with bromodeoxyuridine and derived from cells sorted into specific cell cycle stages by flow cytometry as previously described (Hansen et al. 1993, 1996, 1997). The BrdU-labeled DNA was isolated with an anti-BrdU antibody, and these newly-replicated DNA fractions were then analyzed for the presence of specific loci by semi-quantitative PCR. The sequence tagged site primers used included sat2–384U:781L for satellite 2 sequences and mto8000–8024:mto8369–8345 (Hansen et al. 1997) for mitochondrial sequences. Amplification conditions were similar to those described previously (Hansen et al. 1997), with an annealing temperature of 60°C for sat2–384U:781L.

Results

Southern analysis of satellite 2 hypomethylation in ICF cells

Satellite 2 DNA was one of the early repetitive sequences discovered in humans (Prosser et al. 1986). Megabase-sized pericentromeric regions of chromosomes 1 and 16 contain most of these sequences; the higher order satellite 2 repeat units are larger than 1 kb and are composed of basic repeats of 23–26 bp that contain the common GGAAT and ATC sequences, and a less abundant sequence that is methylateable, viz., CGAAT (Jeanpierre 1994). Although the GC content (~40%) of these sequences is lower than that found in CpG islands (>50%), the proportion of CpGs present is similar to the statistically expected value based on the C and G proportions, a characteristic usually restricted to CpG islands.

To characterize initially the methylation status of satellite 2 sequences in cells from normal individuals and ICF patients, we digested DNAs with methylation-sensitive restriction enzymes and analyzed the products by Southern blot hybridization. As shown in Fig. 1, normal DNAs and those from unaffected parents of ICF patients show a

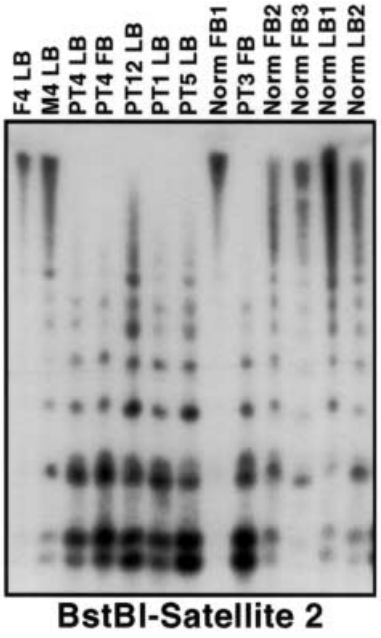

Fig. 1 Southern blot analysis of satellite 2 methylation in ICF and other cell types. Chromosomal DNA was digested by *Bst*BI and hybridized with a satellite 2 PCR probe (sat2–384U:781L). Patients *PT1*, *PT3*, *PT4*, and *PT5* are ICF cases with known mutations in the *DNMT3B* gene. *F4* and *M4* are the father and mother, respectively, of *PT4*. *PT12* is an ICF case with no apparent mutation in the *DNMT3B* gene. *Norm* Cells from phenotypically normal individuals, *LB* lymphoblasts, *FB* fibroblasts

considerable amount of resistance to *Bst*BI digestion, indicating hypermethylation of TTCGAA sequences, whereas DNAs from the ICF patients are extensively digested, indicating hypomethylation. The PT12 ICF cells appeared to differ somewhat from the cells of other ICF patients because, although their satellite 2 DNA was markedly more digested by *Bst*BI than DNA from normal cells, it was not as heavily digested as other ICF DNAs and was therefore somewhat more methylated. This finding is of interest because PT12, although clearly possessing the phenotypic and cytogenetic characteristics of ICF syndrome (Wijmenga et al. 2000; G. Gimelli, unpublished), differed from the other patients in that no mutations were found in the *DNMT3B* gene (Wijmenga et al. 2000). All the other ICF cases were extensively demethylated at satellite 2 and had known mutations in the *DNMT3B* gene that would be expected to affect DNA methyltransferase activity because of their nature and location (Hansen et al. 1999; Wijmenga et al. 2000).

Bisulfite conversion analysis of satellite 2 hypomethylation in ICF cells

Although it is apparent from the results shown in Fig. 1 that satellite 2 DNA sequences in our ICF patients are hypomethylated, it is not clear what the levels and patterns of methylation are in either normal or ICF samples. To determine these levels and patterns, we have developed methods for analyzing the detailed cytosine methylation patterns at satellite 2 by using bisulfite conversion (Frommer et al. 1992). This method involves the sequencing of clones derived from genomic DNA whose unmethylated, but not methylated, cytosines are converted to uracils by bisulfite-catalyzed deamination. Following conversion, sequences corresponding to either the upper or lower DNA strand are amplified by PCR; the products are then size-fractionated and cloned, and the sequences of individual clones determined. Unmethylated cytosines are replaced by thymines in the PCR reaction, whereas methylated cytosines remain unaltered. Thus, this method allows the determination of strand-specific cytosine methylation patterns at specific loci on a single chromosome, and the frequency of those patterns in a cell population.

