Human Molecular Genetics, 2008, Vol. 17, No. 5 ***690–709***
doi:10.1093/hmg/ddm341
Advance Access published on November 27, 2007

DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function

Bilian Jin[1], Qian Tao[2,3], Jinrong Peng[4], Hui Meng Soo[4], Wei Wu[4], Jianming Ying[2,3], C. Robert Fields[1], Amber L. Delmas[1], Xuefeng Liu[5], Jingxin Qiu[6] and Keith D. Robertson[1,*]

[1]Department of Biochemistry and Molecular Biology, UF Shands Cancer Center Program in Cancer Genetics, Epigenetics, and Tumor Virology, University of Florida, PO Box 100245, Gainesville, FL 32610, USA, [2]Cancer Epigenetics Laboratory, State Key Laboratory in Oncology in South China, Sir YK Pao Center for Cancer, Department of Clinical Oncology, Hong Kong Cancer Institute and Li Ka Shing Institute of Health Sciences, Chinese University of Hong Kong, Hong Kong, [3]Johns Hopkins Singapore, Biopolis, Singapore, [4]Laboratory of Functional Genomics, Institute of Molecular and Cell Biology, Proteos, Singapore 138673, [5]Department of Internal Medicine, Wayne State University, Detroit, MI 48201 USA and [6]Department of Pathology, University of Florida, PO Box 100245, Gainesville, FL 32610, USA

Received September 7, 2007; Revised and Accepted November 16, 2007

The data discussed in this publication have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE9499.

Genome-wide DNA methylation patterns are established and maintained by the coordinated action of three DNA methyltransferases (DNMTs), DNMT1, DNMT3A and DNMT3B. DNMT3B hypomorphic germline mutations are responsible for two-thirds of immunodeficiency, centromere instability, facial anomalies (ICF) syndrome cases, a rare recessive disease characterized by immune defects, instability of pericentromeric satellite 2-containing heterochromatin, facial abnormalities and mental retardation. The molecular defects in transcription, DNA methylation and chromatin structure in ICF cells remain relatively uncharacterized. In the present study, we used global expression profiling to elucidate the role of DNMT3B in these processes using cell lines derived from ICF syndrome and normal individuals. We show that there are significant changes in the expression of genes critical for immune function, development and neurogenesis that are highly relevant to the ICF phenotype. Approximately half the upregulated genes we analyzed were marked with low-level DNA methylation in normal cells that was lost in ICF cells, concomitant with loss of repressive histone modifications, particularly H3K27 trimethylation, and gains in transcriptionally active H3K9 acetylation and H3K4 trimethylation marks. In addition, we consistently observed loss of binding of the SUZ12 component of the PRC2 polycomb repression complex and DNMT3B to derepressed genes, including a number of homeobox genes critical for immune system, brain and craniofacial development. We also observed altered global levels of certain histone modifications in ICF cells, particularly ubiquitinated H2AK119. Therefore, this study provides important new insights into the role of DNMT3B in modulating gene expression and chromatin structure and reveals new connections between DNMT3B and polycomb-mediated repression.

*To whom correspondence should be addressed at: Department of Biochemistry and Molecular Biology, University of Florida, College of Medicine Box 100245, 1600 S.W. Archer Rd. Gainesville, FL 32610 USA. Tel: +1 3523921810; Fax: +1 3523922953; Email: keithr@ufl.edu

INTRODUCTION

Methylation at the C-5 position of cytosine within CG dinucleotides is an epigenetic modification critical for normal development, differentiation, gene regulation and control of chromatin structure in mammalian cells (1). Genome-wide DNA methylation patterns are established during embryogenesis and faithfully propagated during cell division by the combined action of three enzymatically active DNA methyltransferases (DNMTs): DNMT1, DNMT3A and DNMT3B. DNMT1 is specialized to carry out most of the maintenance methylation following DNA replication, whereas DNMT3A and DNMT3B are responsible for *de novo* methylation during embryogenesis and germ cell development (2). Mouse knockouts have revealed that *Dnmt1* and *Dnmt3b* are essential for development—embryos deficient for these proteins do not survive to birth. *Dnmt3a* knockout mice are viable but die within a few weeks after birth (3,4). The loss of normal DNA methylation patterns in somatic cells results in impaired cell growth control and is a major and early contributor to tumorigenesis by mediating silencing of critical tumor suppressor genes (5). The loss of normal DNA methylation patterns during human development also contribute to a number of human genetic diseases, such as Prader–Willi, Angelman and Fragile X syndrome (6). Only one human genetic disease is currently known to arise from germline mutations within a DNMT gene—namely, the immunodeficiency, centromere instability, facial anomalies (ICF) syndrome (OMIM #242860) (7). Mutations within the *DNMT3B* gene are responsible for the majority of ICF cases (4,8,9).

Immunodeficiency, centromere instability, facial anomalies syndrome is an extremely rare autosomal recessive disease characterized by profound immunodeficiency because of the absence or significant reduction in at least two immunoglobulin isotypes. Other, more variable, features of ICF syndrome include impaired cellular immunity, mental retardation, intestinal dysfunction, psychomotor impairment, unusual facial features (hypertelorism, flat nasal bridge and macroglossia) and delayed developmental milestones (10,11). The other invariant feature of ICF syndrome is a marked elongation of juxtacentromeric heterochromatin on chromosomes 1 (1qh) and 16 (16qh), and to a lesser extent chromosome 9 (9qh). This abnormality, which often includes the formation of multiradials involving one or more of the decondensed chromosomes, translocations and telomeric associations, is observed almost exclusively in ICF B cells or lymphoblastoid cell lines (LCLs) (10). The molecular defect in 60–70% of ICF patients is mutation of the *DNMT3B* gene on chromosome 20q11.2 (4,8,9). One study suggested that ICF patients lacking *DNMT3B* mutations have a distinctive set of DNA methylation defects, pointing to the existence of distinct ICF subtypes and the possible involvement of another gene (12). Mutations in *DNMT3B* usually target the catalytic domain, but may not result in complete loss of enzyme activity (9). Complete loss-of-function of DNMT3B is likely embryonic lethal in humans, as it is in mice (4). The clinical variability among ICF patients may therefore be due, in part, to residual levels of functional DNMT3B.

One of the invariant features of ICF syndrome, instability of pericentromeric heterochromatin, is likely because of the marked hypomethylation of these sequences in ICF cells. DNA hypomethylation is region specific and affects mainly satellite 2 and 3 repeats, the inactive X chromosome, cancer-testis (C-T) genes and other sporadic repeats (D4Z4 and NBL2) (13–15), suggesting that these sequences are bonafide targets of DNMT3B. Satellite hypomethylation is observed in all cells of ICF patients but chromosomal instability is associated only with lymphocytes, suggesting that proper DNMT3B function is especially important for this cell type. Pericentromeric instability may also be related to the finding that DNMT3B localizes to mitotic chromosomes throughout mitosis and interacts with components of the chromosome condensation machinery (16). Satellite 2 and 3 hypomethylation is also very common in cancer cells (17). Although no elevated cancer incidence has been reported in ICF patients, their small number and short lifespan make it possible that such an increase has gone undetected.

In the present study, we had two main goals (1): to better understand how mutations in *DNMT3B* lead to the phenotypic effects characteristic of ICF syndrome patients and (2) to use ICF syndrome cells as a model system for understanding how DNMT3B is targeted throughout the genome and the interplay between DNMT3B and other epigenetic modifications. In order to address these goals, we made use of expression profiling of nearly the entire human genome in LCLs derived from normal and ICF syndrome individuals to identify genes whose expression is altered by *DNMT3B* mutation. This yielded a list of nearly 800 genes that were up- or downregulated in ICF cells relative to normal cells. Interestingly, a large number of these genes were involved in immune system function, signal transduction and development, and these putative DNMT3B-regulated genes provide new insights into the molecular defects underlying the ICF phenotype. Detailed DNA methylation mapping revealed that a fraction of deregulated genes are methylated at low levels in normal LCLs and lose this DNA methylation in ICF cells. Chromatin immunoprecipitation experiments showed that histone modification patterns present at affected promoters were dramatically altered. The loss of histone H3 lysine 27 trimethylation was the most consistent alteration in ICF syndrome relative to normal LCLs and was paralleled by reduced binding of DNMT3B and polycomb protein SUZ12. Taken together, our analysis provides significant new information of the molecular defects that contribute to ICF syndrome and also shows that the loss of DNMT3B function leads to hypomethylation of non-repetitive regions of the genome and alterations in the histone code.

RESULTS

Global expression profiling identifies genes differentially expressed in ICF versus normal LCLs

We analyzed global gene expression patterns using the Affymetrix U133A/B GeneChip platform in five LCLs derived from normal individuals and three LCLs derived from individuals with ICF syndrome having known mutations in the *DNMT3B* gene [10759: P670T, GM08714: A603T and intron 22 G to A change resulting in insertion of three amino acids (STP), and 4088: V726G]. A total of 44 972 elements are contained on both the U133A and U133B

chips. ANOVA was used to define genes with a two-fold or greater change in expression ($P < 0.05$). This analysis generated a list of 979 probe sets (778 genes); 539 probe sets were upregulated (422 genes, 54.2% of the total) and 440 probe sets were downregulated (356 genes, 45.8% of the total) in ICF cells relative to normal cells (Supplementary Material, Table S1). Principal component analysis (PCA) performed on the most variable 0.01% of these genes indicated that there was some heterogeneity among both normal and ICF lines, although the two groups were clearly distinct (Fig. 1A and B). We performed a one-way ANOVA as a means of identifying genes with a high degree of variability between, but not within, each group (normal or ICF). This yielded a list of 296 probe sets showing a two-fold or greater change [$P < 0.05$, 196 probe sets (152 genes) upregulated and 100 probe sets (97 genes) downregulated] in ICF cells relative to normals (Supplementary Material, Table S2). Hierarchical clustering on gene and sample dimensions highlight the distinctive gene expression profiles of normal and ICF cells and also showed that ICF lines GM08714 and 10759 were more similar in their expression patterns to each other than to the 4088 ICF line (Fig. 1B). The 4088 line was an outlier in a number of respects, which may in part be due to the line having become aneuploid (discussed further later).

We then used the PANTHER gene ontology tools to group the 778 genes listed in Supplementary Material, Table S1 by gene function (18). Of these 778 genes, 472 were classified by PANTHER as being significantly enriched ($P < 0.05$) in certain biological process groups (Fig. 1C). The five categories with the largest number of affected genes were those involved in signal transduction (153 genes), development (106 genes), nucleic acid metabolism (117 genes), mRNA transcription (81 genes), and immunity and defense (69 genes). The two invariant phenotypes associated with ICF syndrome include immune defects and chromosomal instability; neurological defects and facial abnormalities are common but not present in all patients (19). Many of the genes that we identified (Supplementary Material, Tables S1 and S2) are directly or indirectly implicated in properly maintaining these processes in mammals. Genes with altered expression patterns in ICF cells related to immune function were identified (Supplementary Material, Tables S1, S3), consistent with a previous smaller microarray study (20), and include those related to B-cell function and immunoglobulin production [upregulated: *PILRA*, *ID2*; downregulated: Ig heavy chains (*IGHA1*, *IGHA2*, *IGHG1*, *IGHG2*, *IGHG3*, *IGHD*, and *IGHM*), *FCGR2B/2C*, *PIGR*, *FCRL4/5*, *VPREB3*)], the complement cascade (*CR2*, *CD46* upregulated), T-cell function (upregulated: *FKBP5*, *LMO3* and *TNFRSF4*; downregulated: *CD28*, *LTA*, *LMO2* and *CD27/TNFRSF7*), cytokine/chemokine signaling (upregulated: *CXCR4*, *IL1R1*, *IL1R2*, *TNFRSF19*, *CCR7* and *XCL1/2*; downregulated: *CCR6*, *CCR1*, *TNFSF11* and *IL8*), the myeloid/macrophage lineage (upregulated: *CD200*), and immune cells in general (the tyrosine kinases *LCK* and *SYK* and *ALOX5*—all downregulated). Interestingly, PANTHER analysis revealed 21 genes directly implicated in neural function, including those involved in neurotransmitter and synapse function (*BCHE*, *GABRA4*, *CACNB1*, *GRM7* and *SNCA*, Supplementary Material, Table S4) and 32 genes implicated in neurogenesis (Supplementary Material, Table S5). In addition, a large number of genes listed under the 'developmental processes' category (Supplementary Material, Table S6) are important transcriptional regulators implicated in brain (upregulated: *LHX2*, *PRRX1*, *ROBO1*, *HHEX*, *OTP* and *ASCL1*; downregulated: *IGF1*), immune system function (upregulated: *LHX2*, *HHEX*, *ROBO1* and *ID2*; downregulated: *RUNX1*), and cranial/skeletal development (upregulated: *LHX2*, *PRRX1* and *BMP-2*) and their misregulation in ICF cells may explain the other more variable developmental defects of this disease.

We then examined the 778 genes in Supplementary Material, Table S1 for the presence or absence of a CpG island in the 5′ proximal region (within 1500 bp of the transcription start site). Six hundred six out of 778 total genes contained a CpG island (77.9%, including both the less (21) and more rigorous (22) definitions of a CpG island). The distribution of CpG island-containing genes between the up- and downregulated genes was relatively similar, with 80 and 75.4% of up- and downregulated genes, respectively, containing a CpG island. Although the ICF-specific satellite 2/3 hypomethylation affects primarily chromosomes 1, 9 and 16 (19), genes with altered expression patterns were not enriched on these three chromosomes, although there was a non-random enrichment on chromosome 22, followed by chromosomes 3, 5, 18 and 21 (Fig. 1D). In general, the percent of up- and downregulated genes on each chromosome was relatively similar with several exceptions; more genes were downregulated in ICF compared with normal cells on chromosomes 5 and 14, whereas more genes were upregulated on chromosomes 8, 18 and the X chromosome (Fig. 1D). Consistent with our previous data (14), and results of others demonstrating that female ICF cells have defects in X chromosome inactivation (15), some of the most highly upregulated genes in the female ICF line GM08714 include the C-T genes *LAGE-1*, *LAGE-2* and *NY-ESO-1* (increased 18–25-fold, Supplementary Material, Table S1).

