

# Genome-wide DNA methylation analysis in cohesin mutant human cell lines

Jinglan Liu<sup>1</sup>, Zhe Zhang<sup>2</sup>, Masashige Bando<sup>3</sup>, Takehiko Itoh<sup>3</sup>, Matthew A. Deardorff<sup>1,4</sup>, Jennifer R. Li<sup>1</sup>, Dinah Clark<sup>1</sup>, Maninder Kaur<sup>1</sup>, Kondo Tatsuro<sup>5</sup>, Antonie D. Kline<sup>6</sup>, Celia Chang<sup>7</sup>, Hugo Vega<sup>8</sup>, Laird G. Jackson<sup>9</sup>, Nancy B. Spinner<sup>1,4</sup>, Katsuhiko Shirahige<sup>3</sup> and Ian D. Krantz<sup>1,4,\*</sup>

<sup>1</sup>Division of Human Genetics, Abramson Research Institute, <sup>2</sup>Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, <sup>3</sup>Laboratory of Chromosome Structure and Function, Department of Biological Science, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, B2C 4259, Nagatsuta, Midori-ku, Yokohama City, Kanagawa 226-8501, Japan, <sup>4</sup>The University of Pennsylvania School of Medicine, PA 19104, USA, <sup>5</sup>Division of Developmental Disability, Misakaenosono Mutsumi Developmental, Medical, and Welfare Center, Konagai-cho Maki 570-1, Isahaya, 859-0169, Japan, <sup>6</sup>Harvey Institute for Human Genetics, Department of Pediatrics, Greater Baltimore Medical Center, Baltimore, MD 21204, <sup>7</sup>Genomic and Microarray Facility, the Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA, <sup>8</sup>Department of Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA and <sup>9</sup>Department of Obstetrics and Gynecology, Drexel University School of Medicine, Philadelphia, PA 19104, USA

Received January 1, 2010; Revised April 19, 2010; Accepted April 21, 2010

## ABSTRACT

The cohesin complex has recently been shown to be a key regulator of eukaryotic gene expression, although the mechanisms by which it exerts its effects are poorly understood. We have undertaken a genome-wide analysis of DNA methylation in cohesin-deficient cell lines from probands with Cornelia de Lange syndrome (CdLS). Heterozygous mutations in *NIPBL*, *SMC1A* and *SMC3* genes account for ~65% of individuals with CdLS. *SMC1A* and *SMC3* are subunits of the cohesin complex that controls sister chromatid cohesion, whereas *NIPBL* facilitates cohesin loading and unloading. We have examined the methylation status of 27 578 CpG dinucleotides in 72 CdLS and control samples. We have documented the DNA methylation pattern in human lymphoblastoid cell lines (LCLs) as well as identified specific differential DNA methylation in CdLS. Subgroups of CdLS probands and controls can be classified using selected CpG loci. The X chromosome was also found to have a unique DNA methylation pattern in CdLS. Cohesin preferentially binds to hypo-methylated DNA in control LCLs, whereas the differential DNA methylation alters

cohesin binding in CdLS. Our results suggest that in addition to DNA methylation multiple mechanisms may be involved in transcriptional regulation in human cells and in the resultant gene misexpression in CdLS.

## INTRODUCTION

Vertebrate gene expression is tightly regulated at several levels which are mechanistically linked to each other (1). DNA methylation, histone modification and chromatin remodeling are the most well-recognized and closely interwoven epigenetic events (2). Epigenetic regulation during development may occur very early during embryogenesis, driving the formation of different organ systems (3). DNA methylation is maintained by methyltransferase DNMT1 in both mitosis and meiosis, and is considered the most stable epigenetic mark (4).

In mammals, DNA methylation predominantly occurs at CpG dinucleotides by covalent addition of a methyl group to position 5 of the cytosine ring, creating 5-methylcytosine. CpG dinucleotides are under-represented in the human genome with a frequency of 2–5% as compared to the GC content (5). CpG dinucleotides are not equally distributed throughout the human genome; instead, they occur in clusters of large repetitive

\*To whom correspondence should be addressed. Tel: +1 215 590 2931; Fax: +1 215 590 3850; Email: ian2@mail.med.upenn.edu

sequences [such as ribosomal DNA (rDNA), satellite sequences or centromeric repeats] or in short CG-rich DNA stretches, known as CpG islands (CGIs) (6). Dinucleotide clusters of CpGs or 'CpG islands' are present in the promoter and exonic regions of ~40–70% of mammalian genes and these clusters are usually unmethylated (7,8). By contrast, other regions of the mammalian genome contain less CpG dinucleotides and the majority (75%) of these sparsely located CpG dinucleotides are largely methylated (9). A large number of experiments have shown that methylation of promoter CpG islands plays an important role in gene expression, genomic imprinting, X-chromosome inactivation, genomic instability, embryonic development and carcinogenesis (10,11).

Four DNA methyltransferases (DNMTs) (DNMT1, DNMT2, DNMT3A and DNMT3B) and one DNMT-related protein (DNMT3L) have been identified (12). DNMT1 acts as a maintenance methyltransferase, whereas DNMT2 may be an RNA methyltransferase, and DNMT3a and DNMT3b are *de novo* methyltransferases targeting unmethylated DNA. All DNMTs are essential for embryonic viability with homozygous mutant mice dying early in development (13). MBD1-4 proteins or methyl CpG-binding proteins (MeCP2) recognize and bind to methylated DNA. They recruit transcriptional corepressors such as histone-deacetylating complexes and polycomb group (PcG) proteins, associate with chromatin-remodeling complexes and attract chromodomain-binding proteins (13).

DNA methylation and chromatin structure are strikingly altered in many pathological situations, particularly in cancers and various mental retardation syndromes. Altered levels of the methyl donor folate and homocysteine have been repeatedly linked to these disorders. Disease-associated changes in epigenetic modifications can be classified into changes in genes that are epigenetically regulated and in genes that are part of the molecular machinery establishing and propagating the epigenetic modifications through the development and cell divisions. Aberrant methylation patterns have been reported in various neurodevelopmental disorders, including X-linked  $\alpha$ -thalassemia and mental retardation (ATRX), Fragile X, and immune deficiency, centromeric instability and facial abnormalities (ICF) (14). Interestingly, the *ATRX* gene is misexpressed in CdLS [+1.2, false discovery rate (FDR) = 0.07] and the disorder ATRX presents defective sister chromatid cohesion and is considered as one of the 'cohesinopathies' (15).