The satellite 2 region amplified for analysis is about 400 bp and is part of an internal repeat that is present in three copies in the higher-order 1.3-kb repeat. An example of the bisulfite methylation analysis is shown in Fig. 2, where the top line of sequence is a portion of an unmodified strand from the reference sequence gi1220362. Below this line is the sequence predicted to result from the bisulfite conversion method if non-CpG cytosines (indicated by asterisks above) are unmethylated and therefore converted to thymines, and CpG cytosines are methylated and unaltered. In the converted strands from genomic DNA, many unconverted CpGs were found in normal DNA (FB4) that corresponded to CpGs in the reference sequence, whereas the converted strands from ICF (PT4 FB) and sperm (A47) DNAs had mostly TpGs at these sites.

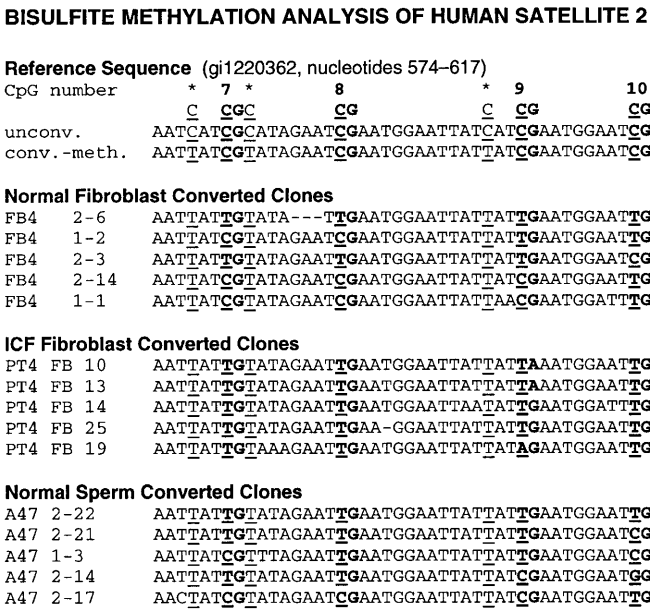

Fig. 2 Examples of bisulfite methylation analysis of human satellite 2. A reference satellite 2 sequence, gi1220362, is aligned with the PCR-amplified sequences derived from the upper strand of the satellite 2 repeat region after bisulfite modification. The sequences correspond to a representative portion of the analyzed region of the repeat of ~330 bp. The *top sequence line* contains bases 574–617 of the unmodified reference sequence (*unconv.*), with *numbers above* corresponding to conserved CpG sites in the region analyzed. *Asterisks* Non-CpG cytosines in the unmodified consensus sequence. The next *line* of sequence is that which would result from bisulfite conversion of the reference sequence if all the CpG sites were methylated (*conv.-meth.*). Sequences from bisulfite-converted samples are shown from five independent clones for each sample, including those derived from *FB4* normal fibroblasts, *PT4* ICF fibroblasts, and *A47* normal sperm. Positions corresponding to conserved cytosines in the unmodified consensus sequence are *underlined*, and positions corresponding to conserved CpG sites are in *bold*

Although the presence of CpGs in the resulting clone indicate sites that were methylated in the unconverted DNA (mCpGs), TpGs could either represent a preexisting TpG variant or an unmethylated CpG that was converted to TpG following bisulfite modification. At least 19 clones were analyzed for each DNA sample.

When the satellite 2 clones are analyzed solely for their mCpG content, the ICF cases are clearly seen to be significantly less methylated than normal ($P<0.01$). A simple graph (Fig. 3) depicting the mean number of mCpGs/clone with the standard deviation shows that the ICF cells with known *DNMT3B* mutations have about one-third the methylation levels of normal cells, with no overlap between the two groups. The values for normal sperm are similar to those for the *DNMT3B* mutant fibroblasts and lymphoblasts. HD14A1 and 14IC3, two hybrid cell cultures resulting from fusion between PT4 lymphoblasts and CHO cells (Hansen et al. 1999), show values significantly higher than that of PT4 lymphoblasts ($P<0.05$), indicating de novo methylation of the satellite 2 repeat had