Independent confirmation of the expression microarray data reveals similarities and differences among ICF syndrome lines and provides a catalog of potential DNMT3B-regulated genes

The microarray analysis described in the previous section provided us with a large list of genes that may be directly or indirectly regulated by the DNA methylation or transcriptional repression functions of DNMT3B. We then chose 100 genes from Supplementary Material, Tables S1 and S2 for follow-up confirmation using semi-quantitative reverse transcriptase–polymerase chain reaction (RT–PCR) on three normal and three ICF LCLs. Criteria for choosing genes for follow-up confirmation included: genes showing the largest fold change, genes potentially contributing to the ICF phenotype, both up- and downregulated genes, and genes with and without CpG islands. The majority of the 100 genes demonstrated the expected differences in expression between normal and ICF cells predicted by the microarray analysis. Our results for 38 upregulated genes and 23 downregulated genes are shown in Fig. 2. For the group of genes that were upregulated in ICF relative to normal cells, we generally observed very low or no expression in the normal LCLs (Fig. 2A). We noted three

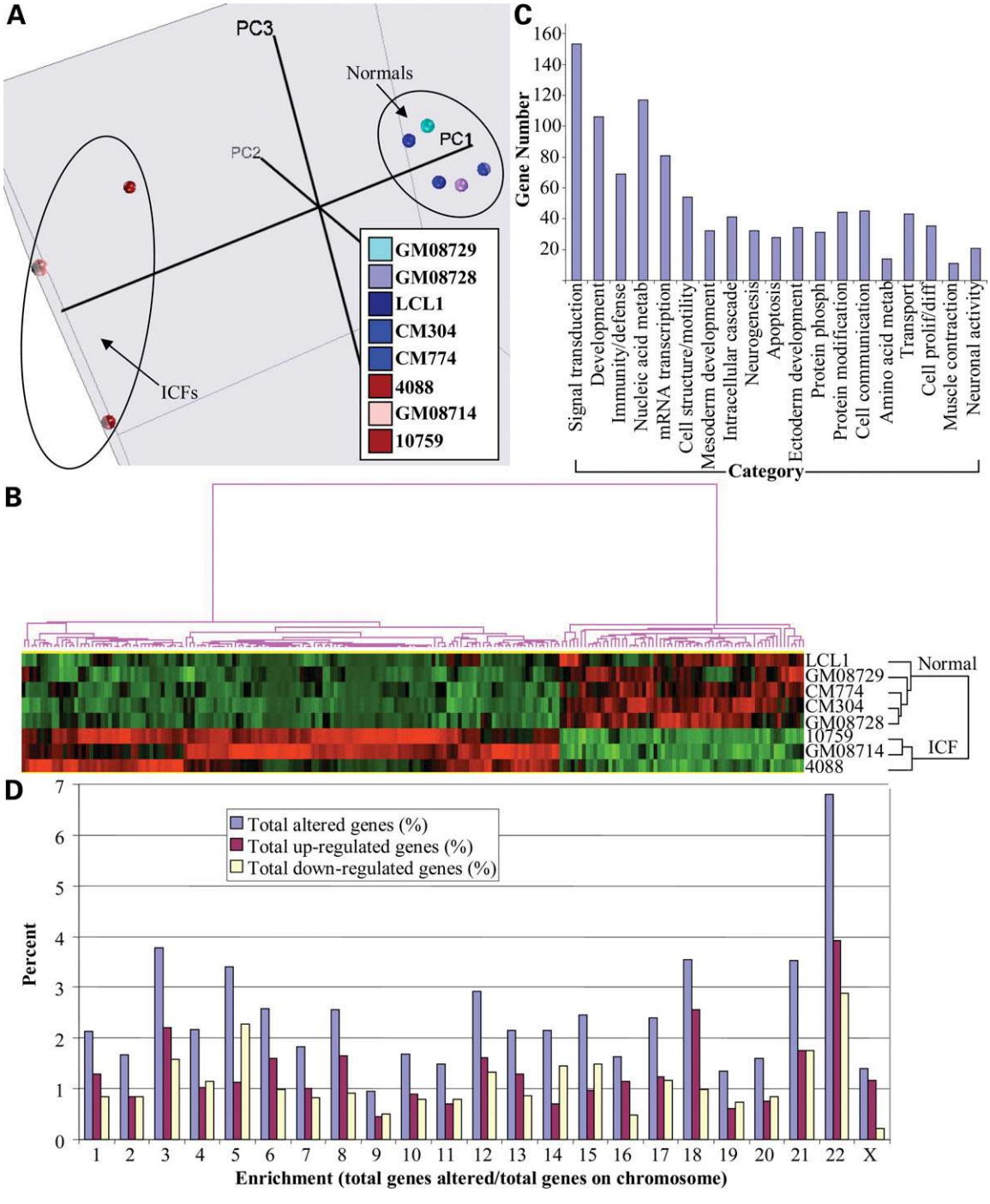

Figure 1. Global expression profiling of ICF syndrome and normal lymphoblastoid cell lines (LCLs). LCLs derived from five normal and three ICF syndrome individuals were used for the analysis. Total RNA was isolated, labeled and hybridized to Affymetrix U133A/B Genechips. Data were analyzed as described in the text. (**A**) Principal component analysis (PCA) of the eight cell lines used for gene expression profiling (represented by the colored balls). (**B**) Hierarchical clustering of the 296 probe sets from both the A and B chips that most reflect differences between normal and ICF lines (from Supplementary Material, Table S2). (**C**) Grouping of genes showing altered expression levels (2-fold or greater, $P < 0.05$) between normal and ICF lines into functional pathways using PANTHER. The number of genes from Supplementary Material, Table S1 in each of 19 different significantly enriched classes is shown on the *y*-axis ($P < 0.05$). (**D**) Genes with altered expression in ICF syndrome cell lines broken down by chromosome. The total percent of all affected genes on each chromosome from Supplementary Material, Table S1 was plotted as the number of up- and/or downregulated genes divided by the total predicted number of genes on each chromosome (as the percent of total affected genes). The total number and the number of up- or downregulated genes are shown. χ^2 test showed that the ICF gene distribution is highly associated ($P < 0.0001$).

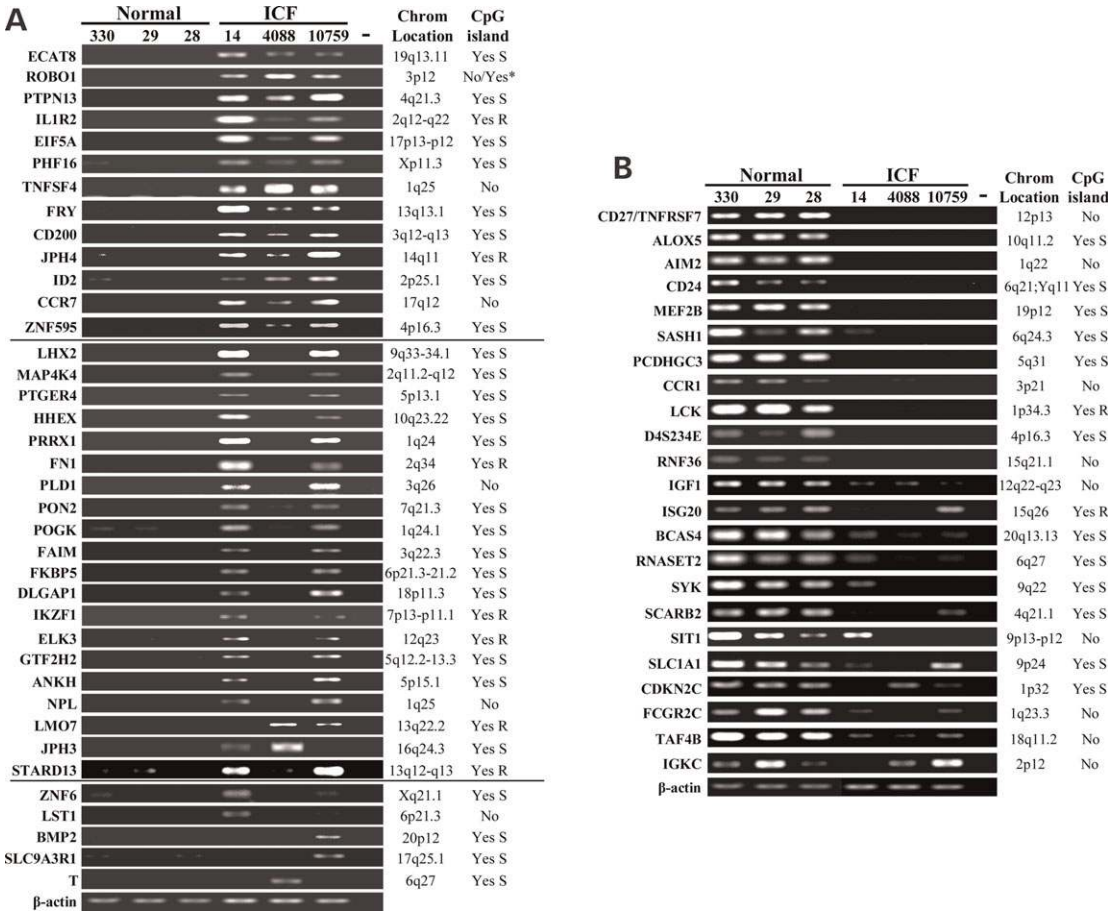

Figure 2. Confirmation of expression changes for a subset of genes identified as differentially expressed in ICF cells relative to normal cells by semi-quantitative RT–PCR. (**A**) Analysis of 38 genes predicted to be upregulated in ICF syndrome relative to normal LCLs. Results for three normal lines (LRL330-'330', GM08729-'29', and GM08728-'28') and three ICF lines (GM08714-'14', 4088 and 10759) are shown. Genes are divided into three classes based on degree of similarity in expression among the three ICF lines. (**B**) Analysis of 23 genes predicted by the microarray analysis to be downregulated in ICF relative to normal LCLs. The gene symbol is shown at the left and the chromosomal location and presence/absence of a CpG island at the 5' end of the gene ('Yes' or 'No') is indicated at the right. 'S' and 'R' indicate the degree of stringency of the CpG island as discussed in Results section, stringent or relaxed, respectively. Beta-actin (bottom gel panel in both parts) serves as a loading control. All reactions were repeated at least three times and results shown are representative. '–' water only negative control. The asterisk indicates that there are two alternative start sites for *ROBO1*, one is CpG-poor and one is a CpG island.

groups of expression patterns: upregulation of the gene in all three ICF lines (the first 13 genes in Fig. 2A), upregulation in two out of three ICF lines (the middle 20 genes in Fig. 2A) and upregulation in one out of three ICF lines relative to the normals (the last five genes in Fig. 2A). Clearly, there was heterogeneity between the three ICF lines. Interestingly, three homeobox genes (*PRRX1*, *LHX2* and *HHEX*) having important roles in multiple aspects of mammalian development (brain, skeletal/craniofacial and the immune system) were markedly upregulated in ICF cells, which may be related to recently described connections between the DNA methylation machinery and the polycomb complex PRC2 (23).

A closer examination of a subset of genes downregulated in ICF relative to normal LCLs by semi-quantitative RT–PCR is shown in Fig. 2B (23 genes). As for the upregulated genes, there was heterogeneity among the ICF and the normal lines as well. In general, though, all genes in this class were not expressed or expressed at low levels in ICF cells. The mechanism(s) by which *DNMT3B* mutations result in reduced gene expression is unclear, but may result from global alterations in transcription factor or repressor binding because of altered heterochromatin structure. Regardless of the mechanism, however, decreased expression of many of the genes we identified may contribute significantly to the ICF phenotype, particularly the immune system defects.