Cornelia de Lange syndrome (CdLS, OMIM#122470, 300590, 610759) is the first identified 'cohesinopathy' which is a heterogeneous dominantly inherited developmental disorder with multiple-organ system involvement (16–19). The majority of CdLS probands were found to have heterozygous mutations in the *NIPBL* gene, whereas a small percentage have mutations in the *SMC1A* and *SMC3* genes. *SMC1A* and *SMC3* are core components of the cohesin complex which controls sister chromatid cohesion during S phase, while *NIPBL* facilitates cohesin loading and unloading (20,21). In addition to cohesin and *NIPBL*'s canonical role in regulating sister chromatid cohesion, they have also been implicated as key regulators

of gene expression over long distances (22,23), and have been shown to preferentially associate to actively transcribed genes and colocalize with RNA polymerase II in *Drosophila* (24). In humans, cohesin colocalizes with CTCF and regulates gene expression at the imprinting *IGF2/H19* locus (25). Moreover, we recently performed a genome-wide transcription and cohesin-binding study on CdLS cell lines and identified a CdLS-specific expression profile from *NIPBL* and cohesin mutant individuals. Our data also suggest that cohesin preferentially binds to transcription start sites (TSSs) and tightly correlates to transcriptional activation in humans. Loss of cohesin binding occurs in CdLS and correlates to dysregulated gene expression (26). We undertook a genome-wide DNA methylation analysis on cell lines from probands with CdLS for the following reasons: First, the colocalization of CTCF and cohesin suggests possible functional overlap between these two proteins. A growing body of evidence suggests that CTCF is involved in regulating DNA methylation in vertebrates (14), and it binds to diverse DNA sequences including most imprinting center regions and many CpG islands. CTCF's binding to chromatin seems sensitive to DNA methylation, because it not only reads DNA-methylation marks but also has a role in determining DNA methylation patterns. Second, *NIPBL* was reported to be involved in chromatin remodeling by direct association with HP1, linking *NIPBL* to multiple enzymes and protein factors that determine the histone modification patterns that tightly correlate with DNA methylation status (27,28). Third, a qualitative analysis of cohesin binding peaks in control and CdLS cells using data obtained from our previous chromatin immunoprecipitation (ChIP)-chip assay (26) has revealed 6.85% of cohesin peaks overlap CpG islands, while 7.77% of the lost peaks in CdLS overlap CpG islands, hence CpG islands are overrepresented by 6.35% ( $P = 0.0018$ ) among cohesin peaks that are lost in *NIPBL* mutant CdLS cells. Fourth, we have identified differentially expressed genes in CdLS from our previous genome-wide studies (using a relatively loosened cutoff of  $FDR < 0.2$ ), multiple of which are critical in DNA methylation. Of these, there are several proteins directly involved in DNA methylation, such as the *de novo* DNA methyltransferase DNMT1 (−1.11,  $FDR = 0.15$ ), the universal methyl-binding domain proteins MBD1 (−1.23,  $FDR = 0.02$ ) and MeCP2 (+1.11,  $FDR = 0.06$ ). In addition, PcG proteins which form the polycomb repressive complexes 1 and 2 (PRC1 and PRC2) are also dysregulated in CdLS. For example, CBX7 (+1.2,  $FDR = 0.13$ ) and BMI1 (+1.36,  $FDR = 0.02$ ) in PRC1, EZH2 (−1.2,  $FDR = 0.05$ ) and SUZ12 (−1.13,  $FDR = 0.199$ ) in PRC2. Of note, Enhancer of Zeste homolog 2 (EZH2) was suggested to serve as a recruitment platform for DNA methyltransferases because it interacts with methyltransferases (DNMTs) and associates with DNMT activity *in vivo* (29). Another example is GADD45A (+1.54,  $FDR = 0.057$ ) ectopic expression of which leads to the reduction in methylation at both specific gene loci and the total cellular 5-methylcytosine content (30). Fifth, SmcHD1-a SMC hinge domain

containing protein, maintains hypermethylation on the inactivated X chromosome in mice (31); and a second SMC-like protein-DMS3 mediates RNA-directed DNA methylation (RdDM) in plants (32), functional identification of these two proteins therefore links cohesin to DNA methylation. Sixth, 'genomic neighborhood diseases' have recently been proposed (33), and are composed of genes that are located in the same chromatin domain, are co-expressed and communicate through three dimensional structures. Dysregulated epigenetic events such as DNA methylation are tightly involved in gene mis-expression and the onset of these diseases. It is not known whether mutations in *NIPBL* or cohesin might be able to alter global DNA methylation and affect genomic organization, which may contribute to transcriptional dysregulation in CdLS.

Very little is known about epigenetic regulation by cohesin. Similarly, the role of epigenetic modulation in CdLS and the majority of other human developmental disorders are poorly characterized. To date, there has been no global assessment of overall DNA methylation in cohesin or *NIPBL* mutant human cells. We applied a comprehensive DNA methylation profiling approach to assess the epigenetic state in CdLS. We asked whether these mutant cells differed from healthy controls in terms of DNA methylation. We used an array-based method to quantitatively measure the methylation levels of 27 578 CpG sites in the regulatory regions of 14 495 genes in the human genome. We have identified differential methylation patterns in CdLS probands and also provided an integrative whole-genome view on DNA methylation, gene expression and cohesin binding in CdLS. We suggest CdLS has its own epigenetic signature that is formed in the early stage of embryonic development, which likely contributes to its clinical features.

## MATERIALS AND METHODS

### Sample collection

Sixty-three lymphoblastoid cell lines (LCLs) from 39 CdLS probands, two Roberts syndrome (RBS) probands and 22 gender- and race-matched healthy controls were tested. In addition, triplicates of one universally methylated DNA control (Zymo Research) and triplicates of each of two universally unmethylated DNA controls were also included (CHEMICON). All the tested CdLS probands have identified gene mutations and well-documented clinical features, including 22 severely affected probands with *NIPBL* mutations, eight mildly affected probands with *NIPBL* mutations, eight mildly affected probands with *SMC1A* mutations and one mildly affected proband with an *SMC3* mutation. These samples include most of the individuals studied in our previous gene expression project (26). All human subjects participating in this study were enrolled under an institutional review board-approved protocol of informed consent at The Children's Hospital of Philadelphia and Misakaenosono Mutsumi Developmental, Medical, and Welfare Center. All

subjects were evaluated by one or more experienced clinicians. Gene mutations were confirmed by sequencing.

### Cell culture and DNA isolation

LCLs were grown uniformly in RPMI 1640 with 20% fetal bovine serum (FBS), 100 U penicillin/ml, 100 µg streptomycin/ml sulfate and 1% L-glutamine as described previously. All 63 cell lines were grown anonymously and processed randomly. Genomic DNA was isolated from LCLs following the manufacturer's instruction (Gentra Systems). The ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) were used to check DNA quality and quantity, respectively.

### Bisulfite treatment of DNA samples

Prior to hybridization, bisulfite conversion of DNA samples was performed using the EZ DNA methylation kit (Zymo Research). Five-hundred nanograms DNA were used; a thermocycling program with a short denaturation step was included for bisulfite conversion (16 cycles of 95°C for 30 s followed by 50°C for 1 h).