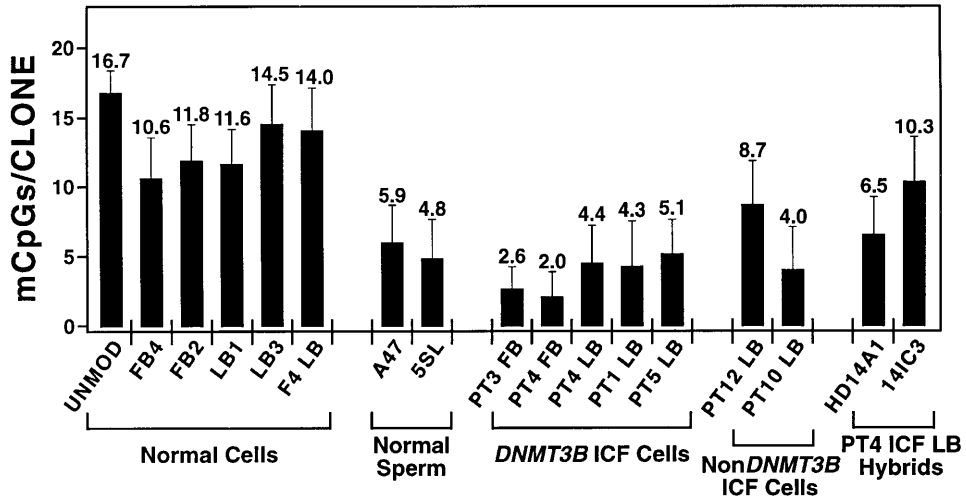

Fig. 3 Average satellite 2 methylation levels in various normal and ICF cells. The average CpGs/clone for bisulfite-modified and unmodified DNAs are depicted in bar graph form and are also shown numerically at the *top* of the *bars*. The value for *UNMOD* clones should represent the maximal value for average CpG methylation because they were derived from the unmodified genomic DNA of FB4. All other clones were derived from bisulfite-modified DNAs. *FB4, FB3* Normal female fibroblasts, *LB1, LB3* normal male lymphoblasts, *F4 LB* lymphoblasts from father of PT4. ICF cases include *PT1, PT2, PT3, PT4, PT5, PT10,* and *PT12. HD14A1, 14IC3* Human-hamster hybrids produced by fusion of PT4 ICF lymphoblasts with Chinese hamster ovary cells, *A47, 5SL* normal sperm cells

occurred, presumably by the action of the hamster homolog of MT3β.

PT12, an ICF patient for whom no *DNMT3B* mutation has been found, has a methylation level much closer to that of normal individuals than to those of the other ICF patients. Although PT10 is another ICF patient apparently lacking *DNMT3B* mutations, methylation levels in PT10 lymphoblasts are as low as in cells with known mutations. The PT5 ICF patient is a compound heterozygote in whom one allele is a missense mutation in the catalytic domain and the other should be a null allele because of an early stop mutation (Wijmenga et al. 2000), although satellite methylation levels in PT5 lymphoblasts are similar to, or even higher than, those in PT1 or PT4 ICF lymphoblasts that have only missense mutations. Our analysis of *DNMT3B* transcription in PT5 lymphoblasts, however, suggests that the predicted null allele is apparently subject to nonsense-associated altered splicing (Maquat 2001), in which the stop mutation is by-passed. Reverse transcription/PCR (RT-PCR) analysis under standard PCR conditions reveals that the catalytic domain is expressed by both alleles as seen by the heterozyous expression of a common polymorphism (data not shown). RT-PCR under "long" PCR conditions, however, results in normally spliced, full-length transcripts that are only derived from the missense allele (data not shown). Much of the satellite 2 methylation in PT5 cells might have resulted, therefore, from the expression of the unmutated catalytic domain of the alternatively spliced "null" allele that has spliced out this mutation. This conclusion has yet to be verified, however, because *DNMT3B* expression levels are not strong enough for us to obtain good Northern blot information on variant transcript size, and our initial attempts at finding abnormal splice junctions by RT-PCR scanning have been unsuccessful.

Although there is no overlap in the mean values for mCpGs/clone between normal and *DNMT3B* ICF cells (Fig. 3), there is considerable overlap in the distributions of individual clone values among these groups (Fig. 4B, C). For example, about 5% of the clones from normal cultures have methylation values near the mean of the *DNMT3B* mutant ICF cases and a similar percentage of mutant ICF clones have values near the normal mean. Normal sperm clones have a distribution similar to *DNMT3B* ICF lymphoblasts (Fig. 4A). Because the repeats are not perfect, one possible explanation for the marked variation in the frequency of methylated sites per clone is that it reflects the variation in frequency of CpGs per clone prior to modification by bisulfite treatment. Such variation in CpG frequency was not found, however, among 41 unmodified clones derived from normal (FB4) DNA where the CpGs per clone was always greater than 13 (Fig. 4A) compared with as low as four in modified clones (Fig. 4B). The marked variation in CpGs per clone in modified sequences thus reflects a heterogeneity between repeats in the level of methylated CpG.

The PT12 non-*DNMT3B* ICF distribution overlapped considerably with that of normal cells, whereas the PT10 non-*DNMT3B* ICF distribution was similar to the *DNMT3B* ICF distributions (Fig. 4D). The de novo methylation of satellites from PT4 ICF lymphoblasts following somatic cell fusion to hamster cells (Fig. 3) results from an increased number of clones with higher methylation values, although the distribution remains wide (Fig. 4E).