Pharmacologic inhibition of DNA methylation and histone deacetylation suggests that genes we identified are regulated by distinct epigenetic mechanisms

In order to begin to determine the role of DNMT3B's DNA methylation or transcriptional repression functions in regulating genes we identified, we treated two normal LCLs with the DNA methylation inhibitor 5-azadC (5 μM for 72 h) or the HDAC inhibitor trichostatin A (TSA, 100 nM for 24 h). We hypothesized that heavily methylated genes would be reactivated by 5-azadC only, whereas sparsely or unmethylated genes would be reactivated by TSA treatment. Interestingly, the majority of genes were upregulated by both TSA and 5-azadC treatment (Fig. 3), suggesting that multiple epigenetic

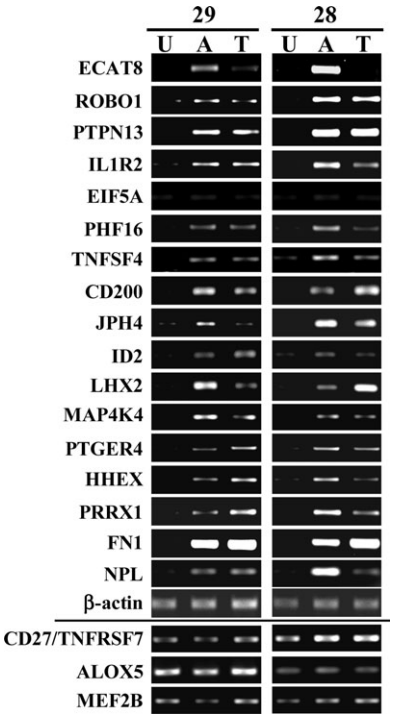

Figure 3. Distinct effects of histone deacetylation and DNA methylation inhibitors on genes showing altered expression in ICF LCLs. Two normal LCLs were treated with 5 μM 5-azadC ('A') for 72 h or with 100 nM TSA ('T') for 24 h, or were mock treated ('U'). Seventeen genes showing increased expression in ICF syndrome cells relative to normal cells were analyzed (see also Fig. 2A) by semiquantitative RT–PCR. Note that these genes were not expressed in normal LCLs or were expressed at low levels under standard growth conditions. Three genes that were downregulated in ICF relative to normal LCLs (see also Fig. 2B) were similarly analyzed (bottom three panels). Beta-actin served as a loading control.

marks are involved. Given that dense DNA methylation is generally sufficient to maintain genes in an inactive state in the face of TSA treatment (24), these results suggest that most genes are only sparsely or completely unmethylated. A smaller group of genes, including *ECAT8*, *JPH4*, *HHEX* and *NPL*, were induced predominantly by 5-azadC, suggesting that they are more heavily methylated. Finally, one gene induced in all three ICF cells, *EIF5A* (Fig. 2A), was not reactivated by 5-azadC or TSA treatment (Fig. 3), suggesting that its upregulation in ICF cells is an indirect effect unrelated to epigenetic modifications. We also examined the expression of three genes that were downregulated in ICF relative to normal cells, *CD27/TNFRSF7*, *ALOX5* and *MEF2B* (Fig. 2B), and observed that neither drug treatment altered their expression (Fig. 3, bottom). Therefore, mutations in *DNMT3B* and/or epigenetic modifications may not be directly responsible for downregulating gene expression for at least a subset of the genes in this category.

Loss of DNMT3B function in ICF syndrome cells results in subtle but significant changes in DNA methylation patterns at certain gene promoters

To obtain a more detailed analysis of DNA methylation patterns that would allow for an assessment of the contribution of DNA methylation in regulating gene expression, we performed bisulfite genomic sequencing (BGS) (25). *JPH4*, *CD200*, *PTPN13*, *ROBO1* and *ECAT8*, five genes upregulated in all ICF lines relative to the normals (Fig. 2A), were analyzed. We designed primers flanking the most CpG-rich regions for each gene. BGS analysis of the *JPH4* CpG island (located within exon 3 and encompassing the translation start site; the *JPH4* promoter was considerably less CpG-rich) revealed considerable DNA methylation in all cell lines with a marginally significant difference ($P = 0.086$) between ICF and normal LCLs (Table 1, Supplementary Material, Fig. S1). Differences in total levels of DNA methylation were assessed by applying ANOVA for nested models to the data set, which assesses differences between disease and normal as well as differences between samples within each class. *P*-values between 0.01 and 0.05 were considered significant, <0.01 very significant, and between 0.05 and 0.1 nearly (or marginally) significant. The *CD200* promoter CpG island was nearly devoid of DNA methylation in all lines although there was a marginally significant difference ($P = 0.084$) between ICF and normal lines (Table 1, Supplementary Material, Fig. S1). In contrast, both the *PTPN13* and *ROBO1*/*DUTT1* promoter CpG islands showed small but very significant differences in DNA methylation levels ($P < 0.0001$ and 0.001, respectively), with the three ICF lines less methylated than the normal LCLs (Fig. 4 and Table 1). *ECAT8* was distinctive because it was heavily hypermethylated in normal LCLs (95%) but lost much of this methylation in the ICF lines (Fig. 4 and Table 1, $P < 0.0001$). We also analyzed the CpG island promoter regions of the *IL1R2*, *EIF5A*, *ID2*, *MAP4K4* and *PRRX1* genes by BGS and of these five, only the *PRRX1* homeobox gene demonstrated marginally significant hypomethylation in ICF cells ($P = 0.051$ for *PRRX1*, $P > 0.1$ for the other four genes, Supplementary Material, Fig. S2, Table 1).

As an additional comparison, we then determined the DNA methylation status of two genes downregulated in ICF cells, *PCDHGC3* and *MEF2B*. Interestingly, the *PCDHGC3* promoter was more methylated in two out of three ICF lines (10759 and GM08714, $P = 0.0162$) although the total DNA methylation levels were still very low (~8%, Fig. 5 and Table 1). The *MEF2B* promoter was essentially devoid of DNA methylation in all lines (Fig. 5 and Table 1, $P = 0.7735$). Therefore, mutations in the *DNMT3B* gene in ICF syndrome cells result in subtle hypomethylation events in some promoter/transcribed regions of the genome. Although the level of DNA methylation in normal cells is relatively low, it is possible that these methylation events contribute to downregulation of the associated gene or mark it for addition of repressive histone modifications. Genes lacking DNA methylation are likely regulated by other aspects of chromatin structure, an idea directly tested in the next section.

Given the importance of homeobox genes to mammalian development and emerging connections between DNA methylation and polycomb proteins, which regulate expression of homeobox genes, we examined the DNA methylation status of one such gene, *LHX2*, in greater detail. We analyzed the *LHX2* promoter CpG island from −515 to +96 with two sets of primers and a portion of the large downstream CpG island (+6051 to +6618 relative to the transcription start, Fig. 6). BGS analysis revealed that the promoter was

Table 1. Summary of all BGS data presented as the percent methylation across all clones and all CpG sites (from Figs. 4 to 6, Supplementary Material, Figs. S1 and S2)

Change in expression	Gene name	Cell line (% methylation)						
		10759	GM08714	4088	ICF Ave.	GM08729	LRL330	Normal Ave.
Increased in ICF cells	JPH4	19.8	17.8	13.3	17.0	11.7	13.8	12.7*
	CD200	0.7	0.4	0.4	0.5	1.8	1.6	1.7*
	PTPN13	0.3	4.9	0.6	1.9	11.4	8.5	10.0***
	ROBO1-iso b	0.5	3.8	0.0	1.4	9.3	8.3	8.8***
	LHX2-region 1	0.0	0.0	0.7	0.2	6.3	6.3	6.3***
	LHX2-region 2	0.4	0.0	0.4	0.2	7.1	7.5	7.3***
	LHX2-region 3	8.4	1.8	2.5	3.0	20.5	21.5	21.0***
	IL1R2	5.7	38.6	17.0	20.4	8.0	14.8	11.4
	EIF5A	1.6	0.5	0.5	0.9	0.0	1.1	0.6
	ID2	0.4	0.4	0.0	0.3	0.8	1.3	1.0
	MAP4K4	0.5	0.2	0.8	0.5	0.3	0.8	0.6
	PRRX1	0.0	1.4	6.3	2.6	9.7	9.7	9.7*
	ECAT8	13.0	38.3	48.6	33.3	95.0	95.0	95.0***
Decreased in ICF cells	PCDHGC3	8.0	8.7	0.4	3.0	0.0	2.0	1.0**
	MEF2B	0.4	0.0	0.0	0.1	0.0	0.4	2.0

***$P < 0.01$, very significant; **$P = 0.01–0.05$, significant; *$P = 0.05–0.1$, nearly significant, by ANOVA.

methylated at a low level (~6–7% CpG's methylated) in normal LCLs and significantly hypomethylated in ICF cells (<1% CpG's methylated, $P < 0.0001$ for *LHX2* region 1 and $P = 0.0005$ for *LHX2* region 2) (Fig. 6 and Table 1). As we moved into the transcribed CpG island in intron 3 of *LHX2*, the DNA methylation differences became more pronounced; ~21% of the CpG sites were methylated in normal LCLs compared with only 3% in the ICF lines for region 3 (Fig. 6 and Table 1, $P < 0.0001$). These findings suggest that DNA methylation may have a role in suppressing transcription of the *LHX2* gene in normal LCLs and this effect could be mediated at the promoter or from within the transcribed region. Methylation within transcribed regions has been shown to negatively regulate the transcription (26). Taken together, the BGS data reveal three distinct methylation signatures for the genes we analyzed (1), low-level DNA methylation in normals and significant hypomethylation in ICF cells (e.g. *PTPN13* and *LHX2*) (2), high-level DNA methylation in normals and significant hypomethylation in ICFs (*ECAT8*), and (3) genes showing no difference in methylation between normal and ICF LCLs (most of which are largely unmethylated).

Altered gene expression in ICF syndrome cells is associated with pronounced changes in the histone code

Our analysis of DNA methylation patterns of 13 genes revealed subtle but significant DNA methylation differences in six genes (Table 1). Although these differences were statistically significant, total levels of methylation in the normal LCLs remained low (<20%, with the exception of *ECAT8*); therefore, it is unclear whether this level of DNA methylation alone would suffice to suppress expression since most aberrantly silenced genes in tumor cells are densely methylated (24). For most other genes examined we found little or no methylation, suggesting that aspects of the histone code also contribute to their regulation. To examine this in greater detail, we performed ChIP analyses using six antibodies, two specific for histone modifications characteristic of transcriptionally active/permissive chromatin [histone H3 trimethylated on lysine 4 (3XMe-H3K4), H3 acetylated on lysines 9 and 18 (Ac-H3K9/K18)] and four specific for transcriptionally repressed chromatin [histone H3 trimethylated on lysine 9 (3XMe-H3K9), H3 di- and trimethylated on lysine 27 (2XMe-, 3XMe-H3K27), and histone H4 trimethylated on lysine 20 (3XMe-H4K20)] (27–29). We analyzed two regions of the *JPH4* gene, one flanking the relatively CpG-poor transcription start site and one within the exon 3 CpG island (the latter was also analyzed by BGS, Supplementary Material, Fig. S1). The *JPH4* promoter showed relatively little difference in H3K4 and H3K27 di-/trimethylation levels, and H3 acetylation level, whereas the repressive H3K9 trimethylation mark was completely absent in ICF cells (Fig. 7, Supplementary Material, Fig. S3). The combination of both repressive and active modifications within the same region has been reported for certain genes in T cells (28). In contrast, however, the *JPH4* exon 3 region showed dramatic differences in all histone modifications between ICF and normal LCLs except for H3K27 dimethylation and H4K20 trimethylation. The exon 3 region of *JPH4* showed elevated H3K4 trimethylation and H3K9/K18 acetylation and complete loss of H3K9 and H3K27 trimethylation in ICF cells compared with normal LCLs (Fig. 7, Supplementary Material, Fig. S3). Normal rabbit IgG served as a negative control and demonstrated that background binding in our ChIP experiments was very low.

We then used ChIP to examine the promoter regions of both *ROBO1* isoforms. H3K4 trimethylation was actually elevated at the isoform a promoter in normal compared with ICF LCLs, despite the gene being repressed, whereas there was little H3K4 trimethylation at the isoform b promoter in all four LCLs (Fig. 7). H3K27 dimethylation was unchanged and markedly reduced, respectively, at *ROBO1* isoform a and b promoters (Supplementary Material, Fig. S3). Both *ROBO1* promoters showed drastic reductions in H3K9 and H3K27 trimethylation levels in ICF syndrome cells, consistent

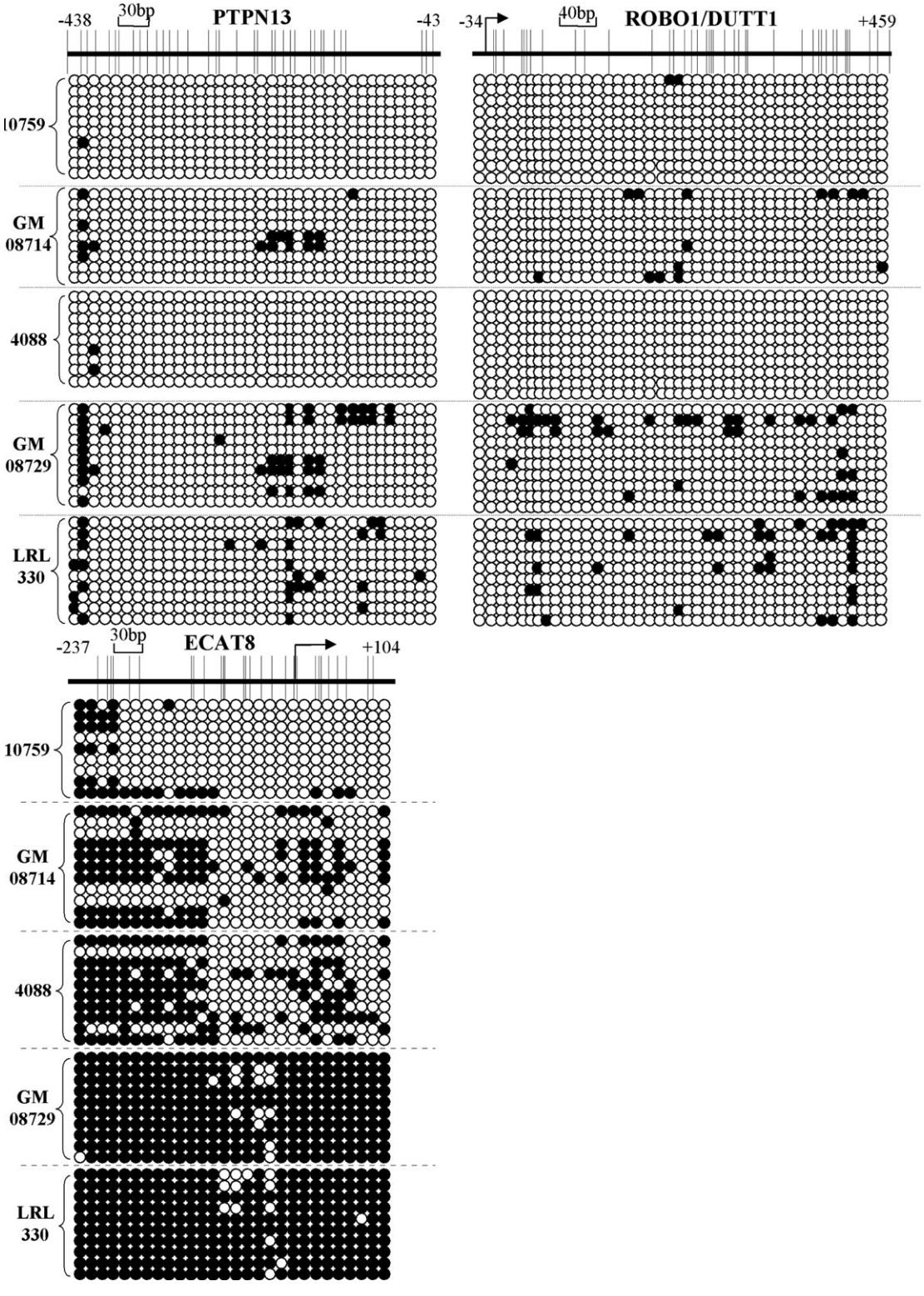

Figure 4. Bisulfite genomic sequencing (BGS) analysis of genes demonstrating increased expression in ICF syndrome cells. The CpG-rich regions of three genes, *PTPN13*, *ROBO1*/*DUTT1* and *ECAT8* were analyzed for differences in DNA methylation patterns in three ICF cell lines and two normal cell lines. A CpG plot of the region analyzed is at the top of each panel and circles below this indicate methylation status; filled circles, methylated CpGs, open circles, unmethylated CpGs. Numbering is relative to the transcription start site as defined using NCBI MapViewer. The bent arrow, when shown, is the transcription start site. Each row of circles corresponds to one bacterial clone. Following bisulfite treatment and PCR, resulting products were TA-cloned and at least eight independent colonies were sequenced. A summary of the total percent methylation for all genes is given in Table 1.