### Methylation assays by Infinium HumanMethylation27 BeadChip hybridization

After bisulfite treatment, each sample was whole-genome amplified (WGA) and enzymatically fragmented, and then applied to the BeadChips using Illumina-supplied reagents and conditions at Genomic Facility at the the Wistar Institute. HumanMethylation27 DNA Analysis BeadChip (Illumina), which carries 27 578 highly informative CpG sites derived from the well-annotated NCBI CCDS database (Genome Build 36) and spans more than 14 495 genes, was used for this experiment. Allele-specific primer annealing was followed by single-base extension using DNP- and Biotin-labeled ddNTPs. After extension, the array was fluorescently stained, scanned and the intensities of the unmethylated and methylated bead types were measured by a BeadArray Reader (34). Each methylation data point was represented by fluorescent signals from the M (methylated) and U (unmethylated) alleles and was recorded via a Methylation Module in BeadStudio software. DNA methylation values were described as beta ( $\beta$ ) values which computed from the two alleles:  $\beta = M/(U+M)$ . The  $\beta$ -value therefore reflects the fractional methylation level of each CpG site. DNA methylation  $\beta$ -values are continuous variables between 0 and 1, representing the ratio of the intensity of the methylated bead type to the combined locus intensity. We quantified methylation level using  $\beta$ -value, and performed the statistic analysis based on the value of  $\text{Log}_2[\beta/(1-\beta)]$  for more linearized data. We arbitrarily defined CpGs with  $\text{Log}_2[\beta/(1-\beta)] > 0$  [also equals to  $\beta > 0.5$  (50% of reads)] as hyper-methylated, CpGs with  $\text{Log}_2[\beta/(1-\beta)] < -2$  [also equals to  $\beta < 0.2$  (20% of reads)] as hypo-methylated, and CpGs with  $-2 < \text{Log}_2[\beta/(1-\beta)] < 0$  as medium methylated.

### Statistical analysis

Data processing and statistical analyses were performed within R statistical environment ([www.r-project.org](http://www.r-project.org)); all CpG sites were included in the analysis. Array data from the methylated and unmethylated alleles were first processed separately by LOESS normalization across 63 LCL samples (the three sets of triplicated artificially methylated and unmethylated DNA controls were excluded) and then put together to calculate  $\beta$ -values. Differential methylation between control–CdLS or male–female groups was evaluated by a two-way analysis of variance (ANOVA) model with disease status and sample gender as the two tested factors. FDR was estimated by a procedure that randomly shuffled the sample labeling and repeated the ANOVA test 100 times.

### Bisulfite sequencing validation

To validate the data obtained using the HumanMethylation27 BeadChip, samples from four to six healthy controls and four to six severely affected probands with *NIPBL* mutations were evaluated by bisulfite sequencing (BS). The genomic addresses and sequence information of each CpG dinucleotide on the BeadChip were downloaded from the company's database (Illumina);  $\pm 200$  bp surrounding the target CpG of *CAPN2* and *LMO2* were retrieved from the UCSC genome database (<http://genome.ucsc.edu/>) and used as template sequence to design polymerase chain reaction (PCR) primers. Genomic DNA (1  $\mu$ g) was bisulfite-converted and recovered as described above. Primers were designed by Methyl Primer Express v1.0 software (Applied Biosystems) using default settings. Primer sequences are available on request. Hot-start touchdown PCR was done in 25  $\mu$ l reaction containing 0.25 mM dNTPs, 1 $\times$  buffer, 0.4  $\mu$ M forward and reverse primers and 1 U of Zymo *Taq*<sup>TM</sup> DNA Polymerase (Zymo Research). The PCR conditions were as the following: 94°C for 15 min, then 14 cycles with a gradual decrease of annealing temperature from 62°C to 55°C with 0.5°C reduction per cycle followed by 72°C for 1 min per cycle. After that, amplification was continued with 36 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min, then ended with 72°C for 15–20 min. PCR products were verified by gel electrophoresis, 2  $\mu$ l PCR product was subsequently cloned into pGEM-T vector and transformed into JM109 cells according to the manufacturer's instruction (Promega). Ten to twelve individual clones of each PCR fragment were selected for sequencing using an ABI Prism 377 automatic sequencer (Applied Biosystems) with T7 primer. Sequencher and MacVector were used to align obtained sequences with reference sequences from the UCSC genome browser (<http://genome.ucsc.edu/>). C $\rightarrow$ T changes at CpG sites were documented.

### Affymetrix expression array hybridization and ChIP microarray analysis

All procedures were performed as describe in our previous publication (26).

### Online data mining

EpiGRAPH (<http://epigraph.mpi-inf.mpg.de/WebGRAPH/>) is an online software to analyze genomic and epigenomic features enriched in a group of given DNA fragments (35–37). EpiGRAPH was used to analyze DNA sequences harboring the differentially methylated CpG sites in CdLS in terms of the specific DNA sequence patterns, the overlap with specific genomic regions (e.g. CpG islands, repetitive regions and SNPs) and histone modification makers. Galaxy (<http://galaxy.psu.edu/>) was used to format downloaded genomic sequences from the UCSC genome browser (<http://genome.ucsc.edu/>). ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) was used to identify consensus sequence around CpG sites whose methylation was significantly changed in CdLS.

### Accession numbers

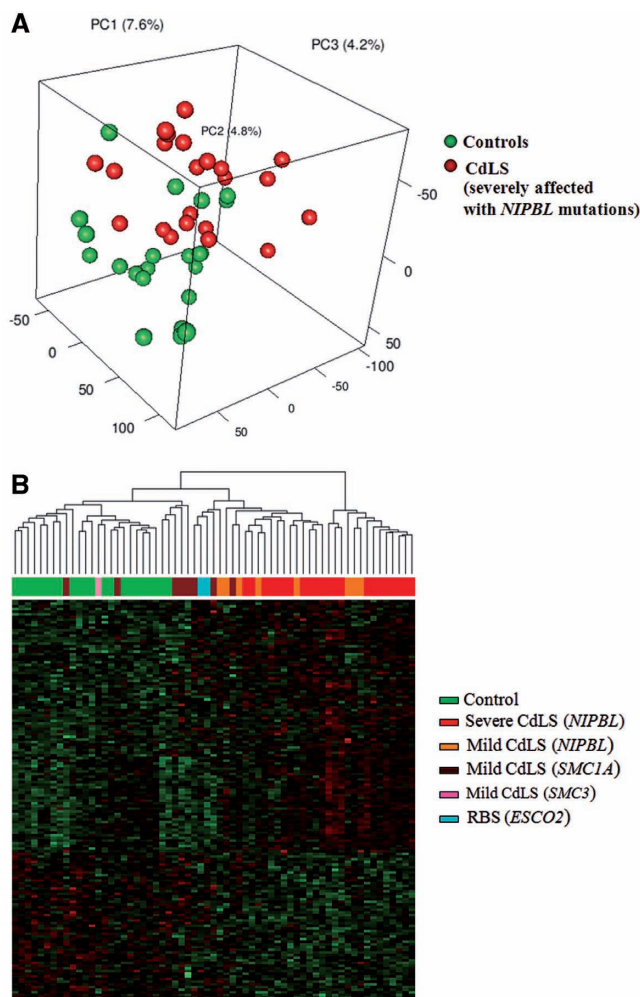
Genomic sequences reported in this manuscript have been submitted to NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>): methylation data are under accession number GSE 18458, gene expression data are under accession number GSE 12408 and ChIP–chip data are under accession number GSE 12603.