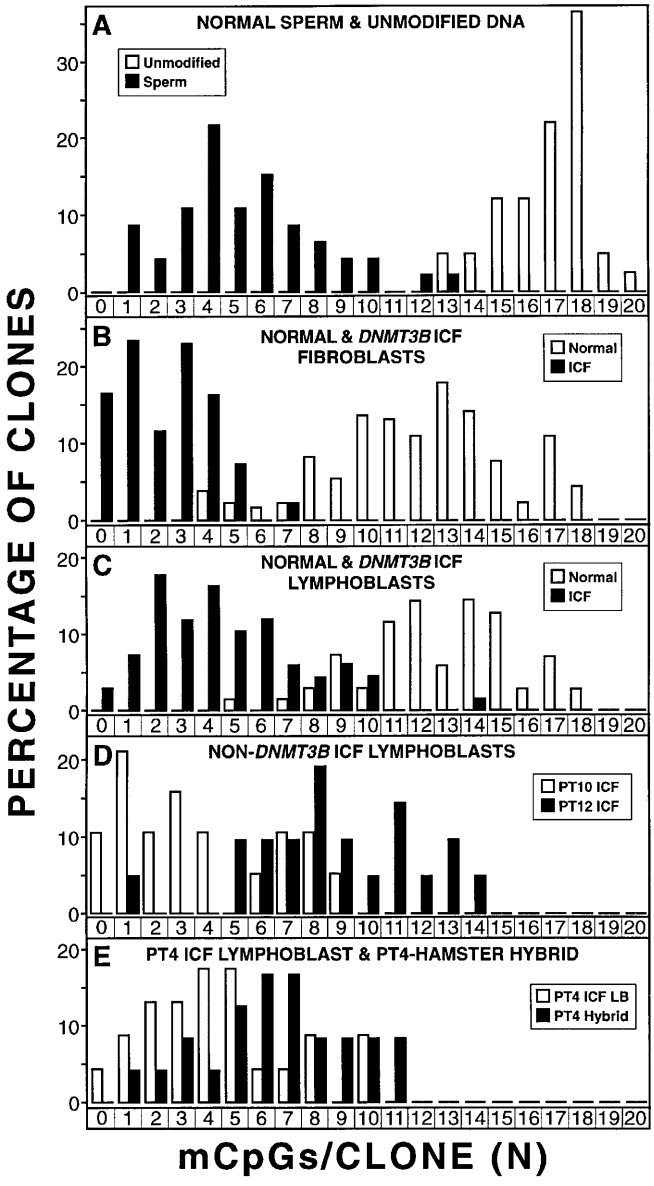

Fig. 4A–E Frequency of *n* sites methylated/clone among clones derived from various normal and ICF cells (*n*=0–20). The percentage of total clones with a given number of mCpGs was determined for each DNA sample. Shown are the average values for fibroblasts and lymphoblasts from normal and *DNMT3B* ICF cases, and those from sperm. Values for non-*DNMT3B* ICF lymphoblasts and an ICF-hamster hybrid clone are also shown. Clones from unmodified DNA (FB4) were also analyzed for CpG content as a reference distribution of methylateable (CpG) sites. **A** Normal sperm (5SL, A47) and bisulfite-unmodified DNA. **B** Normal fibroblasts (FB2, FB4) and ICF fibroblasts (PT3, PT4). **C** Normal lymphoblasts (F4, LB1, LB3) and *DNMT3B* ICF lymphoblasts (PT1,

Methylation site specificity in satellite 2 sequences

Because the bisulfite methylation analysis allows for the determination of all methylated CpG sites, we examined whether there is any tendency for some sites to be preferentially more (or less) methylated than other sites. One problem with such an analysis is the uncertainty, as discussed above, in the assignment of bisulfite-converted TpGs as deriving from either unmethylated CpGs or from preexisting TpGs. Because this uncertainty results from sequence heterogeneity in satellite 2 repeats, we attempted to determine the common TpG and CpG distributions in the unmodified satellite sequences by amplifying, cloning, and sequencing the same region of ~330 bp in unmodified DNA from a normal fibroblast culture (FB4). A consensus sequence containing 18 conserved CpG sites was produced from 41 clones (Fig. 5); 35 clones have more than 90% identity to the consensus, five clones have between 73% and 88% identity, and one clone has 57% identity. The clones apparently derive from different repeat loci, because only two of the 41 sequences are fully identical. Such variability in satellite 2 repeats has been reported earlier (Jeanpierre 1994). Also similar to previous reports, the disassembly of our consensus sequence into internal repeats yielded fifteen 22-mers to 26-mers that are highly similar to each other (Fig. 5).

In these unmodified genomic sequences, the average number of CpGs per clone is 16.7±1.6 S.D. (Fig. 3). Because of the individuality of the repeats, the frequency of CpGs at conserved sites varies considerably. For example, eight of the 41 clones had frequencies of greater then 90% CpG at these sites (one had 100%), and three clones had frequencies of less than 70%. As expected from the increased rate of mutation at methylated CpGs, most of the non-CpGs observed at these sites were TpGs and CpAs. The relative TpG and CpG frequencies at conserved sites in the unmodified sequences were used to help evaluate methylation data in bisulfite-converted sequences (Fig. 6A). The percentages of CpGs at conserved sites relative to CpG + TpG (sites that could produce TpG in bisulfite-converted DNA) are mostly greater than 75% (except for site 1, which is 56%). Twelve of the18 sites are more than 95% CpG among all the unmodified clones.

In bisulfite-modified DNA from normal fibroblast cultures (FB4 and FB2), the mCpG frequency at a given site (Fig. 6B) indicates that no site is either always methylated or always unmethylated. A trimodal methylation pattern is apparent, however, with consistent low points at sites 2, 7, 13, 17, and 18 (site 1 is excluded because of high TpG/CpG degeneracy in unmodified DNA). This pattern cannot be accounted for by sequence variation, because it is also seen when the methylation levels are adjusted to account for variable TpG levels in the unmodified consensus sites (Fig. 6 and data not shown). This trimodal pattern is also evident in the other cells analyzed, including normal lymphoblasts, ICF lymphoblasts and fibroblasts, normal sperm, and ICF-derived hybrids (Fig. 6C, D, and data not shown). The trimodal pattern is even seen in the two ICF cases without *DNMT3B* mutations, PT10 and PT12 (data not shown).