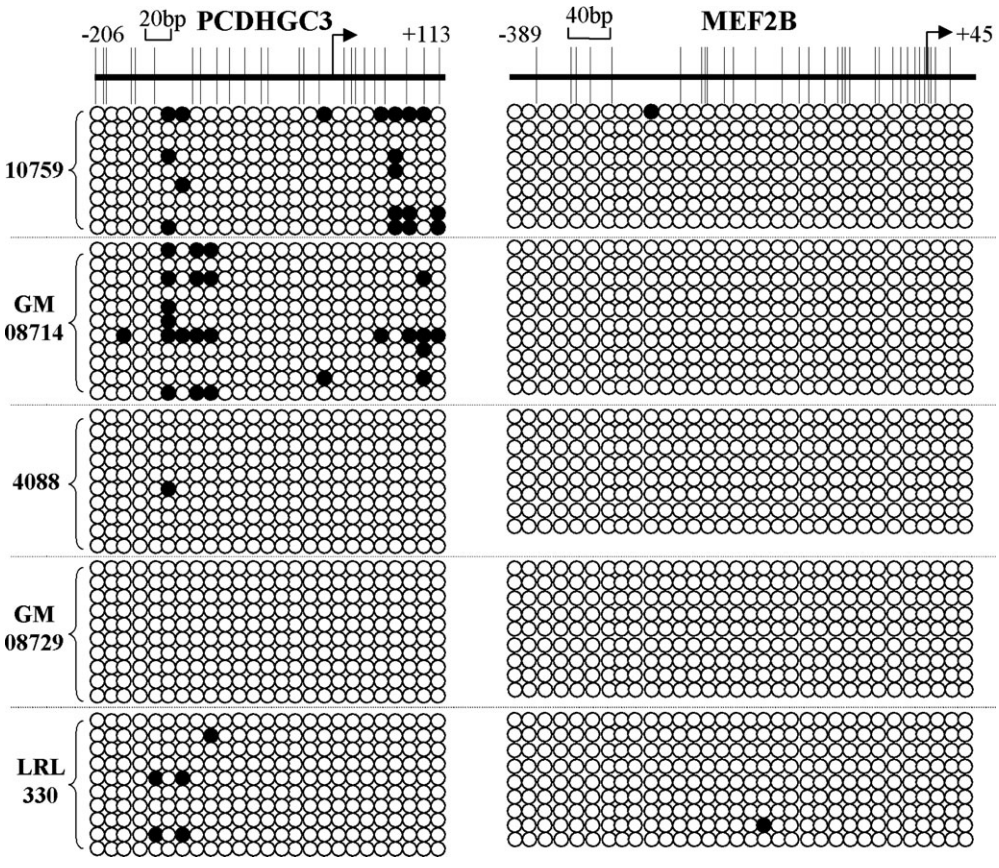

Figure 5. Bisulfite genomic sequencing analysis of genes demonstrating decreased expression in ICF syndrome cells. The CpG-rich regions of two genes, *PCDHGC3* and *MEF2B*, were analyzed for differences in DNA methylation patterns in three ICF cell lines and two normal cell lines. Symbols and labels are the same as in Fig. 4. A summary of the total percent methylation for all genes is provided in Table 1.

with these promoters being transcriptionally active in both ICF lines (Fig. 7). We next examined two regions of the *PTPN13* promoter by ChIP and both consistently demonstrated marked increases in levels of H3K4 trimethylation and H3 acetylation and decreased levels of H3K9 and H3K27 di-/trimethylation in ICF relative to normal LCLs (Fig. 7, S3).

We next examined the CpG island promoter regions of two genes, *CD200* and *ID2*, that were upregulated in all three ICF lines, but which showed almost no DNA methylation in their promoters (Table 1). For the *CD200* promoter, we observed little difference in H3K4 trimethylation, H3 acetylation, H3K27 dimethylation and H4K20 trimethylation levels in normal compared with ICF lines; however, H3K9 and H3K27 trimethylation levels were dramatically reduced in ICF cells (Fig. 7). A similar examination of the *ID2* promoter revealed that H3 acetylation levels were similar across all four cell lines while H3K4 trimethylation was slightly reduced in the ICF cells. In contrast, H3K9, H3K27 and H4K20 methylation were all strongly reduced or absent at the *ID2* promoter in ICF syndrome cells (Fig. 7, S3). Taken together, ChIP analysis of five genes upregulated in ICF syndrome cells revealed marked changes in levels of several histone modifications that were consistent with these genes going from an inactive or low expression state in normal LCLs to a moderate to high-level expression state in ICF syndrome cells. Interestingly, the most consistently altered modification was H3K27 trimethylation, which is associated with transcriptional repression mediated by the EZH2 subunit of the PRC2 polycomb complex (30).

ChIP analysis of the *ECAT8* promoter revealed a distinctive situation. *ECAT8* is a member of a group of genes, including *OCT3/4* and *NANOG*, that are highly expressed in undifferentiated ES and male germ cells and may play a role in maintaining pluripotency (31). In normal LCLs, we detected the repressive H3K9, H3K27 and H4K20 methylation marks, in keeping with the dense DNA methylation in these cells. Interestingly, however, in ICF cells where the gene is reactivated, levels of H3K4 trimethylation and H3 acetylation marks dramatically increased without any apparent reduction in repressive modifications (Supplementary Material, Fig. S4).

We identified and confirmed that three homeobox-containing transcriptional regulators, *LHX2*, *PRRX1* and *HHEX*, were upregulated in ICF syndrome cells and that two of these, *LHX2* and *PRRX1*, had low to modest levels of DNA methylation in normal LCLs that were lost in ICF LCLs. Given the importance of this class of genes to proper mammalian development and emerging connections between DNA methylation and polycomb complexes, we analyzed the histone modification patterns at multiple regions of these three genes. ChIP results for three regions within the *LHX2*

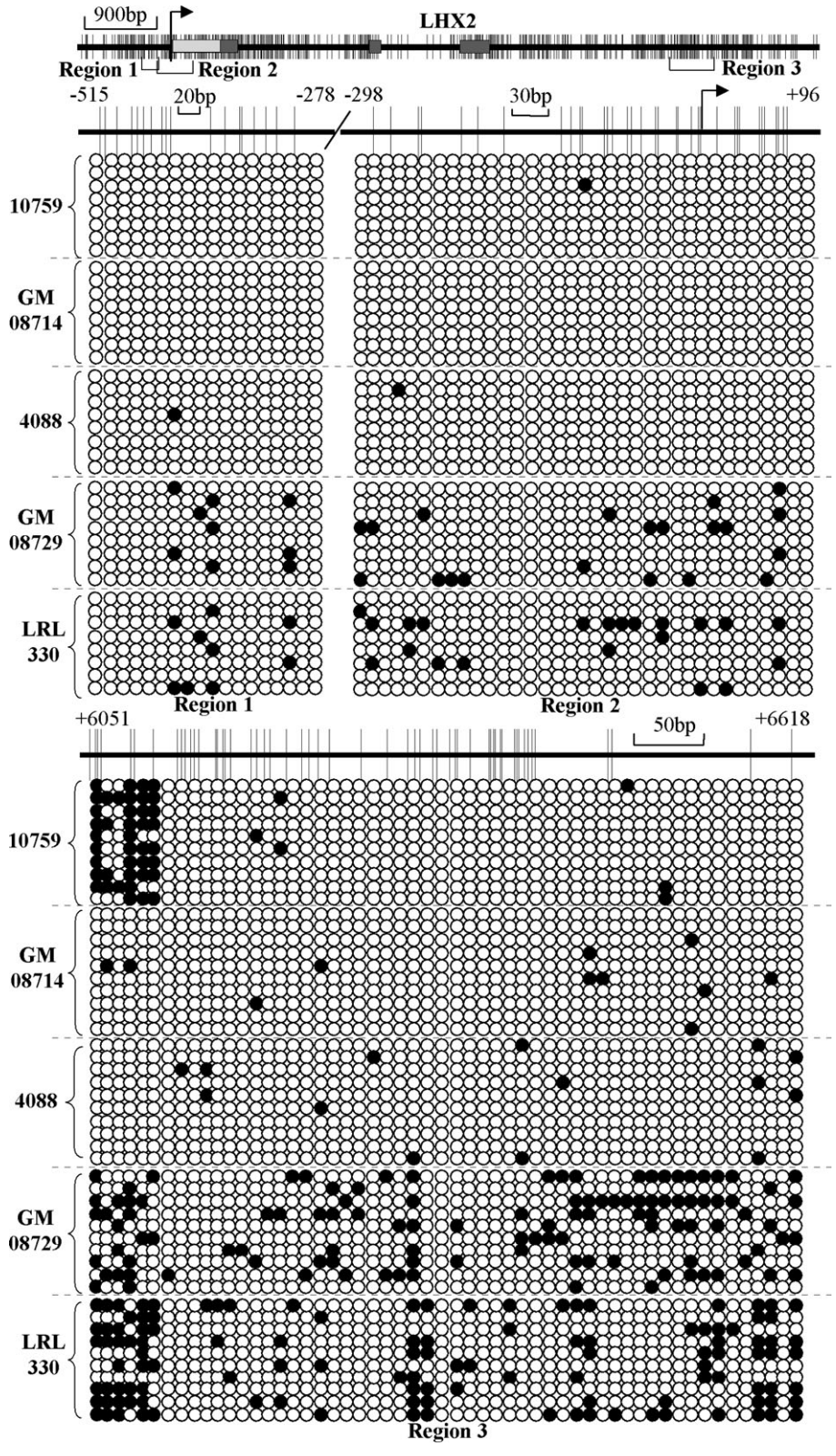

Figure 6. Bisulfite genomic sequencing analysis of the *LHX2* homeobox gene whose expression is increased in ICF relative to normal LCLs. Three CpG-rich regions of *LHX2* were analyzed (denoted regions 1–3 below the CpG plot at the top) in the same five cell lines used in Figs. 4 and 5; two within the promoter/5′ region and one within a very CpG-rich domain in intron 3. Symbols and labels are the same as in Fig. 4. Boxes represent exons; dark gray, translated; light gray, 5′-UTR. A summary of the total percent methylation for all regions is provided in Table 1.