## RESULTS

### The overall whole-genome DNA methylation patterns in human LCLs in healthy controls

Whole-genome DNA methylation studies were conducted on 72 sodium bisulfite-converted DNA samples obtained from 63 sample LCLs, two sets of triplicated artificially de-methylated DNA controls and one set of triplicated artificially fully methylated DNA controls. Infinium HumanMethylation27 BeadChip that carries 27 578 CpG dinucleotides located within the promoter region of 14 495 unique genes in the human genome was used for hybridization. Hybridization signal and  $\beta$ -value were examined as described in 'Material and Methods' section,  $\beta < 0.2$  (20% of reads) was considered as low level of methylation, while  $\beta > 0.5$  (50% of reads) was considered as high level of methylation. Linearized  $\text{Log}_2 [\beta / (1 - \beta)]$  was used for statistic analysis. Samples were selected from Caucasian individuals and gender and age were closely matched. The 63 experimental samples included 22 healthy controls, 22 severely affected CdLS probands with *NIPBL* protein truncating mutations, eight mildly affected CdLS probands with *NIPBL* missense mutations, eight mildly affected CdLS probands with *SMC1A* mutations, one mildly affected proband with the *SMC3* mutation and two RBS probands with homozygous mutations in *ESCO2* (Supplementary Table S1). After normalization, the six replicated unmethylated and three replicated methylated artificial DNA control samples showed corresponding low and high  $\beta$ -values, indicating an efficient bisulfite conversion of DNA samples (Supplementary Figure S1). The methylation density peaks of all of the 63 experimental samples have  $\text{Log}_2 [\beta / (1 - \beta)] < 0$  and overlap the peaks of the unmethylated

artificial controls, indicating that the DNA from human LCLs is hypo-methylated globally (Supplementary Figure S1). There are 26486 CpG dinucleotides on the human autosomes (chromosomes 1–22), the average DNA methylation levels of these dinucleotides are highly variable among the 22 control LCLs, but the majority of them are hypo-methylated. Bimodal distribution of DNA methylation was seen on all CpGs among the 22 controls, suggesting two types of genomic contents may be produced by DNA methylation. This bimodal distribution further suggests that it may represent the overall hypo-methylation of CpG sites inside CpG islands and the overall hyper-methylation of CpG sites located within the CpG poor genomic regions in human LCLs (Supplementary Figure S1).



**Figure 1.** CdLS specific DNA methylation profile revealed by unsupervised and supervised classification analysis. (A) Unsupervised principal component analysis (PCA) of DNA methylation on 26486 autosomal CpG sites. Twenty-two severely affected CdLS probands with *NIPBL* mutations and 22 healthy controls are clearly separated into two groups except five controls.  $\text{Log}_2[\beta/(1-\beta)]$  was calculated for each probe in each sample. (B) Hierarchical clustering of 63 experimental samples based on 152 differentially methylated autosomal CpG sites ( $P < 0.001$ ). Individuals with different clinical presentations are color-coded, the corresponding gene mutations are labeled inside the brackets.

### Identification of CdLS specific DNA methylation patterns in LCLs

*Differential DNA methylation identified in CdLS probands.* Methylation levels of 26486 probes on autosomes (chromosomes 1–22) were quantified by  $\beta$ , afterwards  $\text{Log}_2[\beta/(1-\beta)]$  of each probe in each sample was calculated and used for unsupervised principal component analysis (PCA). Figure 1A depicts that the 22 severely affected CdLS probands with *NIPBL* protein truncating mutations can be separated from the 22 healthy controls solely based on DNA methylation levels on autosomes except five healthy control samples. Control ‘AGS-222-S’, ‘27574’, ‘CDL-145-03S’, ‘27572-B’ and ‘95-3986-S’ are discordant to the rest of other samples in control group. This unsupervised PCA result indicates that the genome-wide DNA methylation profile is significantly different in CdLS as compared to healthy individuals. Probes on X and Y were excluded due to the gender effect on DNA methylation.

In order to define the differentially methylated DNA loci in CdLS, we performed a two-way ANOVA analysis of the total 27578 probes on the control group (22 individuals) and the severely affected *NIPBL* mutant CdLS group (22 individuals). There are 924 CpG sites (corresponding to 902 cognate genes) differentially methylated in CdLS with  $P < 0.01$  (FDR = 0.222), out of these 924 CpG sites, methylation levels were decreased on 361 sites (356 genes) and increased on 563 (546 genes) sites in CdLS (Supplementary Table S2). We selected 152 differentially methylated CpG sites on the autosomes with  $P < 0.001$  (Supplementary Table S2) and further performed clustering analysis on all of the 63 samples in our cohort which are varied both clinically and genotypically (Figure 1B). Samples from each subgroup could be clustered together, although with a few outliers. Control samples and severely affected *NIPBL* mutant CdLS probands are evidently separated from each other, whereas mildly affected CdLS probands stay in between. Within the mildly affected CdLS probands, individuals with *NIPBL* mutations stay closer to the *NIPBL* mutant subgroup that has severe manifestations, whereas CdLS probands with other gene mutations (*SMC1A* and *SMC3*) tend to cluster together and stay closer to healthy controls. Interestingly, the two RBS probands are clustered side by side with *SMC1A* mutant individuals. RBS is an autosomal recessive genetic disorder due to homozygous or compound heterozygous mutations in the *ESCO2* gene, *ESCO2* has acetyltransferase activity and is involved in the establishment of sister chromatid cohesion.

*The X chromosome has a unique DNA methylation pattern in CdLS.* We further separately analyzed the hypo-methylated CpG sites ( $\text{Log}_2[\beta/(1-\beta)] < -2$ ) and hyper-methylated CpG sites ( $\text{Log}_2[\beta/(1-\beta)] > 0$ ) in controls. In addition, autosomal and X-linked sites were also analyzed separately (Supplementary Table S3). In general, the average DNA methylation levels of all of the hyper-methylated CpG sites and the X-linked hypo-methylated CpG sites have been increased in

CdLS, whereas the autosomal hypo-methylated CpG sites remain unchanged (Figure 2A). Statistically, the average methylation level (combining both females and males) of X-linked hypo-methylated CpG sites has increased significantly in CdLS as compared to the increase of autosomal hypo-methylated CpG sites ( $P = 1.51E-08$ ) (Figure 2A). Since gender could play a major role affecting DNA methylation levels, we then repeated the above analysis by analyzing female and male samples separately. We saw more significant methylation change of hypo-methylated CpG sites on the X chromosome than on the autosomes in CdLS females ( $P = 3.5E-17$ ). Significant increase in DNA methylation of hypo-methylated CpG sites than that of hyper-methylated CpG sites was also found on X chromosome in female CdLS probands ( $P = 1.2E-12$ ) (Figure 2B). Surprisingly, significantly reduced methylation levels were found for the above hypo-methylated CpG sites on the X chromosomes in male CdLS probands (Figure 2B).