Examination of the sequences surrounding the low methylation sites did not reveal any obvious patterns that were different from high methylation sites. Such analyses included examinations, with different window sizes, of various single and dinucleotide base compositions. The

Fig. 5 A satellite 2 consensus sequence for the region analyzed for cytosine methylation. ClustalW 1.8 (http://www.ebi.ac.uk/clustalw/) was used to align cloned sequences derived from PCR-amplified genomic DNA (bisulfite-unmodified), and a consensus sequence was obtained. This sequence was disassembled into 21-bp to 26-bp internal repeats and compared with a 26-bp consensus internal repeat. The starting nucleotide position of each internal repeat is shown *left*. Common CpG sites analyzed for methylation are numbered *above*; *underlined* CpG sites are frequently unmethylated in both normal and ICF DNAs. A *black spot* is shown above nucleotides that lie in the internal repeat and that differ from the consensus nucleotide (also in *bold*); *dashes* denote gaps in the alignment. A 16-bp motif, 5′-ATCGAATGGAATCGAA-3′, is *underlined* in the consensus internal repeat; this sequence is frequently found at sites with higher methylation levels, as ascertained by MEME analysis (http://meme.sdsc.edu/meme/website/meme.html)

only obvious correlation was the poly-A region (A_7) that is unique to site 13 (Fig. 5), the site that is usually the lowest in mCpG. Deconstruction of the unmodified consensus sequence into internal repeats also did not reveal any consistent similarities among the low methylation sites that were not found at other sites (Fig. 5). A multiple expectation-maximization motif elicitation analysis was performed separately on groups of high and low methylation sites (http://meme.sdsc.edu/meme/website/meme.html), and the motif 5′-ATCGAATGGAATCGAA-3′ was found to be common to the high sites (Fig. 5). An additional feature of the trimodal methylation pattern is that it derives from a periodic low in methylation at about every 100 bp, i.e., within every fourth internal repeat (Fig. 5).

The satellite 2 methylation pattern in the human-hamster hybrid HD14A1 (Fig. 6D) is of interest with respect to the question of de novo methylation site preferences, because these satellites were derived from hypomethylated PT4 ICF lymphoblasts and became more methylated in the hamster background. Assuming that this de novo methylation is attributable to the action of the hamster MT3β, the preferential methylation of sites 1–3 and site 12 in these cells supports the idea that there can be regional and/or site preference for MT3β de novo methylation activity within these sequences.

Centromeric decondensation frequencies and satellite 2 methylation in ICF cells

Under normal culture conditions, we found that our ICF lymphoblastoid lines exhibited only a low frequency (~1% of cells) of centromeric decondensation of chromosome 1 (data not shown). As in previous studies of ICF fibroblasts (Maraschio et al. 1989), growing the ICF lymphoblasts in folate-deficient medium plus FudR gave higher yields of pericentromeric decondensation but no rearrangements (11%–69%). These data, detailed in Table 1, suggest that pericentromeric decondensation levels among ICF cases are not strictly correlated with satellite 2 methylation levels. The PT12 example provides good support for this conclusion because the satellite 2 methylation level is closer to that of normals than the levels in ICF cases with known mutations in *DNMT3B* (PT1, PT3, PT4, and PT5), and yet the mean level of pericentromeric decondensation is similar to that of cases such as PT4 that

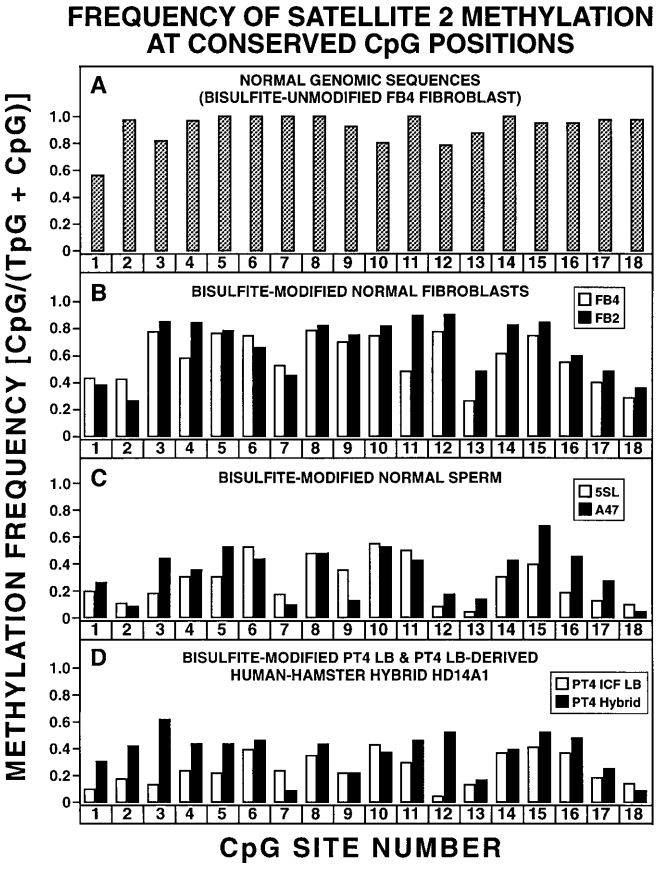

Fig. 6A–D Satellite 2 methylation frequencies at conserved CpG sites among normal and ICF cells. For each of 18 conserved positions, the data are presented as the fraction of clones with CpG versus the total of those with either CpG or TpG at that position. **A** Unmodified normal fibroblast, FB4, which represents an estimate of the maximal methylation at each position. **B** Bisulfite-modified normal fibroblasts (*FB4*, *FB2*). **C** Bisulfite-modified normal sperm cells (*A47, 5SL*). **D** Bisulfite-modified *PT4* ICF lymphoblasts and derived CHO hamster hybrid *HD14A1*

Table 1 Centromeric decondensation of chromosome 1 in lymphoblasts. Centromeric decondensation was only observed on chromosome 1 and never on both homologs; all cells were treated with FudR

Sample	Total no. of cells	No. cells with a decondensed chromosome 1	% of cells with a decondensed chromosome 1
Control	53	0	0
PT1	73	15	20
PT4	293	33	11
PT5	26	18	69
PT12	196	26	13

have much lower levels of satellite 2 methylation (13% versus 11%). Although these cases differ in their genetic bases, the common ICF cytogenetic abnormalities in lymphoid cells apparently derive from tissue-specific factors whose function can be altered by partial hypomethylation of satellites to some threshold level.