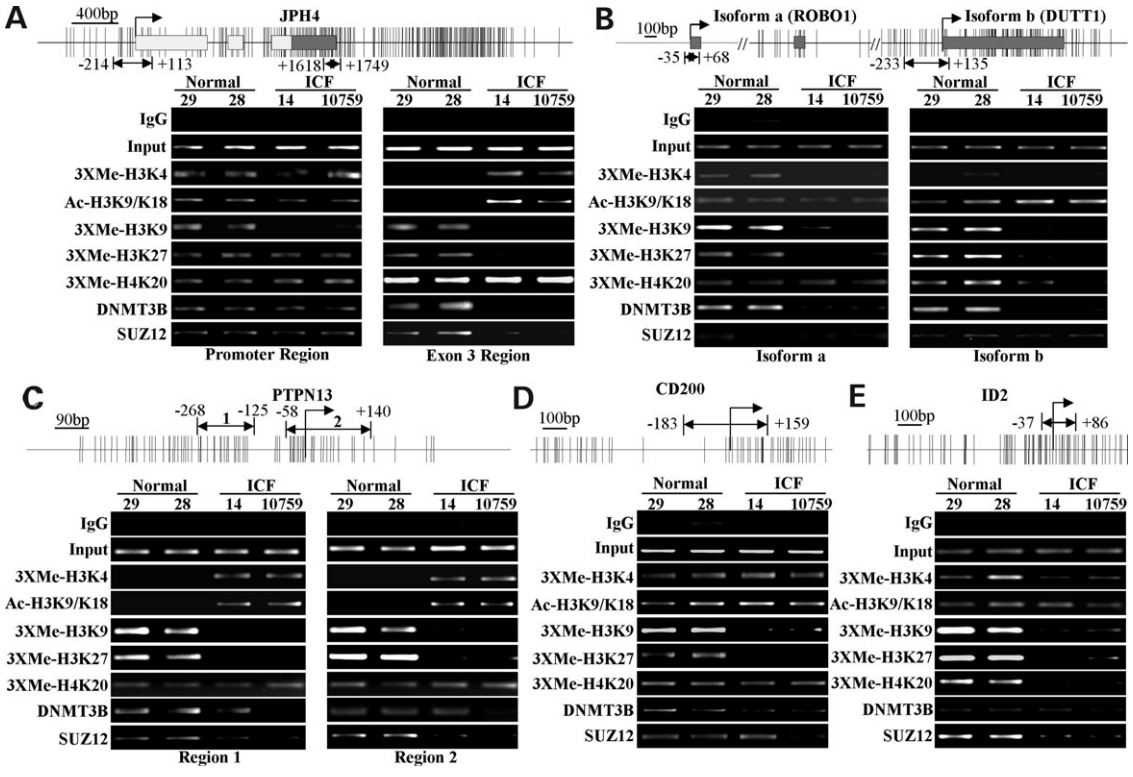

Figure 7. Chromatin immunoprecipitation (ChIP) analysis of five genes upregulated in ICF syndrome cells: (**A**) *JPH4*, (**B**) *ROBO1*/*DUTT1* (isoform a/isoform b), (**C**) *PTPN13*, (**D**) *CD200* and (**E**) *ID2*. The top of each panel is a CpG plot for each gene with locations of exons (boxes), transcription start site (bent arrow) and ChIP PCR primers (double headed arrows, numbering is relative to the transcription start site). Sonicated chromatin was prepared from two normal and two ICF LCLs and used in immunoprecipitations with the antibodies listed at the left of each panel. Normal rabbit IgG served as a negative control for non-specific binding. Input chromatin (1% of the ChIP reactions) is also shown as a loading control. Antibodies were directed against trimethylated ('3XMe') forms of H3K4, H3K9, H3K27 and H4K20, acetylated ('Ac') H3K9/K18, as well as DNMT3B and the PRC2 subunit SUZ12. Immunoprecipitated DNA was used with the primers indicated in semiquantitative PCR reactions and then analyzed on 2% agarose gels. Reactions were repeated at least three times from two independent ChIP reactions and the results were consistent (not shown). ICF line 4088 was not analyzed because we were unable to grow it in quantities sufficient for ChIP. Note that *ROBO1* isoform b originates from within intron two of the isoform a transcript. Regions analyzed by ChIP correspond closely to those analyzed by BGS (Fig. 4, Supplementary Material, Figs. S1 and S2).

locus were remarkably consistent and showed dramatic increases in H3K4 trimethylation and H3 acetylation, and similarly large decreases in H3K9, H3K27, and to a lesser extent H4K20 trimethylation in ICF relative to normal LCLs (Fig. 8). H3K27 dimethylation levels were consistently reduced in ICF cells only at the *LHX2* promoter (Supplementary Material, Fig. S3). Three regions of the *PRRX1* promoter were then analyzed by ChIP and all showed differences between normal and ICF syndrome cells similar to what we observed for *LHX2*, with the exception of H4K20 trimethylation (Fig. 8). Interestingly, H3K27 dimethylation was again reduced only at the *PRRX1* transcription start site (region 3, Supplementary Material, Fig. S3). Lastly, we examined two regions at the 5' end of the *HHEX* gene and both consistently demonstrated dramatically decreased H3K9 and H3K27 trimethylation levels, but little change in H3K27 dimethylation and H4K20 trimethylation. H3K4 trimethylation and H3 acetylation levels were not consistently altered (Fig. 8, Supplementary Material, Fig. S3). Taken together, the ChIP analyses presented in this section demonstrate that histone modifications are dramatically altered in the promoters of many derepressed genes in ICF syndrome cells, with the most consistent change being loss of H3K27 trimethylation.

Changes in expression and histone H3K27 trimethylation levels correlate with binding of polycomb protein SUZ12 and DNMT3B

We next used ChIP to determine whether DNMT3B bound to genes showing altered histone modifications and if this binding was altered in ICF syndrome cells. ChIP revealed three patterns of DNMT3B binding. The first pattern, exemplified by the *JPH4* promoter, *ID2*, and *PRRX1* region 3, was characterized by very low DNMT3B binding that was unaltered in ICF syndrome cells (Figs. 7 and 8, 'DNMT3B' panels). The second pattern, represented by *PTPN13*, *LHX2* region 2, *PRRX1* region 2, and *HHEX* region 2, showed complete or near complete loss of DNMT3B binding in ICF line 10759 (Figs. 7 and 8). Finally, the third pattern of DNMT3B ChIP result, observed at the majority of regions analyzed, was characterized by pronounced loss of DNMT3B binding in both ICF syndrome lines (Figs. 7 and 8). Collectively, these results suggest that ICF-specific *DNMT3B* mutations influence its ability to bind certain DNA sequences within the context of chromatin *in vivo* and that DNMT3B binds at genes repressed in normal LCLs even if they had a low level of DNA methylation (e.g. *CD200*). This data, combined

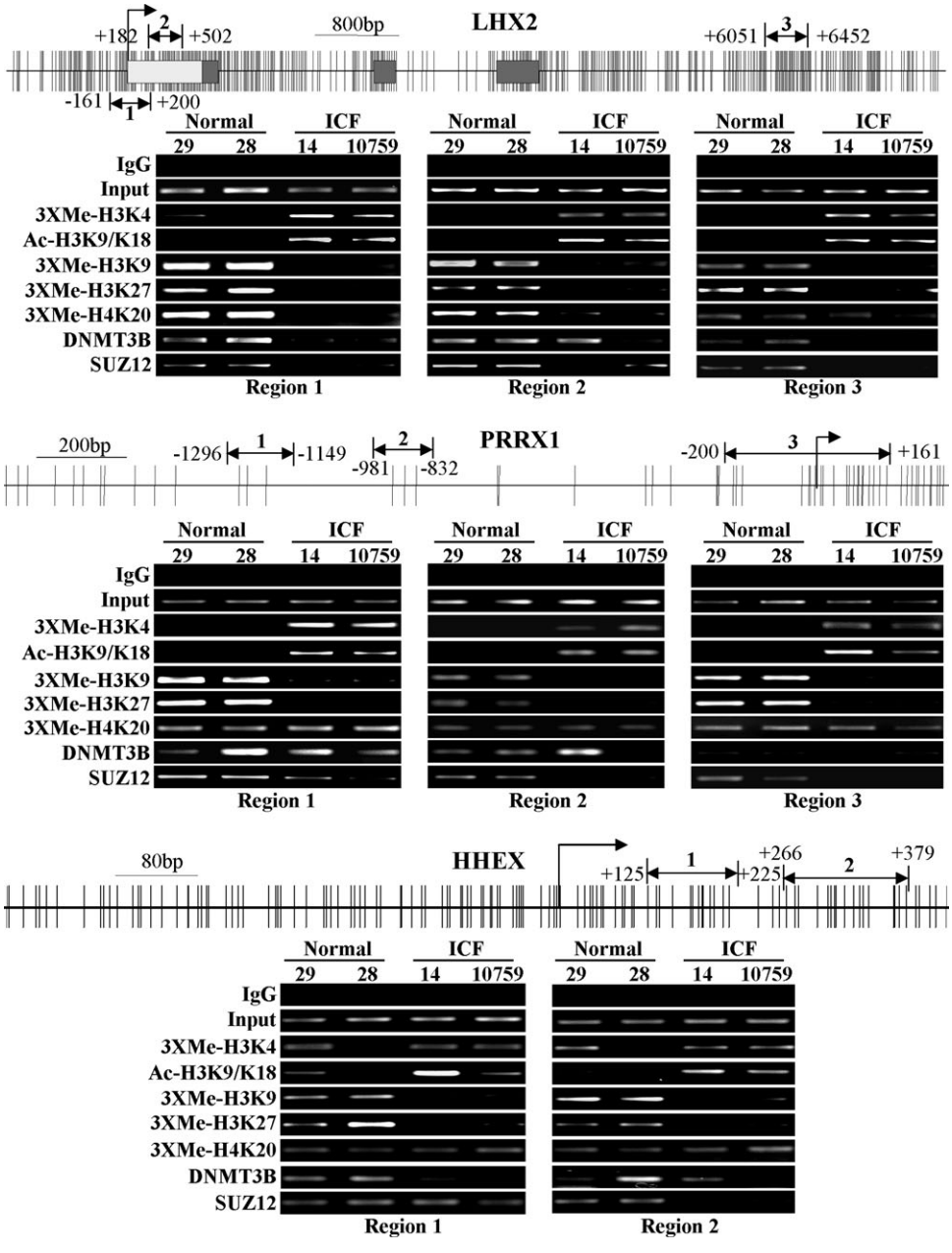

Figure 8. ChIP analysis of multiple regions of three homeobox genes demonstrating increased expression in ICF syndrome cells, *LHX2*, *PRRX1* and *HHEX*. Labels, cell lines and antibodies are as described in Fig. 7. Regions analyzed by ChIP correspond closely to those analyzed by BGS for *LHX2* and *PRRX1* (Fig. 6, Supplementary Material, Fig. S2).

with our pharmacologic inhibition experiments (Fig. 3), also suggest that DNMT3B-mediated recruitment of HDACs or histone methylases, in the absence of *de novo* DNA methylation, is an important contributor to gene regulation.

Given our ChIP results showing that the most consistently altered histone modification in ICF syndrome cells was H3K27 trimethylation, we examined PRC2 binding using an antibody against SUZ12. The correlation between H3K27 trimethylation levels and SUZ12 binding by ChIP was remarkably consistent; all regions, with the exception of the *CD200* promoter (GM08714) and *HHEX* region 1, that showed reduced H3K27 trimethylation also exhibited markedly reduced SUZ12 binding (Figs. 7 and 8, 'SUZ12' panels). The promoters of both *ROBO1* isoforms bound SUZ12 at very low levels and no differences were detected between normal and ICF syndrome cell lines (Figs. 7 and 8). Changes in SUZ12 binding paralleled those in DNMT3B binding in many, but not all cases, suggesting that binding of these two transcriptional regulators is not strictly linked. Interestingly, for the *ECAT8* promoter, binding of DNMT3B and SUZ12 was only detected in ICF cells (Supplementary Material, Fig. S4). The reasons for this are unclear at

present, although work from others has shown that polycomb complexes do indeed bind to genes that are actively transcribed (32).

Global levels of histone modifications in ICF and normal LCLs

Given our results that certain histone modifications were commonly altered in ICF syndrome cells at specific genomic regions examined by ChIP, we wondered whether this was a reflection of global changes in histone modifications. To examine this, we isolated total nuclear protein from two normal and two ICF syndrome lines and established that the four protein samples were equally loaded using an antibody to total histone H3 in western blotting and by staining a comparably loaded gel with coomassie blue (Fig. 9A, bottom two panels). Next, we probed the extracts with eight different histone modification antibodies, including all of the antibodies used in the previous ChIP experiments. Levels of H3K9/K18 acetylation, dimethylated H3K4 and trimethylated H3K9 and H4K20 showed no difference in ICF compared with normal LCLs (Fig. 9). H3K27 trimethylation was mildly reduced and H3K27 dimethylation was markedly decreased in ICF syndrome cells. Interestingly, we observed a pronounced increase in the level of ubiquitinated H2AK119, a mark mediated by the PRC1 polycomb complex. H3K4 trimethylation levels were also significantly increased in ICF cells (Fig. 9). Quantitation of changes in levels of select modifications was performed by densitometry scanning of western blots run in triplicate (Fig. 9B) and confirmed that levels of trimethylated H3K4, dimethylated H3K27 and monoubiquitinated H2AK119 were significantly altered ($P \leq 0.05$). Differences in global histone modifications were not attributable to altered expression of DNMT3B or the PRC2 complex (SUZ12) as these were relatively constant across all four cell lines. PRC1 levels (BMI1) were mildly increased in ICF cells, consistent with the increased H2AK119 ubiquitination (Fig. 9). Taken together, these results suggest that the loss of normal DNMT3B function leads to perturbations in certain histone modifications, particularly those mediated by the polycomb complexes.

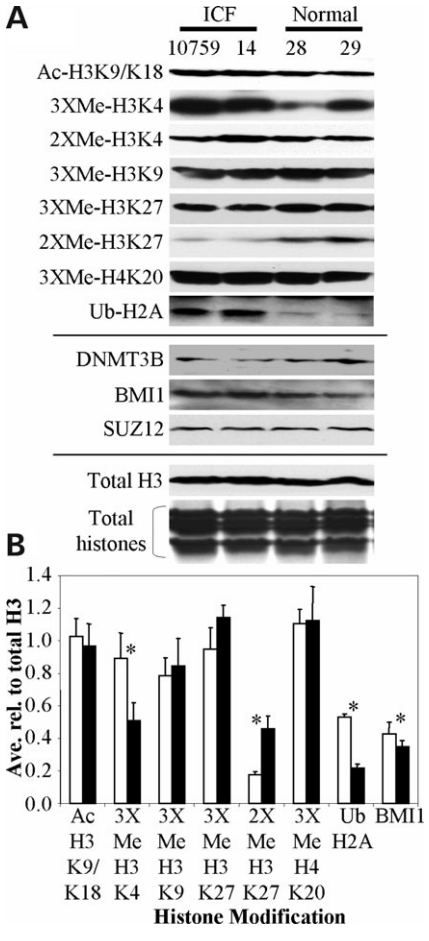

Figure 9. Global levels of histone modifications in ICF and normal LCLs assessed by western blotting. (**A**) Isolated nuclear protein (soluble nuclear plus chromatin-enriched fractions) derived from two normal and two ICF LCLs was separated on SDS–PAGE gels, transferred to PVDF membrane and probed with the antibodies indicated at the left. Protein equivalent to 2×10^6 cells is loaded in each lane. Equal loading of all four samples is demonstrated by western blotting for total histone H3 and by coomassie blue staining ('total histones') of the core histones (bottom-most two panels). Antibody abbreviations are the same as in Fig. 7. 'Ub'-H2A monoubiquitinated at K119, 2XMe, dimethylated form; 14, GM08714; 28, GM08728; 29, GM08729; BMI1 is a subunit of the PRC1 polycomb complex, SUZ12 is a subunit of the PRC2 polycomb complex. (**B**) Densitometry quantitation of select histone modifications from (A). Each protein sample was run in triplicate, scanned and quantitated. White bars, ICF cell lines; black bars, normal LCLs. Bars are the average signal intensity of six independent western blots (three from each cell line) normalized to the total histone H3 signal for each cell line. Error bars are the standard deviation from the mean. Asterisks indicate statistical significance (*t*-test, $P \leq 0.05$, only the BMI1 result yielded a $P = 0.05$, all others were <0.05).