#### **The correlation between the differential DNA methylation and the dysregulated gene expression in CdLS**

*General correlation between DNA methylation and gene expression in control human LCLs.* The expression of genes in LCLs and the methylation levels of their associated CpG sites were first examined in 22 control samples. We took advantage of the whole-genome expression analysis from our previous study (26), in which we have identified 10 378 out of 15 162 unique refSeq genes expressed in LCLs using Affymetrix HG\_U133plus2.0 arrays. Each of these genes was mapped to one and only one transcription starting site (TSS). CpG dinucleotides were mapped to within  $\pm 1$ -kb regions surrounding TSSs to obtain 22 351 TSS-CpG pairs from 12 081 unique genes and 21 110 unique CpG sites. Consistent with the well-documented literature, an inverse correlation between DNA methylation and gene expression was also found in LCLs in our study. DNA methylation suppresses gene expression and the reduced methylation level correlates to higher probability of gene expression (Figure 3A, Supplementary Figure S2A). Figure 3A shows the percentage of expressed genes and the corresponding CpG methylation levels. The overall percentage of genes expressed in LCLs is 69.9% (8449/12 081), which indicates that 69.9% of human genes are likely to be expressed in LCLs. While the level of DNA methylation is increased, fewer genes are expressed. Using  $(\text{Log}_2[\beta/(1-\beta)] = 0)$  as cutoff, Fisher's exact test shows highly significant interdependence between gene expression and DNA methylation in the promoter region ( $P = 4.5E-315$ ) (Figure 3A). The elevated expression of a subgroup of genes that is associated with hyper-methylation ( $\text{Log}_2[\beta/(1-\beta)] > 0$ ) seems to be unexpected, suggesting mechanisms other than DNA methylation are involved in their transcriptional regulation; or in another word, DNA methylation alone is not enough to downregulate gene expression in human LCLs (Figure 3A, Supplementary Figure S2A). We subdivided the 21 110 unique sites into two groups based on the CpG island mapping information from the UCSC genome browser (<http://genome.ucsc.edu>) as CpG sites

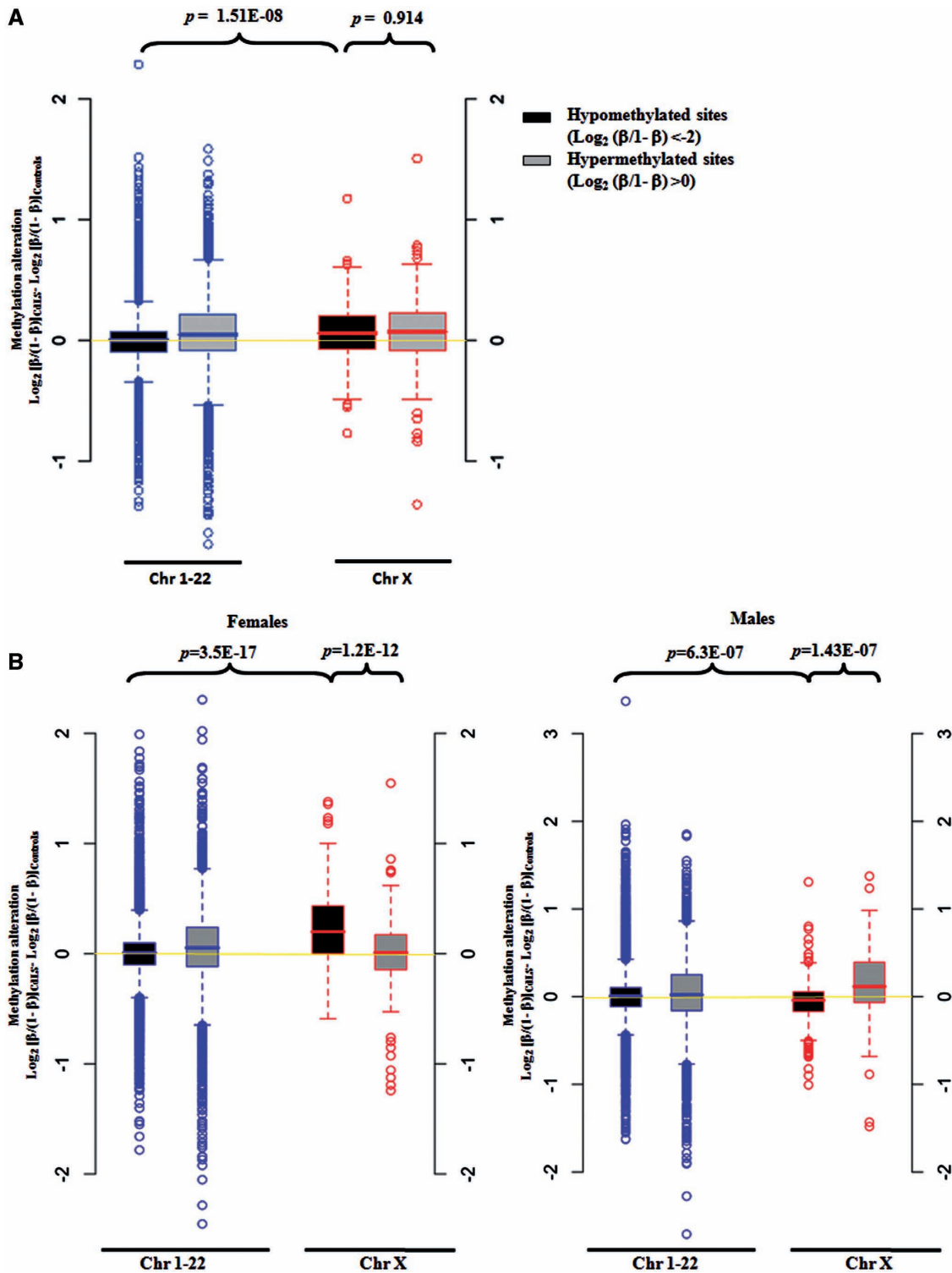
located within the CpG island (CGI sites) and outside the CpG island (non-CGI sites). There was no linear correlation between the level of DNA methylation and the transcriptional activity of the cognate gene observed for both CGI sites and non-CGI sites, which further indicated the direct impact from DNA methylation on gene expression might not be predominant (Supplementary Figure S2B).

We also found that the relative location of the CpG site to the TSS impacts the association between DNA methylation and the expression of the downstream gene (Figure 3B). While 44.5% (1224/2752) of the genes with hyper-methylated CpG sites around their TSSs were expressed in LCLs, the closer the sites were located to the TSSs, the less likely the genes were to be expressed. When there was a hyper-methylated site located within  $\pm 50$  bases around a TSS, the downstream gene had only a 27.6% (93/337) chance of being expressed. However, such an effect of the hypo-methylated sites was less evident. The negative correlation between DNA methylation and gene expression is stronger on autosomes than on the X chromosome (Supplementary Figure S2C). The probability of genes to be expressed on autosomes and the X chromosome is 82.3% and 61.4%, respectively ( $P = 6.0E-06$ ) when there is a hypo-methylated site around the TSS in female samples, and 36.8% and 44.5% ( $P = 0.08$ ) when there is a hyper-methylated site. This observation suggests that the association between DNA methylation and gene expression has different mechanisms on autosomes and the X chromosome.