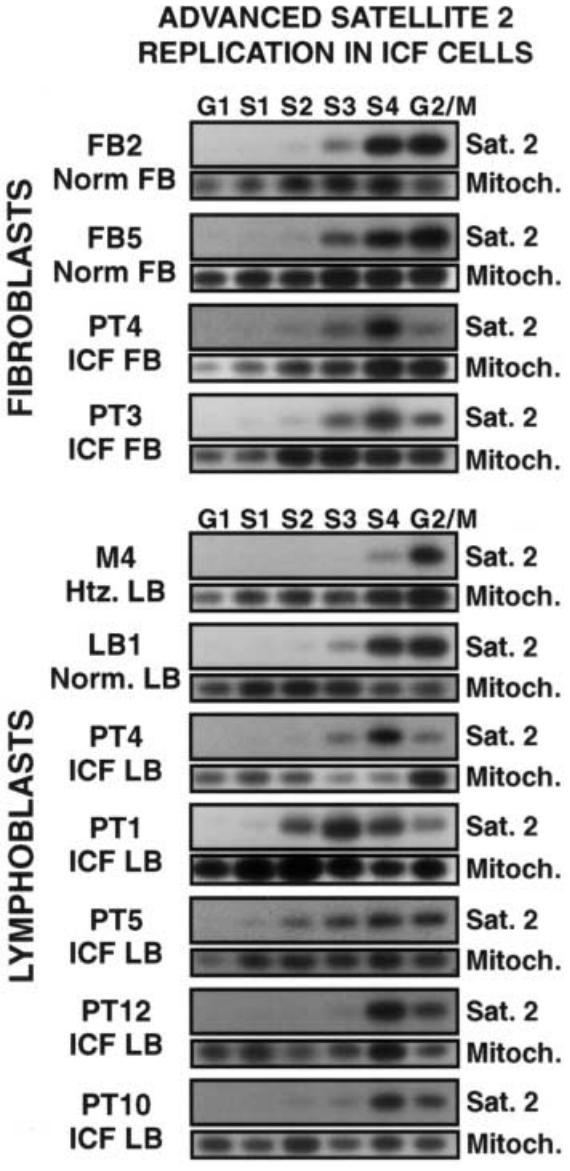

Fig. 7 Replication timing patterns for satellite 2 sequences in normal and ICF cells. Replication of satellite 2 sequences was determined by exponential amplification of newly replicated DNA (384U:781L primer set) in each of six cell cycle fractions (*G1*, *S1–S4*, *G2/M*). These sequences replicate quite late (*S4, G2*) in normal cells (*FB2* and *FB5* normal fibroblasts, *LB1* normal lymphoblasts) and in *M4* ICF carrier lymphoblasts (mother of *PT4*), but their replication is advanced (*S3*, *S4*) in ICF cases (*PT4* ICF fibroblasts and lymphoblasts, and *PT1*, *PT5*, *PT12*, and *PT10* ICF lymphoblasts). Mitochondrial DNA replication patterns (*Mitoch.*) are also shown as these sequences replicate during all portions of the cell cycle (Bogenhagen and Clayton 1977), thereby serving as an internal control for normalization of the cell cycle fractions (Hansen et al. 1997; Strehl et al. 1997)

Satellite 2 replication time in ICF cells

We considered it possible that alteration of normal replication timing for the satellite regions might be involved in

the production of cytogenetic abnormalities in ICF cells. We examined satellite 2 replication time with a standard method involving the assay of specific sequences in newly replicated DNA, labeled with bromodeoxyuridine, from cells that were fractionated by flow cytometry into different portions of the cell cycle (Hansen et al. 1993, 1997).

Examples of the replication profiles for satellite 2 in normal and ICF cultures are shown in Fig. 7. Mitochondrial DNA replication patterns are also shown; these sequences replicate throughout the cell cycle (Bogenhagen and Clayton 1977) and can thereby serve as an internal control for the normalization of cell cycle fractions (Hansen et al. 1997; Strehl et al. 1997). Similar to the inactive X and other heterochromatic regions, satellite 2 DNA normally replicates late in the cell cycle (S4 and G2/M). In ICF cells, satellite 2 replication is advanced compared with that in normal cells (primarily S4 in both lymphoblasts and fibroblasts). The degree of advanced replication appears to correlate only loosely with the level of satellite hypomethylation. For example, the partially methylated PT12 LB satellite 2 sequences replicate later (S4 and G2/M) than the minimally methylated PT1 LB sequences (S3 and S4) and at about the same time as the minimally methylated PT10 LB sequences.