ICF-specific changes in expression of transcriptional regulators identified by microarray analysis allow for prediction of alterations in relevant downstream targets

Many gene expression changes we identified can be predicted to directly lead to aspects of the ICF phenotype, such as downregulation of immunoglobulin heavy-chain genes as a contributor to the immune system defects. A large number of the differentially expressed genes are direct regulators of transcription. Therefore, we sought to determine whether upregulation of such genes in ICF syndrome cells leads to subsequent changes in downstream target genes that may also contribute to the ICF phenotype but are not directly regulated by DNMT3B function. We chose two genes to test this notion, *MAP4K4* and *PRRX1* (Fig. 2A), which regulate expression of downstream targets with potential functional relevance to ICF syndrome. *MAP4K4* (*NIK*, mitogen activated protein 4 kinase 4) is a serine/threonine kinase activated by the c-jun N-terminal kinase pathway that suppresses transcription of *C/EBPα* (33). *C/EBPα* is a key transcription factor in many tissues, including the hematopoietic system. Consistent with this prediction, expression of C/EBPα was reduced in ICF syndrome cell lines (Fig. 10A). It has also been reported that mice lacking *prrx1* and the related homeobox gene *prrx2* showed large decreases in sonic hedgehog expression (Shh, a critical developmental regulator), suggesting that the *prrx* genes activate *Shh* expression at critical sites during development (34). ICF cell lines with upregulated PRRX1 expression also demonstrated strongly upregulated SHH levels (Fig. 10A).

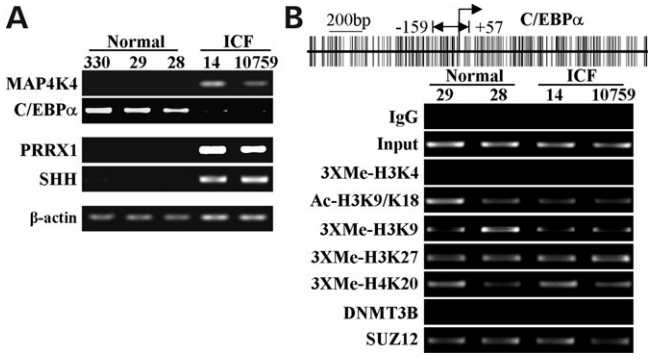

Figure 10. Use of microarray expression data to predict changes in potentially relevant downstream target genes in ICF syndrome cells. *C/EBPα* and *SHH* (sonic hedgehog) are downstream targets of the transcriptional regulators *MAP4K4* and *PRRX1*, respectively, that were upregulated in ICF syndrome cells. (**A**) Semi-quantitative RT-PCR analysis for the gene listed at the left in three normal and two ICF LCLs. Beta-actin (bottom panel) serves as a loading control. (**B**) ChIP analysis of the *C/EBPα* promoter region. Antibodies, conditions and abbreviations are the same as described in Fig. 7. A CpG plot is shown at the top with the location of the ChIP primers indicated with a double-headed arrow.

This observation is particularly intriguing in light of the relatively common craniofacial defects seen in ICF patients, hypertelorism and flat nasal bridge, because artificially elevated expression of SHH in an embryonic chick model system resulted in a similar phenotype (35). If regulation of *C/EBPα* and *SHH* is indeed downstream of other transcriptional events occurring in ICF cells, we hypothesized that patterns of histone modifications and/or DNMT3B binding would be distinct compared with direct targets. Consistent with this, we observed that histone modifications at the *C/EBPα* CpG island promoter were different from what we observed previously, with most modifications showing little change between ICF and normal LCLs (Fig. 10B). Interestingly, although SUZ12 bound equally well in normal and ICF lines, we were unable to detect DNMT3B binding to the *C/EBPα* promoter in any of the cell lines (Fig. 10B). Therefore, the microarray data are useful not only for identifying direct targets of DNMT3B repression, but also for predicting potential downstream target genes whose misregulation may contribute to ICF syndrome.

Heterogeneity among ICF cell lines and genomic instability associated with *DNMT3B* mutation

Analyses presented in Figures 1 and 2 revealed heterogeneity in patterns of gene expression among the three ICF syndrome LCLs, but the 4088 cell line consistently demonstrated the greatest difference relative to GM08714 and 10759. The 4088 cells also grew more slowly and were more sensitive to serum concentration (data not shown). These differences may have arisen from the differing spectrum of *DNMT3B* mutations present in each cell line; however, we were interested in determining whether other factors might contribute. We first examined levels of apoptosis under standard laboratory conditions. ICF 4088 had a mildly elevated level of apoptosis as monitored by annexin V staining (data not shown). A far more distinguishing characteristic of the 4088 cell line was revealed upon cell cycle analysis. Although there were small variations in the percentage of cells in G1, S and G2/M phases of the cell cycle among all lines, 4088 was remarkable in that nearly 70% of the cells were tetraploid (Fig. 11). This property does not appear to be common among ICF LCLs; however, too few may have been analyzed in this manner to conclude that 4088 is unique in this regard. A much earlier study using this cell line reported aneuploidy as well, although no data were shown (36), suggesting this is an inherent property of the 4088 line. The tetraploid state of ICF 4088 is particularly intriguing in light of other studies in which the *Dnmt3b* gene was conditionally inactivated in mouse embryo fibroblasts (MEFs), revealing that >50% of the *Dnmt3b* knockout MEFs had a 4N DNA content (37). Taken together, at least some of the differences in the 4088 cell line's gene expression pattern may arise from the large fraction of cells in the culture that are tetraploid.

DISCUSSION

Our studies of ICF syndrome presented here provide a better understanding of how DNA methylation interfaces with the histone code. We identified nearly 800 genes with an altered pattern of expression in ICF relative to normal cells, including new genes highly relevant not only to the immune deficits of ICF syndrome, but also to the developmental and neurological defects observed in many ICF patients. BGS demonstrated that some genes with altered expression had no change in DNA methylation, but importantly, we show for the first time that nearly half of the genes we examined are methylated at low levels (5–20%) in normal cells and this methylation is lost in ICF cells. The importance of histone modification patterns was also apparent following ChIP analysis, which revealed that genes upregulated in ICF cells lose histone modifications characteristic of repressed chromatin (particularly histone H3K27 trimethylation) and gain modifications characteristic of transcriptionally active chromatin. Binding of DNMT3B and polycomb protein SUZ12 was also lost in most locations examined. These studies, therefore, further emphasize the interrelatedness of DNA methylation and histone modifications and the ability of DNMT3B to repress transcription in the absence of *de novo* methylation. This work also provides a firm basis for better understanding the molecular defects that give rise to the ICF syndrome phenotype.

Owing to the rarity of ICF syndrome, we are limited to using immortalized LCLs for our studies. As such, the relevance of the gene expression and chromatin structure changes occurring within genes that have roles in immune function may be readily related to the immune defects of ICF patients as they are derived from the lymphoid compartment. We studied other genes, however, whose functions are unrelated to the immune system, but are important for aspects of development and CNS function. Owing to the scarcity of primary tissue from ICF syndrome patients, we are unable to prove that misregulation of the genes we have studied in lymphocytes also occurs in the tissue (neurons or the developing fetus for example) where they exert their primary effects. Such confirmation will likely require the use of the mouse model of ICF syndrome (38). We believe,

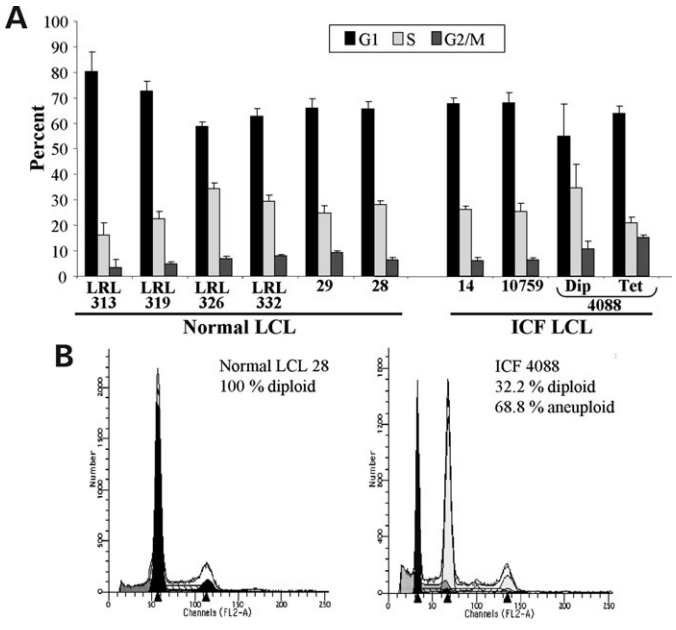

Figure 11. Cell cycle analysis of normal and ICF LCLs. (**A**) Summary of cell cycle profiles for five normal LCLs and three ICF LCLs showing the percent cells in G1 (black bars), S (light gray bars) and G2/M (dark gray bars) phases of the cell cycle. Cell lines were analyzed in triplicate with the bar showing the average and the error bar the standard deviation. (**B**) Representative raw cell cycle data for a normal LCL (GM08728) and the 4088 ICF LCL emphasizing the near tetraploid state of the 4088 cells. Dip, diploid; Tet, tetraploid.

however, that the alterations in gene expression in ICF lymphocytes we have defined are relevant and will be mirrored in other cell types of ICF patients. Support for this notion comes from a recent microarray study also using LCLs, which detected sex-specific differences in the expression of genes involved in reproductive tissues (ovary and testis) which were not simply because of differential expression of genes on the X or Y chromosome (39).

A previous study analyzed gene expression profiles in several normal and ICF LCLs, including a subset of the ICF lines we have used here. Of the 5600 genes analyzed, 32 demonstrated reproducible differential expression and were further analyzed. We found 14 of these 32 genes in the present study, including *IGHA1*, *IGHA2*, *IGHD*, *IGHG3*, *IGHM*, *CD27*/*TNFRSF7*, *TFRC*, *GUCY1B3*, *GUCY1A3*, *PTPN13*, *CNN3*, *MARK3*, *ID3* and *SLC1A1* (20). COBRA DNA methylation analysis of several of these genes revealed no changes between normal and ICF cells; however, this method interrogates the methylation status of only a few CpG sites and may have missed the relatively sparse methylation we detected with BGS. Alternatively, these particular genes are not differentially methylated, which is also in keeping with a number of the genes we analyzed, such as *ID2* and *MAP4K4*.

Our expression data have revealed a wealth of potential new genes, whose misregulation in ICF syndrome cells could contribute its main clinical features, the immune and neurologic defects and peculiar facial features. Regarding the immune defects, PANTHER ontology analysis yielded 69 genes implicated in immune function with altered expression in ICF syndrome cells, including most of the immunoglobulin heavy chains. We also identified genes implicated in T-cell and myeloid/macrophage function, the complement cascade and cytokine signaling. Multiple genes under the 'Development' category, such as homeobox gene *HHEX*, are also implicated in development of the immune system. Therefore, there is clearly an overlap between these pathway classifications. We focused our in-depth characterization on genes not previously studied that may contribute to the clinical features of ICF syndrome. For example, the *ID2* gene is an important dominant-negative regulator of basic helix–loop–helix transcription factors and is critical for development of peripheral lymphoid organs, T-cell maturation and class switch recombination to IgE (40). This latter role for ID2 suggests that reduced levels of IgE in some ICF patients may be because of elevated ID2 expression (40). IL1R2, often co-expressed with IL1R1 on T and pre-B cells, is one of the two interleukin 1 (IL-1) receptors. IL-1 is a major macrophage-derived proinflammatory cytokine important for thymocyte proliferation and B-cell growth and differentiation. IL1R2 is a potent natural inhibitor (decoy) of IL-1 that competes with IL-1 for binding to IL1R1, the receptor responsible for transducing the activating effects of IL-1. IL1R2 may therefore buffer the activity of IL-1 and too much IL1R2 may mute the effects of IL-1 and blunt certain immune responses (41).

Additional genes misregulated in ICF syndrome cells relevant to immune function that we identified include *PTPN13*, *CD200*, *LCK*, *SYK* and *CD24*. PTPN13 (PTPL1 and FAP-1), a ubiquitiously expressed tyrosine phosphatase, was also identified in a prior ICF syndrome microarray study (20). PTPN13 interacts with Fas in the cytoplasm, reducing Fas export to the cell surface with consequent reduction in Fas-mediated apoptosis (42). LCK, a Src-family tyrosine kinase downregulated in ICF syndrome cells, is expressed in T cells where it associates with the CD4 and CD8 coreceptors and participates in the signal transduction cascade that ultimately culminates in T-cell activation (43). Similarily, SYK, a non-receptor tyrosine kinase expressed in all hematopoietic cells, is essential for lymphocyte development and signal transduction via the immune receptors, such as the B- and T-cell receptors. Reduced expression of LCK and SYK could therefore contribute to defects in B- and T-cell function and maturation (44). We also showed that CD24, a GPI-anchored protein expressed on many cell types, including B cells where it is expressed in a differentiation-dependent manner, is downregulated in ICF cells. Crosslinking of CD24 induces apoptosis in B-cell precursors and inhibits proliferation in mature resting B cells in response to certain stimuli (45). Altered expression of CD24 may therefore also negatively affect immune function.