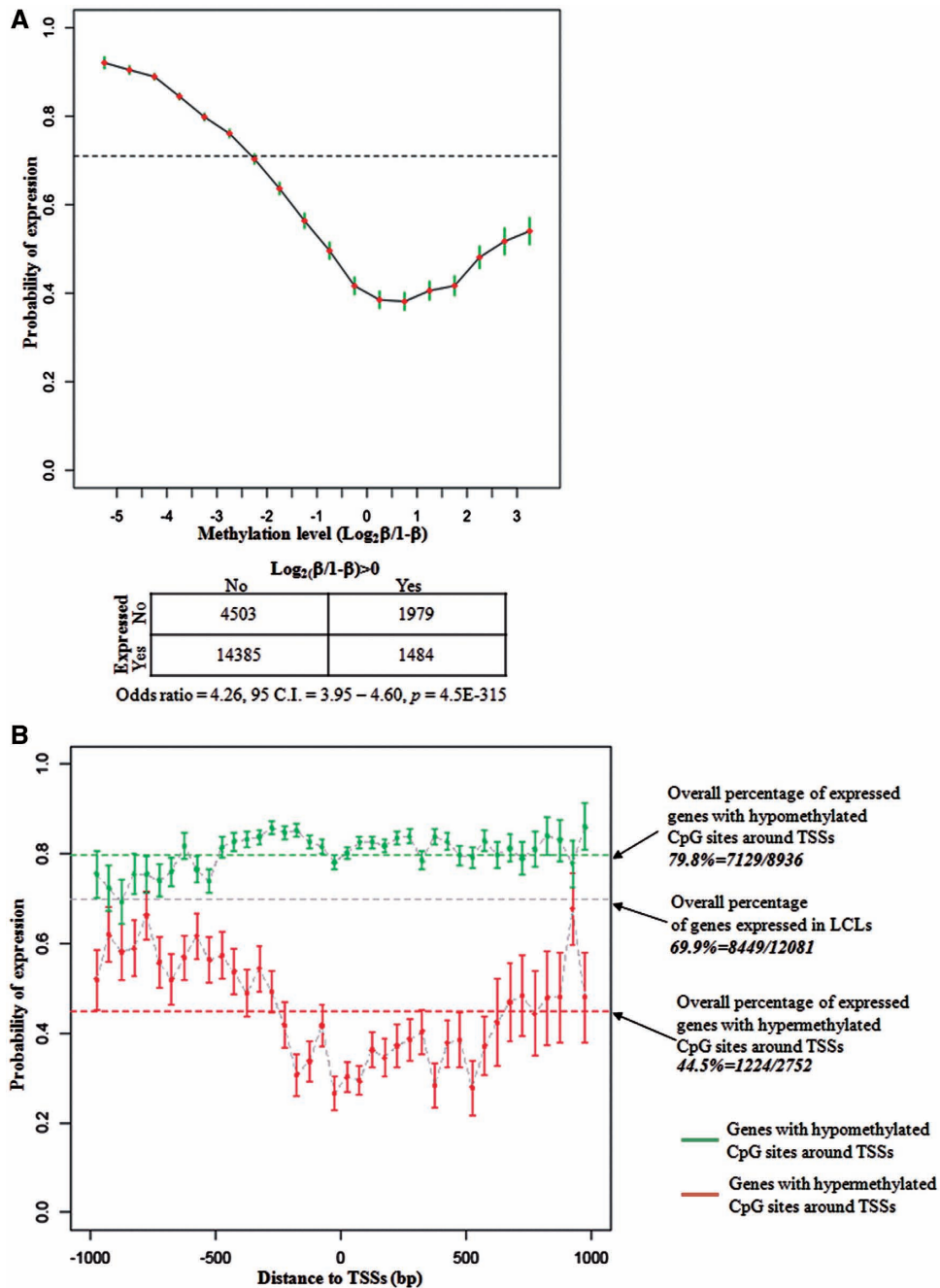
*The correlation between differential DNA methylation and altered gene expression in CdLS.* As described earlier, 12 081 unique refSeq genes are associated with CpG sites with DNA methylation information available from this study. Out of these 12 801 genes, 8449 genes are expressed in LCLs. We compared the expression level and the methylation level for each of these 8449 genes between 22 controls and 22 severely affected *NIPBL* mutant CdLS probands, then we plotted the alteration of expression to the alteration of methylation for each gene. In general, there is no strong evidence supporting that global differential DNA methylation correlates to global transcriptional dysregulation in CdLS (Figure 4). The increased DNA methylation level correlates to the transcription downregulation for only six genes in CdLS ( $P = 0.003$ ), whereas the decreased DNA methylation level correlates to transcriptional upregulation for only two genes but without statistic significance (Figure 4, Supplementary Table S4). There is no link seen between the altered gene expression of the rest of 8441 genes and the altered DNA methylation in CdLS (Figure 4). We therefore conclude that the differential DNA methylation may not directly contribute to the transcriptional dysregulation in CdLS.

#### **The correlation between the differential DNA methylation and cohesin binding in CdLS**

*Cohesin preferentially binds to hypomethylated DNA in control human LCLs.* The average DNA methylation



**Figure 2.** Box plots showing DNA methylation changes for hypo-methylated CpG sites on chromosome X are different in CdLS. (A) DNA methylation levels of X-linked hypo-methylated CpG sites have significantly increased in CdLS as compared to autosomal hypo-methylated CpG sites. Twenty-two controls and 22 severely affected CdLS probands with *NIPBL* mutations were included. (B) DNA methylation levels of X-linked hypo-methylated CpG sites have significantly increased in female CdLS probands but decreased in male CdLS probands as compared with methylation changes of the autosomal CpG sites. Twenty-two controls and 22 severely affected CdLS probands with *NIPBL* mutations were included, and females (11 controls and 11 probands) and males (11 controls and 11 probands) were analyzed separately. The label of Y-axis is defined as 'CdLS-Control' =  $\text{Log}_2 (\beta/1-\beta)_{\text{CdLS}} - \text{Log}_2 (\beta/1-\beta)_{\text{Control}}$ , indicating the alteration of methylation levels between CdLS and controls.

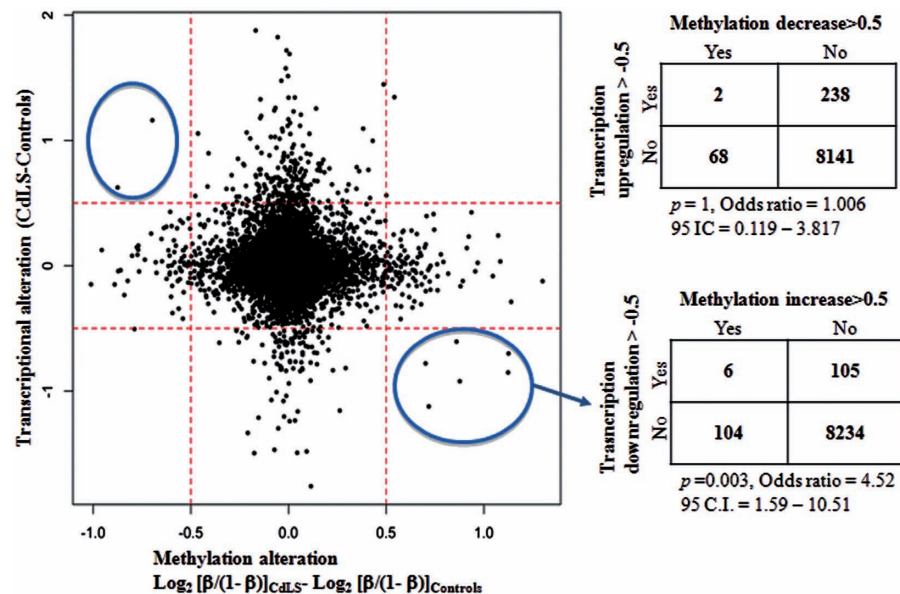


**Figure 3.** Correlation between the DNA methylation and the gene expression in human LCLs. (A) Increased level of DNA methylation correlates to reduced gene expression. A total of 22 351 TSS–CpG pairs from 12 081 unique genes were analyzed. The broken line indicates that 8449 (69.9%) of these genes were expressed in LCLs. (B) Location of the CpG sites impacts the association between DNA methylation and gene expression. Among the genes having hyper-methylated CpG sites around their TSSs, the closer the CpG sites are located to TSS, the less possible the downstream gene is expressed (red lines). This effect is not obvious among genes with hypo-methylated CpG sites around their TSSs (green lines).