Discussion

We have shown that ICF cells with known mutations in the *DNMT3B* gene retain about one-third the normal level of satellite 2 methylation, which is comparable with the level found in normal sperm. That satellite DNA is hypomethylated in gametes has been known for some time (Pages and Roizes 1982; Ponzetto-Zimmerman and Wolgemuth 1984; Sanford et al. 1984; Sturm and Taylor 1981; Zhang et al. 1987); although satellite 2 methylation has not been reported for human oocytes, sperm and oocyte methylation levels of mouse repeats are comparable (Sanford et al. 1984), and human satellites are likely to be similar. It is possible, therefore, that the satellite 2 methylation levels in the ICF cases represent residual methylation contributed by the gametes to the zygote and that these ICF cases have null mutations. We suggest that this explanation is unlikely, as outlined below.

First, a *Dnmt3b* knockout in the mouse is an embryonic lethal (Okano et al. 1999), whereas the *DNMT3B* mutations in humans have less severe effects. Second, the "genome-wide" demethylation known to occur in the early embryo (Kafri et al. 1992) probably includes the pericentromeric satellite sequences. If true, this demethylation in ICF is followed by a partial remethylation by the mutant MT3β (or possibly some other methyltransferase). Third, the existence of residual activity is supported by the marked variability in mCpG/clone in ICF (Fig. 4), with over 5% of satellite 2 clones exhibiting normal methylation values. As pointed out earlier, even though there is no overlap in the mean number of mCpGs/clone between normal and *DNMT3B* ICF cells, there is considerable overlap at the level of individual clones. This methylation variability may reflect a limited time period in early development when MT3β de novo activity is highest, thus resulting in a large stochastic variability in methylated CpGs per satellite sequence.

Although the bisulfite methodology is not currently capable of examining large amplimers, it would be of interest to investigate the larger context of the patterns that we found in order to determine whether methylation clustering extends over large satellite regions. Given that there are probably several thousand higher order repeats per chromosomal region, does this mean that every chromosomal satellite 2 region normally has a few percent of its repeats with very low levels of methylated CpGs or are there only a few percent of chromosomes within the cell population that have low levels of methylated CpGs throughout the satellite? The latter possibility has several implications. For example, in normal cells, an ICF-like satellite could lead to ICF-like chromosomal abnormalities, whereas in ICF cells, a normal-like satellite might prevent such abnormalities. The same variability is seen in sperm distributions, in which about 10% of the individual clones have values well into the normal somatic range. The hypomethylation of centromeric DNA in gametes is probably required for synapsis-related decondensation; if a small percentage of individual sperm had somatic levels of methylated satellite 2 DNA, this could interfere with normal meiosis and lead to segregation errors.

The most distinctive feature of ICF syndrome is the centromeric decondensation of chromosomes 1, 9, and 16 and the rearrangements involving these regions, which are present in leukocytes. In other cells, especially fibroblasts, these abnormalities are much reduced in frequency and sometimes can only be seen under special inducing conditions, such as low folate concentrations, the addition of FudR, or the addition of excess thymidine (Maraschio et al. 1989). As both ICF leukocytes and fibroblasts have hypomethylated satellites, the rearrangements specific to leukocytes presumably derive from the differential presence of a factor that either destabilizes or promotes satellite 2 heterochromatin in one cell type or the other. For example, a leukocyte-specific recombinase or nuclease might recognize the satellites when hypomethylated and bring about such rearrangements.

The advanced replication that we have observed is consistent with other observations of hypomethylation-associated advanced replication (Hansen et al. 1996, 2000; Selig et al. 1988). Advanced replication of satellite 2 could be responsible for promoting cytogenetic abnormalities and for potential gene-silencing defects caused by alteration of these regions. ICF decondensation in chromosomes 1, 9, and 16 may reflect a chromatin state that has not progressed to mitotic condensation because the advance in replication time has not been compensated for by a proportional advance in the timing of recondensation signals following replication. The cell cycle kinetics in fi-

broblasts may be more tolerant of such changes than the kinetics in leukocytes.

It is also likely that the normal interphase chromatin structure of these regions is altered in ICF because of hypomethylation. The advanced satellite replication in ICF could be a consequence of greater accessibility to replication factors, and such advances may also hinder the association of heterochromatic proteins that normally assemble onto late-replicating sequences. One intriguing possibility is that the advanced replication time in an otherwise heterochromatic island of pericentromeric satellite may promote CENP-A deposition and neocentromere formation (Ahmad and Henikoff 2001; Lo et al. 2001), thus leading to chromosomal abnormalities.

The processes of mitotic condensation and heterochromatization are probably genetically complex and MT3β, which acts indirectly through methylation, is probably only one of a number of factors that control these processes. Indeed, a *Drosophila* gene discovered almost 20 years ago specifically affects condensation of centromeric heterochromatin (Gatti et al. 1983) and, more recently, mutations in the *Drosophila* replication origin recognition complex have been shown to produce defects in replication timing and chromosome condensation (Loupart et al. 2000). Mutations in such genes could be the basis of the ICF phenotype exhibited by PT10 or PT12, which are apparently normal at *DNMT3B*; alternatively, the defect in these cells could be in another methyltransferase or in a methyltransferase-targeting factor.