Unlike previous studies, we identified important regulators of neurogenesis, brain function and development as having altered expression in ICF cells, such as *ROBO1*, *JPH4*, *FRY*, *MAP4K4*, *PCDHGC3* and *IGF1*. *ROBO1* (roundabout homolog 1) encodes a transmembrane receptor that binds to the Slit family of secreted proteins. ROBO1/Slit interaction regulates axon branching and neuronal migration, whereas in leukocytes this interaction inhibits chemotaxis, suggesting that improper expression of ROBO1 could contribute to two aspects of ICF syndrome (46,47). In addition to *ROBO1*,

both *JPH3* and *JPH4*, which were upregulated in ICF syndrome cells, have roles in neuronal function. The junctophilin family of proteins, JPH1-4, contributes to the formation of junctional membrane complexes by spanning the sarcoplasmic reticulum membrane and binding with the plasma membrane in muscle cells. They are also expressed in the brain, and *Jph3*/*Jph4* double knockout mice display severe growth retardation, an aberrant foot-clasping reflex indicative of a neurologic defect, and impaired calcium signaling in neurons (48). It should be noted that many of the studies deciphering gene function cited here made use of knockout mice, which contrasts our results that generally show increased expression of brain developmental genes in ICF cells. Effects of overexpression may have different functional and phenotypic consequences; however, given that brain development is a complex highly ordered process, it is likely that too much of a gene product may be as detrimental as too little. *IGF1* (insulin-like growth factor-1), downregulated in ICF syndrome cells, is an essential factor for growth and development of the mammalian brain. It stimulates neurogenesis during embryonic development and inhibits apoptosis of neurons and oligodendrocytes postnatally (49). Assuming that our finding of reduced IGF1 expression in ICF LCLs can be extrapolated to cells of the brain, the reduced IGF1 levels during development may contribute to the neurological deficits found in a significant proportion of ICF syndrome patients.

One of the more significant findings from this work relating in a general sense to the role of DNMTB in development and its potential interplay with polycomb proteins was the number of non-clustered homeobox-containing transcriptional regulators with altered expression in ICF syndrome cells. These include three of the genes we characterized in detail, *LHX2*, *PRRX1* and *HHEX*, as well as *ONECUT2*, *SIX3*, *CBX8*, *HMX2* and *ISL2* (Supplementary Material, Table S1). *LHX2* (LIM homeobox gene family) is highly expressed in B cells at early stages of differentiation and in the developing brain (50). Knockout of the *Lhx2* gene in mice revealed it is essential for development of the eye, cerebral cortex, liver, and for erythropoiesis (51). *HHEX* (hematopoietically expressed homeobox), expressed in hematopoietic progenitors, myeloid and liver cells, also plays a multifunctional role in development. It is involved in B-cell differentiation (52), forebrain patterning and development of cells of the liver, thyroid and hematopoietic lineages (53). Mice deficient in *Hhex* display reduced numbers of mature and pre-B cells, with B-cell development almost completely arrested at the pro-B cell stage (52). *PRRX1* (*Mhox*/*Phox*, paired-related homeobox gene-1) is expressed in the postmigratory cranial mesenchyme of all facial prominences and is required for proper formation of the proximal first arch derivatives (the precursor of the lower jaws). *Prrx1* null mice display various skeletal defects, particularly in the craniofacial region (34). PRRX1 and PRRX2 may regulate cell proliferation in critical areas during development via sonic hedgehog (34). We found that, similar to the effects of ICF *DNMT3B* mutations on PRRX1 expression, SHH expression was also markedly upregulated and it will be of interest to study the integrity of this pathway in ICF cells in future studies. Taken together, these results suggest that the loss of DNMT3B function causes derepression of a subset of important developmental regulators.

Many homeobox genes are repressed in pluripotent cells, which is necessary to prevent differentiation. In this state, they are bivalently marked with both active H3K4 trimethylation and repressive H3K27 trimethylation marks and so are thought to be poised for full activation or silencing (54). During differentiation, these genes are temporally and spatially activated then repressed with or without the addition of promoter DNA methylation. Polycomb complexes PRC1 and PRC2 are key developmental regulators during embryogenesis, repressing large numbers of genes (~8–14% depending on cell type) including homeobox genes and genes involved in developmental processes (32,55,56). The PRC2 complex, consisting of EED, EZH2, SUZ12 and RbAp46/48, mediates H3K27 trimethylation via its EZH2 subunit. This histone mark is thought to recruit the PRC1 complex, composed of RING1/2, HPH proteins, BMI1 and HPC proteins. HPC proteins contain a chromodomain that specifically binds to the H3K27 trimethylation mark to establish long-term silencing. In addition, the PRC1 complex mediates H2AK119 monoubiquitination via the RING1 and BMI1 subunits (57).

Exciting connections are emerging that link regulation of DNA methylation to polycomb complexes. For example, DNMT1, DNMT3A and DNMT3B interact with EZH2 and this interaction is capable of recruiting DNA methylation to certain genomic regions (23). Connections between DNMTs and polycomb may be more complicated, however, since EZH2 knockdown does not result in loss of DNA methylation from promoters in tumor cells (58) and DNMT1 and DNMT3A interact with the BMI1 and HPC2 subunits, respectively, of the PRC1 complex (59,60). Genes regulated by polycomb appear to be preferential targets for tumor-specific *de novo* methylation, potentially because of the polycomb–DNMT interactions or because they mistakenly acquire other repressive histone marks, such as H3K9 methylation (61,62). Consistent with this hypothesis, several genes we studied, such as *PTPN13*, *ROBO1*, *LHX2* and *PRRX1*, acquire dense promoter hypermethylation in tumor cells (63–65). In keeping with the role of PRC1 and PRC2 in repression, we found that all genes upregulated in ICF syndrome cells that we examined were bound by PRC2 (represented by SUZ12) in normal LCLs and this binding was reduced or absent in ICF cells regardless of whether there was a loss of DNA methylation. Also consistent with recent reports, many of the genes upregulated in ICF syndrome cells that we characterized, such as homeobox genes *HHEX*, *LHX2* and *PRRX1* and non-homeobox genes like *PTPN13*, *JPH4*, *ROBO1* and *ALOX5*, were occupied by PRC2 and/or PRC1 complexes in human ES or embryonic fibroblasts (32,56). Likely, these genes are activated during critical periods of development and, at least in lymphocytes, become repressed at a later stage of differentiation. Mutations in *DNMT3B* may directly affect the repressive functions of PRC1 and/or PRC2 complexes by altering their recruitment or their ability to stably maintain repression. Alternatively, the loss of full DNMT3B function may have indirect effects on polycomb function though disruption of pericentromeric heterochromatin where components of the PRC1 complex localize (66). In either case, both local and global changes in histone modifications result.

In conclusion, using gene expression profiling of nearly the entire human transcriptome in ICF and normal LCLs, we

identified a large number of genes that are aberrantly expressed in ICF syndrome cells related to immune function, development and neurogenesis, which provide important new insights into the molecular defects underlying the ICF syndrome phenotype. This study revealed for the first time that there are marked changes in patterns of histone modifications and loss of PRC2 and DNMT3B binding at promoters of genes upregulated in ICF cells. Therefore, this work further strengthens the emerging connections between the DNA methylation and histone modification machineries and polycomb repression complexes. It also suggests that normal DNMT3B DNA methylation and/or repression functions are required for proper gene repression and polycomb binding.

MATERIALS AND METHODS

Cell lines, drug treatments and reagents

Epstein–Barr virus-immortalized LCLs used in this study include the ICF syndrome lines GM08714, 4088 and 10759 (see next section for description of origin and *DNMT3B* mutation) and cell lines derived from normal individuals GM08729, GM08728, LCL1, CM304, CM774, LRL-332, LRL-330, LRL-313, LRL-319 and LRL-326 ('LRL' lines were provided by Dr. R.F. Ambinder). ICF LCLs were grown in RPMI1640 media (Mediatech) supplemented with 2 mM L-glutamine (Mediatech) and 20% heat-inactivated fetal bovine serum (Hyclone). Normal LCLs were grown in the same media except 10% fetal bovine serum was used. 5-Aza-2′-deoxycytidine and TSA were purchased from Sigma. For drug treatments, 5-azadC was added to cultures at a final concentration of 5 μM for 3 days (fresh drug added each day); TSA treatments were done for 24 h at 100 nM final concentration.

Microarray expression profiling

RNA samples, probe preparation, hybridization and scanning. Lymphoblastoid cell lines were grown under the conditions described above. ICF syndrome cell lines are GM08714 [female, heterozygous A603T and intron 22 G to A mutation resulting in insertion of three amino acids (STP) in *DNMT3B*, Coriell Cell Repository], 4088 (male, homozygous V726G mutation in *DNMT3B*, provided by Dr. D. Smeets) and 10759 (male, P670T mutation in *DNMT3B*, provided by Dr. R.S. Hansen) (8,10,14). LCLs derived from normal individuals include GM08729 and GM08728 (father and mother, respectively, of ICF patient GM08714, Coriell Cell Repository), and LCL1, CM304 and CM774 (14). Total RNA from each cell line was prepared using Trizol according to the manufacturer's protocol (Invitrogen). Three independently prepared replicates of each cell line harvested within 2 weeks of each other were used for analysis. Each replicate RNA was independently labeled and hybridized to both the Affymetrix U133A and U133B GeneChips according to the manufacturer's instructions (Affymetrix) and as we have described previously (67).

Data analysis

Probe Profiler software (ver. 2.1.0) (Corimbia Inc.) was used to convert Microarray Suite Version 5 software (Affymetrix)-generated*.cel files into quantitative estimates of gene expression (EScores). All expression values were globally scaled to 100. On each chip, genes were ranked according to their expression value. The signal value associated with the highest ranked absent call was assigned as the signal value of all genes with an absent call and thus served as the absent/present threshold. Genes not expressed in any of the samples ($P > 0.05$) were removed from the data set. Genes were ranked according to the variability (coefficient of variation) of their signal values across all arrays. Values for each triplicate sample were averaged. PCA was performed on the most variable 0.01% of genes to assess heterogeneity among patients (GeneLinker Gold). A one-way ANOVA was performed for two treatment groups: diseased ($n = 3$) and normal ($n = 5$). Informative genes were identified as those with a significant treatment effect ($P < 0.05$) and having greater than a 2-fold difference in treatment means. The expression value of these informative genes was normalized by performing a *Z*-transformation. Hierarchical clustering was performed on the *Z*-transformed expression values with TreeView. Gene ontology and pathway analysis was performed using PANTHER (http://www.pantherdb.org/) (18).

Reverse transcriptase–polymerase chain reaction

Reverse transcriptase–polymerase chain reaction was carried out as described previously (67). Briefly, untreated and treated cell lines were homogenized in Trizol and the RNA purified according to the manufacturer's instructions (Invitrogen). First-strand cDNA synthesis was carried out using Superscript III RT (Invitrogen). Subsequently, the cDNA was used in semi-quantitative PCR with the primers listed in Supplementary Material, Table S7. All reactions were performed at least three times. Following PCR, reaction products were resolved on 2% agarose gels and photographed using a BioRad gel documentation system.

Bisulfite genomic sequencing

Bisulfite genomic sequencing was performed as described previously (67). For a complete listing of PCR primer sequences used for BGS, refer to Supplementary Material, Table S7. TaqGold (ABI) thermostable DNA polymerase was used for all reactions. The band was purified from the agarose gel using the Qiaex II gel extraction kit (Qiagen) and cloned using the TA Cloning Kit (Invitrogen). Products from at least two independent PCR reactions were cloned and sequenced in a 96-well plate format using the M13 reverse and/or forward primers. All sequencing was performed at the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR).

Chromatin immunoprecipitation (ChIP) and western blotting

ChIP was performed essentially as described previously (68). Cross-linked chromatin was sheared by sonication using a Branson Sonifier 450 (tip model 102, output 5, 50% duty, 12 × 1 min bursts). The supernatant from the irrelevant antibody served as a positive control ('input', 1% of the ChIP

material). Purified, immunoprecipitated DNA was analyzed by semi-quantitative PCR. All PCR reactions were repeated at least three times. ChIP PCR primer sequences are provided in Supplementary Material, Table S7. For western blotting, cells were fractionated as described previously (69) and the pooled soluble nuclear and chromatin fractions were used. Antibodies used in ChIP and western blotting experiments (for ChIP 'C', amount used/reaction; for western blotting 'W', the dilution used) are: normal rabbit IgG (Pierce, C-10 μg, non-specific antibody control for ChIP), DNMT3B (Novus, C-10 μg, W-1:500), Ac-H3K9/K18 (Upstate, C-15 μl, W-1:2000), 3X-MeH3K4 (Abcam, C-10 μg, W-1:500), 3X-MeH3K9 (Abcam, C-5 μg, Upstate, W-1:1000), 3X-MeH3K27 (Abcam, C-10 μg, Upstate W-1:5000), 3X-MeH4K20 (Upstate, C-10 μg, W-1:500), SUZ12 (Upstate, C-10 μl, W-1:1000), Ub-H2AK119 (Upstate, C-10 μg, W-1:500), total histone H3 (Upstate, W-1:500), BMI1 (Santa Cruz, W-1:500), 2XMe-H3K4 (Upstate, W-1:2000) and 2XMe-H3K27 (Abcam, C-6 μg, W-1:800).