level of 21 110 CpG sites associated to the above 12 081 genes in 22 controls were analyzed with  $\text{Log}_2[\beta/(1 - \beta)]$ . Cohesin binding signal was obtained from our previous ChIP–chip assay (26), re-analyzed quantitatively and mapped to each of 21 110 CpG site. The majority of cohesin signals were concentrated at chromatin regions with a low level of DNA methylation which is  $\text{Log}_2[\beta/(1 - \beta)] < -2$  ( $\beta < 0.2$ ), indicating that cohesin preferentially binds to hypo-methylated chromatin;

meanwhile, an increased DNA methylation level prohibits cohesin from binding to the chromatin (Supplementary Figure S3A and B). From our previous study (26), we have shown evidence that cohesin preferentially binds to gene promoters especially at TSSs. We therefore split the 12 081 genes into three groups according to the methylation levels of the CpG sites at their TSSs, and looked at cohesin binding intensities around promoters. As described above, CpGs with  $\text{Log}_2[\beta/(1 - \beta)] > 0$  were

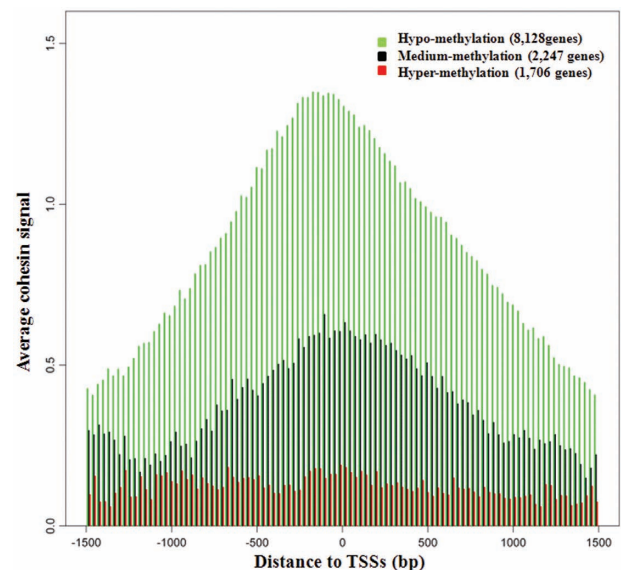




**Figure 4.** The change of DNA methylation does not generally affect gene expression in CdLS. The 8449 genes that were expressed in LCLs were examined in 22 controls and 22 severely affected probands. The increased methylation shows a significant suppression effect on transcription for only six genes in CdLS with  $P = 0.003$  in Fisher's test. The decreased DNA methylation reveals no significant correlation with transcriptional upregulation in CdLS. Y-axis represents the gene expression changes in CdLS calculated with signal intensity on expression arrays; x-axis represents the methylation alteration in CdLS using  $\text{Log}_2(\beta/1 - \beta)$ . The 0.5 cutoff value was set up arbitrarily.

defined as hyper-methylated, while CpGs with  $\text{Log}_2[\beta/(1 - \beta)] < -2$  were defined as hypo-methylated, and CpGs with  $-2 < \text{Log}_2[\beta/(1 - \beta)] < 0$  were defined as medium methylated. A similar cohesin-binding pattern was identified for the 8128 genes with hypo-methylated CpGs and the 2247 medium methylated CpGs. Cohesin preferentially binds to the vicinity of their TSSs especially around the core promoter region (from  $-200$  bp to  $+1$  bp) with each peak clearly seen (Figure 5). On the contrary, very little cohesin binds to the 1706 genes with hyper-methylated CpGs, and there is no quantitative difference of cohesin binding to different regions surrounding the TSSs (Figure 5). Although cohesin binding correlates to gene expression (26), we were not able to identify a direct correlation between gene expression and DNA methylation based on our above analyses. In combination with our previous studies, these results suggest the association between cohesin and promoters are tightly correlated to gene expression, but DNA methylation may only be one of several upstream events that affect cohesin binding. The transcriptional regulation does not appear to solely depend on DNA methylation but rather multiple mechanisms or pathways are likely functioning together to control the expression of genes in human LCLs.

*Differential DNA methylation alters cohesin binding in CdLS.* In addition to respectively examining genes with hypo-methylated promoters and hyper-methylated promoters in CdLS, we also wanted to examine whether the X chromosome or autosome location of the CpG dinucleotide will differently affect the correlation between DNA methylation and cohesin binding in CdLS, hence only the female samples from the control



**Figure 5.** Cohesin preferentially binds to CpG sites at promoter regions with low levels of DNA methylation. Very little cohesin binds to hyper-methylated promoters regardless of the distance to TSSs.  $\text{Log}_2(\beta/1 - \beta) > 0$  was defined as hyper-methylation,  $\text{Log}_2(\beta/1 - \beta) < -2$  was defined as hypo-methylation,  $-2 < \text{Log}_2(\beta/1 - \beta) < 0$  was defined as medium methylation. Cohesin binding signal was quantified from previous ChIP-chip study and log-transformed before the analysis.

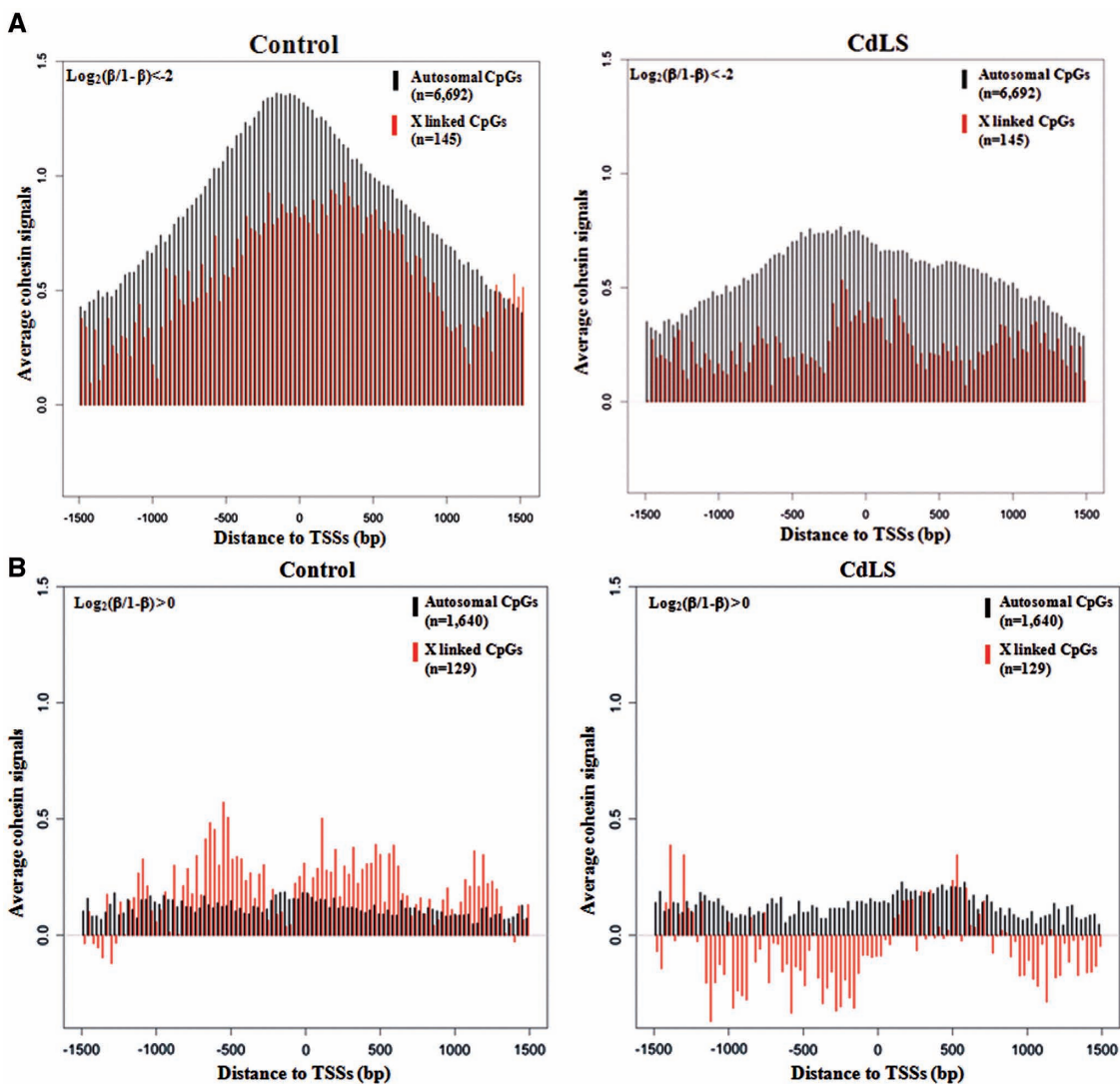
group and severely affected *NIPBL* mutant CdLS individuals were analyzed. For the genes with hypo-methylated promoters, binding of cohesin is enriched around TSSs regardless of whether they are located on autosomes (6692 genes) or X chromosome (145 genes); however,