Acknowledgements Theresa Canfield, Ping Luo, and Kartik Varadarajan provided excellent technical assistance. Cisca Wijmenga and Cory Weemaes kindly provided ICF patient cultures. We also thank Charles Laird and Lester Goldstein for their helpful suggestions and editorial comments on the manuscript. This work was supported by grants from the National Institutes of Health (HD16659, GM53805, and AG13280).

References

Ahmad K, Henikoff S (2001) Centromeres are specialized replication domains in heterochromatin. J Cell Biol 153:101–110

Bogenhagen D, Clayton DA (1977) Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell 11:719–727

Brown MG, Lawce H (1997) Peripheral blood cytogenetic methods. In: Barch MJ, Spurbeck JL, Knutsen T (eds) The AGT cytogenetics manual, 3rd edn. Raven, New York, pp 107–110

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89: 1827–1831

Gatti M, Smith DA, Baker BS (1983) A gene controlling condensation of heterochromatin in *Drosophila melanogaster*. Science 221:83–85

Gimelli G, Varone P, Pezzolo A, Lerone M, Pistoia V (1993) ICF syndrome with variable expression in sibs. J Med Genet 30: 429–432

Hansen RS, Ellis NA, Gartler SM (1988) Demethylation of specific sites in the 5′ region of the inactive X-linked human phosphoglycerate kinase gene correlates with the appearance of nuclease sensitivity and gene expression. Mol Cell Biol 8:4692–4699

Hansen RS, Canfield TK, Lamb MM, Gartler SM, Laird CD (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell 73:1403–1409

Hansen RS, Canfield TK, Fjeld AD, Gartler SM (1996) Role of late replication timing in the silencing of X-linked genes. Hum Mol Genet 5:1345–1353

Hansen RS, Canfield TK, Fjeld AD, Mumm S, Laird CD, Gartler SM (1997) A variable domain of delayed replication in *FRAXA* fragile X chromosomes: X inactivation-like spread of late replication. Proc Natl Acad Sci USA 94:4587–4592

Hansen RS, Canfield TK, Stanek AM, Keitges EA, Gartler SM (1998) Reactivation of *XIST* in normal fibroblasts and a somatic cell hybrid: abnormal localization of XIST RNA in hybrid cells. Proc Natl Acad Sci USA 95:5133–5138

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CMR, Gartler SM (1999) The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci USA 96:14412–14417

Hansen RS, Stoger R, Wijmenga C, Stanek AM, Canfield TK, Luo P, Matarazzo MR, D'Esposito M, Feil R, Gimelli G, Weemaes CM, Laird CD, Gartler SM (2000) Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum Mol Genet 9:2575–2587

Jeanpierre M (1994) Human satellites 2 and 3. Ann Génét (Paris) 37:163–171

Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, Razin A (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev 6:705–714

Lo AW, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KH (2001) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J 20:2087–2096

Loupart M-L, Krause SA, Heck MS (2000) Aberrant replication timing induces defective chromosome condensation in *Drosophila* ORC2 mutants. Curr Biol 10:1547–1556

Maquat LE (2001) The power of point mutations. Nat Genet 27: 5–6

Maraschio P, Tupler R, Dainotti E, Piantanida M, Cazzola G, Tiepolo L (1989) Differential expression of the ICF (immunodeficiency, centromeric heterochromatin, facial anomalies) mutation in lymphocytes and fibroblasts. J Med Genet 26:452–456

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99:247–257

Pages M, Roizes G (1982) Tissue specificity and organisation of CpG methylation in calf satellite DNA I. Nucleic Acids Res 10:565–576

Ponzetto-Zimmerman C, Wolgemuth DJ (1984) Methylation of satellite sequences in mouse spermatogenic and somatic DNAs. Nucleic Acids Res 12:2807–2822

Prosser J, Frommer M, Paul C, Vincent PC (1986) Sequence relationships of three human satellite DNAs. J Mol Biol 187: 145–155

Sanford J, Forrester L, Chapman V, Chandley A, Hastie N (1984) Methylation patterns of repetitive DNA sequences in germ cells of *Mus musculus*. Nucleic Acids Res 12:2823–2836

Seabright M (1971) A rapid banding technique for human chromosomes. Lancet II:971–972

Selig S, Ariel M, Goitein R, Marcus M, Cedar H (1988) Regulation of mouse satellite DNA replication time. EMBO J 7: 419–426

Stöger R, Kajimura TM, Brown WT, Laird CD (1997) Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene *FMR1*. Hum Mol Genet 6:1791–1801

Strehl S, LaSalle JM, Lalande M (1997) High-resolution analysis of DNA replication domain organization across an R/G-band boundary. Mol Cell Biol 17:6157–6166

Sturm KS, Taylor JH (1981) Distribution of 5-methylcytosine in the DNA of somatic and germline cells from bovine tissues. Nucleic Acids Res 9:4537–4546

Wijmenga C, Hansen RS, Gimelli G, Bjorck EJ, Davies G, Valentine D, Belohradsky BH, Dongen JJ van, Smeets DFCM, Heuvel LPWJ van den, Luyten JAFM, Strengman E, Weemaes C, Pearson PL (2000) Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Hum Mutat 16:509–517

Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402:187–191

Zhang XY, Loflin PT, Gehrke CW, Andrews PA, Ehrlich M (1987) Hypermethylation of human DNA sequences in embryonal carcinoma cells and somatic tissues but not in sperm. Nucleic Acids Res 15:9429–9449