Cell cycle analysis and apoptosis

Cell cycle profiles were determined by analyzing DNA content using propidium iodide (PI) staining and flow cytometry. Cells were harvested, washed with 1× PBS, and fixed with ice cold 70% ethanol overnight. Fixed cells were washed once with 1× PBS and resuspended at 1×10^6 cells/ml in PI staining solution (0.2 mg/ml PI, and 0.2 mg/ml RNAase in PBS) and incubated in the dark at room temperature for 15 min before analysis. Cell cycle profiles were determined using fluorescence-activated cell sorting (FACS) with a Becton Dickinson FACSort. For each sample, 3×10^4 events were recorded. Data were analyzed by ModFit Cell Cycle Analysis Software (Verity) to determine the percentage of cells in each phase. Apoptosis was measured with the Annexin V-PI kit according to the manufacturer's protocol (Trevigen).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank Drs. Richard Ambinder, R. Scott Hansen, Dominique Smeets, Chan Soh-Ha and Wen-son Hsieh for providing cell lines, and Li Liu, Mick Popp and Jason Li for assistance with the microarray data analysis.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by NIH grants K22CA084535, R01CA114229 (K.D.R.), a Johns Hopkins Singapore Grant, and the Chinese University of Hong Kong (Q.T.).

REFERENCES

1. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.*, **3**, 662–673.
2. Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
3. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
4. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases *Dnmt3a* and *Dnmt3b* are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
5. Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
6. Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
7. Ehrlich, M., Jackson, K. and Weemaes, C. (2006) Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). *Orphanet J. Rare Dis.*, **1**: 2.
8. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M.R. and Gartler, S.M. (1999) The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA*, **96**, 14412–14417.
9. Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Pequignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187–191.
10. Smeets, D.F.C.M., Moog, U., Weemaes, C.M.R., Vaes-Peeters, G., Merkx, G.F.M., Niehof, J.P. and Hamers, G. (1994) ICF syndrome: a new case and review of the literature. *Hum. Genet.*, **94**, 240–246.
11. Ehrlich, M. (2003) The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clin. Immunol.*, **109**, 17–28.
12. Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hulten, M., Jonveaux, P., Maraschio, P., Megarbane, A. *et al.* (2005) DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum. Mut.*, **25**, 56–63.
13. Kondo, T., Bobek, M.P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., Bourc'his, D., Viegas-Pequignot, E., Ehrlich, M. and Hanash, S.M. (2000) Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats *D4Z4* and *NBL2*. *Hum. Mol. Genet.*, **9**, 597–604.
14. Tao, Q., Huang, H., Geiman, T.M., Lim, C.Y., Fu, L., Qiu, G.-H. and Robertson, K.D. (2002) Defective *de novo* methylation of viral and cellular DNA sequences in ICF syndrome cells. *Hum. Mol. Genet.*, **11**, 2091–2102.
15. Hansen, R.S., Stoger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D' Esposito, M., Feil, R., Gimelli, G. *et al.* (2000) Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum. Mol. Genet.*, **9**, 2575–2587.
16. Geiman, T.M., Sankpal, U.T., Robertson, A.K., Chen, Y., Mazumdar, M., Heale, J.T., Schmiesing, J.A., Kim, W., Yokomori, K., Zhao, Y. *et al.* (2004) Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.*, **32**, 2716–2729.
17. Narayan, A., Ji, W., Zhang, X.-Y., Marrogi, A., Graff, J.R., Baylin, S.B. and Ehrlich, M. (1998) Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int. J. Cancer*, **77**, 833–838.
18. Mi, H., Guo, N., Kejariwal, A. and Thomas, P.D. (2007) PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.*, **35**, D247–D252.
19. Ehrlich, M. (2002) DNA hypomethylation, cancer, the Immunodeficiency, Centromeric Region Instability, Facial Anomalies Syndrome and chromosomal rearrangements. *J. Nutr.*, **132**, 2424S–2429S.
20. Ehrlich, M., Buchanan, K.L., Tsein, F., Jiang, G., Sun, B., Uicker, W., Weemaes, C.M.R., Smeets, D., Sperling, K., Belohradsky, B.H. *et al.* (2001) DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphomagenesis genes. *Hum. Mol. Genet.*, **10**, 2917–2931.
21. Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.

22. Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosome 21 and 22. *Proc. Natl. Acad. Sci. USA*, **99**, 3740–3745.
23. Vire, E., Brenner, C., Delpus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Van der winden, J.-M. *et al.* (2006) The polycomb group protein EZH2 directly controls DNA methylation. *Nature*, **439**, 871–874.
24. Cameron, E.E., Bachman, K.E., Myohanen, S., Herman, J.G. and Baylin, S.B. (1999) Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat. Genet.*, **21**, 103–107.
25. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (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.
26. Lorincz, M.C., Dickerson, D.R., Schmitt, M. and Groudine, M. (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Mol. Biol.*, **11**, 1068–1075.
27. Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Nailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., III, Gingeras, T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
28. Roh, Y.-Y., Cuddapah, S., Cui, K. and Zhao, K. (2006) The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. USA*, **103**, 15782–15787.
29. Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
30. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. and Reinberg, D. (2003) Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of Zeste protein. *Genes Dev.*, **16**, 2893–2905.
31. Imamura, M., Miura, K., Iwabuchi, K., Ichisaka, T., Nakagawa, M., Lee, J., Kanatsu-Shinohara, M., Shinohara, T. and Yamanaka, S. (2006) Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.*, **6**.
32. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. and Helin, K. (2006) Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.*, **20**, 1123–1136.
33. Tang, X., Guilherme, Z., Chakladar, A., Powelka, A.M., Konda, S., Virbasius, J.V., Nicoloro, S.M.C., Straubhaar, J. and Czech, M.P. (2006) An RNA interference-based screen identifies MAP4K4/NIK as a negative regulator of PPARγ, adipogenesis, and insulin-responsive hexose transport. *Proc. Natl. Acad. Sci. USA*, **103**, 2087–2092.
34. ten Berge, D., Brouwer, A., Korving, J., Reijnen, M.J., van Raaij, E.J., Verbeek, F., Gaffield, W. and Meijlink, F. (2001) Prx1 and Prx2 are upstream regulators of sonic hedgehog and control cell proliferation during mandibular arch morphogenesis. *Development*, **128**, 2929–2938.
35. Hu, D. and Helms, J.A. (1999) The role of sonic hedgehog in normal and abnormal craniofacial morphogenesis. *Development*, **126**, 4873–4884.
36. Narayan, A., Tuck-Muller, C., Weissbecker, K., Smeets, D. and Ehrlich, M. (2000) Hypersensitivity to radiation-induced non-apoptotic and apoptotic death in cell lines from patients with the ICF chromosome instability syndrome. *Mutat. Res.*, **456**, 1–15.
37. Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N. and Li, E. (2005) Inactivation of *Dnmt3b* in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J. Biol. Chem.*, **280**, 17986–17991.
38. Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K. and Li, E. (2006) Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development*, **133**, 1183–1192.
39. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
40. Sugai, M., Gonda, H., Kusunoki, T., Katakai, T., Yokota, Y. and Shimizu, A. (2003) Essential role of Id2 in negative regulation of IgE class switching. *Nat. Immunol.*, **4**, 25–30.
41. Subramaniam, S., Stansberg, C. and Cunningham, C. (2004) The interleukin 1 receptor family. *Dev. Comp. Immunol.*, **28**, 415–428.
42. Ivanov, V.N., Bergami, P.L., Maulit, G., Sato, T.-A., Sassoon, D. and Ronai, Z. (2003) FAP-1 association with Fas (Apo-1) inhibits Fas expression on the cell surface. *Mol. Cell Biol.*, **23**, 3623–3635.
43. Palacios, E.H. and Weiss, A. (2004) Function of the src-family kinases, lck and fyn, in T-cell development and activation. *Oncogene*, **23**, 7990–8000.
44. Turner, M., Schweighoffer, E., Colucci, F., Di Santo, J.P. and Tybulewicz, V.L. (2000) Tyrosine kinase SYK: essential functions for immunoreceptor signalling. *Immunol. Today*, **21**, 148–154.
45. Suzuki, T., Kiyokawa, N., Taguchi, T., Sekino, T., Katagiri, Y.U. and Fujimoto, J. (2001) CD24 induces apoptosis in human B cells via the glycolipid-enriched membrane domains/rafts-mediated signaling system. *J. Immunol.*, **166**, 5567–5577.
46. Andrews, W., Liapi, A., Plachez, C., Camurri, L., Zhang, J., Mori, S., Murakami, F., Parnavelas, J.G., Sundaresan, V. and Richards, L.J. (2006) Robo1 regulates the development of major axon tracts and interneuron migration in the forebrain. *Development*, **133**, 2243–2252.
47. Wu, J.Y., Feng, L., Park, H.-T., Havlioglu, N., Wen, L., Tang, H., Bacon, K.B., Jiang, Z.-h., Zhang, X.-c. and Rao, Y. (2001) The neuronal repellent Slit inhibits leukocyte chemotaxis induced by chemotactic factors. *Nature*, **410**, 948–952.
48. Moriguchi, S., Nishi, M., Komazaki, S., Sakagami, H., Miyazaki, T., Masumiya, H., Saito, S.-y., Watanabe, M., Kondo, H., Yawo, H. *et al.* (2006) Functional uncoupling between Ca^{2+} release and afterhyperpolarization in mutant hippocampal neurons lacking junctophilins. *Proc. Natl. Acad. Sci. USA*, **103**, 10811–10816.
49. D'Ercole, A.J., Ye, P., Calikoglu, A.S. and Gutierrez-Ospina, G. (1996) The role of the insulin-like growth factors in the central nervous system. *Mol. Neurobiol.*, **13**, 227–255.
50. Xu, Y., Baldassare, M., Fisher, P., Rathbun, G., Oltz, E.M., Yancopoulos, G.D., Jessell, T.M. and Alt, F.W. (1993) LH-2: A LIM/homeodomain gene expressed in developing lymphocytes and neural cells. *Proc. Natl. Acad. Sci. USA*, **90**, 227–231.
51. Porter, F.D., Drago, J., Xu, Y., Cheema, S.S., Wassif, C., Huang, S.-P., Lee, E., Grinberg, A., Massalas, J.S., Bodine, D. *et al.* (1997) Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development. *Development*, **124**, 2935–2944.
52. Bogue, C.W., Zhang, P.-X., McGrath, J., Jacobs, H.C. and Fuleihan, R.L. (2003) Impaired B cell development and function in mice with a targeted disruption of the homeobox gene. *Hex. Proc. Natl. Acad. Sci. USA*, **100**, 556–561.
53. Martinez-Barbera, J.P., Clements, M., Thomas, P., Rodriguez, T., Meloy, D., Kioussis, D. and Beddington, R.S.P. (2000) The homeobox gene *Hex* is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development*, **127**, 2433–2445.
54. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
55. Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K. *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, **441**, 349–353.
56. Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-i. *et al.* (2006) Control of developmental regulators by polycomb in human embryonic stem cells. *Cell*, **125**, 301–313.
57. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. and Cavalli, G. (2007) Genome regulation by polycomb and trithorax proteins. *Cell*, **128**, 735–745.
58. McGarvey, K.M., Greene, E., Fahrner, J.A., Jenuwein, T. and Baylin, S.B. (2007) DNA methylation and complete transcriptional silencing of cancer genes persists after depletion of EZH2. *Cancer Res.*, **67**, 5097–5102.
59. Hernandez-Munoz, I., Taghavi, P., Kuijl, C., Neefjes, J. and van Lohuizen, M. (2005) Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol. Cell. Biol.*, **25**, 11047–11058.
60. Li, B., Zhou, J., Liu, P., Hu, J., Jin, H., Shimono, Y., Takahashi, M. and Xu, G. (2007) Polycomb protein CBX4 promotes SUMO modification of *de novo* DNA methyltransferase Dnmt3a. *Biochem. J.*, **405**, 369–378.
61. Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W. *et al.* (2007) A stem

cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.*, **39**, 237–242.
62. Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I. *et al.* (2007) Epigenetic stem cell signature in cancer. *Nature Genet.*, **39**, 157–158.
63. Dallol, A., Forgacs, E., Martiez, A., Sekido, Y., Walker, R., Kishida, T., Rabbitts, P., Maher, E.R., Minna, J.D. and Latif, F. (2002) Tumour specific promoter region methylation of the human homologue of the *Drosophila* roundabout gene DUTT1 (ROBO1) in human cancers. *Oncogene*, **21**, 3020–3028.
64. Ying, J., Li, H., Cui, Y., Wong, A.H.Y., Langford, C. and Tao, Q. (2006) Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter. *Leukemia*, **20**, 1173–1175.
65. Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S.K., Kernstine, K.H., Riggs, A.D. and Pfeifer, G.P. (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc. Natl. Acad. Sci. USA*, **104**, 5527–5532.
66. Saurin, A.J., Shiels, C., Williamson, J., Satijn, D.P.E., Otte, A.P., Sheer, D. and Freemont, P.S. (1998) The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J. Cell. Biol.*, **142**, 887–898.
67. Kim, T.-Y., Zhong, S., Fields, C.R., Kim, J.H. and Robertson, K.D. (2006) Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res.*, **66**, 7490–7501.
68. Zhong, S., Fields, C.R., Su, N., Pan, Y.-X. and Robertson, K.D. (2007) Pharmacologic inhibition of epigenetic modifications, coupled with gene expression profiling, reveals novel targets of aberrant DNA methylation and histone deacetylation in lung cancer. *Oncogene*, **26**, 2621–2634.
69. Mendez, J. and Stillman, B. (2000) Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.*, **20**, 8602–8612.