less cohesin binds to the X chromosome than to autosomes and the peak shifts upstream of the TSS. In CdLS, the amount of both autosomal and X chromosome bound cohesin has decreased, although a smaller peak remains in a narrowed region surrounding the TSS (Figure 6A). For the genes with hyper-methylated promoters, the amount of cohesin bound is quite low for both autosomal and X-linked genes (1640 and 129, respectively) (Figure 6B), and cohesin binding is equally distributed along  $\pm 1.5$ -kb region of TSSs with no obvious peak seen at TSSs. More cohesin is associated with the X chromosome than with the autosomes in controls but this tendency was reversed in CdLS which shows less cohesin associated with the X chromosome than with autosomes (Figure 6B). A slightly reduced amount of cohesin binds to the X chromosome in CdLS,

but there is little change in cohesin binding to autosomal genes in CdLS (Figure 6B). In summary, cohesin binding has changed remarkably at hypo-methylated promoters in CdLS, with unknown mechanisms in addition to DNA methylation likely involved in regulating cohesin binding to X-linked and hyper-methylated promoters.

### BS validation for microarray results and evidence of varied DNA methylation in control LCLs

Bisulfite conversion and sequencing described by Susan Clark (38) has always been considered the gold standard by which to measure CpG methylation. We selected two genes, *LMO2* and *CAPN2*, which demonstrated differential methylation between CdLS and control cells, for BS validation. In our previous studies (26), these two

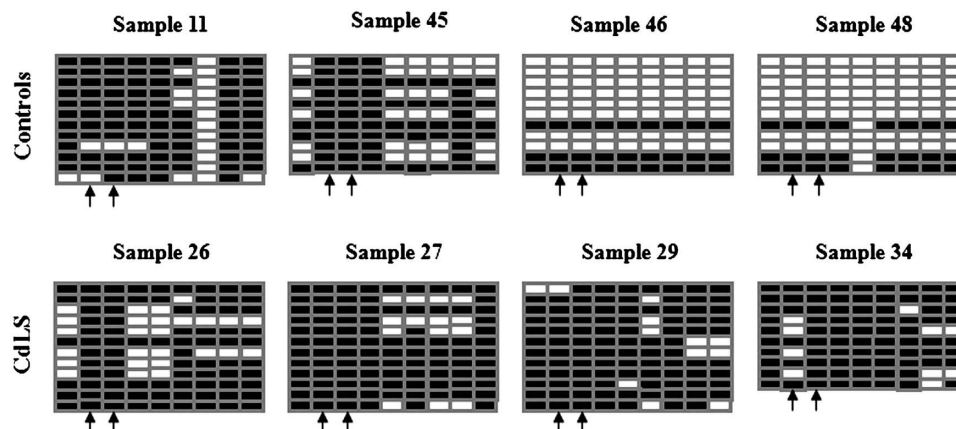


**Figure 6.** Reduced cohesin binding at hypo-methylated promoters in CdLS with minimal changes for hyper-methylated promoters. (A) Cohesin binding to genes with hypo-methylated promoters is reduced in CdLS, and less cohesin binds to promoters on the X chromosome than on autosomes; 6692 of such genes are on autosomes, while 145 are on the X chromosome. (B) Less cohesin binds to genes with hyper-methylated promoters in both controls and probands; however, more cohesin binds to X-linked promoters than to autosomal promoters in controls and this X-linked cohesin binding is reduced in CdLS.  $\text{Log}_2(\beta/1-\beta) < -2$  was defined as hypo-methylation and  $\text{Log}_2(\beta/1-\beta) > 0$  was defined as hyper-methylation. Cohesin signal was quantified and log-transformed before the analysis. Only 11 female samples from controls and 11 female samples from severely affected CdLS probands were analyzed.

genes showed reduced transcription and loss of cohesin binding to TSSs in CdLS cells. There is a big CpG island embedded at the *CAPN2* promoter region; two probes, cg01566170 and cg14972271 on the HumanMethylation27 array, were located 455 bp and 465 bp upstream of the TSS of *CAPN2*, respectively. The 252-bp fragment (chr1:221,966,278-221,966,529) amplified by primers CAPN2 329/330 covers these two probe CpG sites in addition to seven other CpG dinucleotides included in this fragment. DNA isolated from four controls and four severely affected CdLS probands with *NIPBL* mutations tested on methylation arrays above were tested again by BS. Ten to twelve clones were sequenced for each sample. Figure 7 shows the increased number of 5-methylcytosine at CpG sites in CdLS probands, indicating elevated DNA methylation levels in this chromatin region at the *CAPN2* promoter which is consistent with the methylation microarray findings. In addition, the BS result from control samples also illustrated varied levels of DNA methylation in the healthy population. Sample 11 obviously has much higher methylation level than sample 46 and 48, although, on average, 48.8% of CpGs in control and 80.3% in CdLS are methylated. A second promoter region, the promoter of the *LMO2* gene, was also tested by BS in six control and six CdLS individuals. A 358-bp fragment that is located at the *LMO2* promoter region was amplified by primers LMO2 337/338. This examined fragment is located on chr11:33,870,140-33,870,497 surrounding probe cg11822932 on the HumanMethylation27 array, which is 120 bp upstream of the TSS of *LMO2*. The *LMO2* promoter is very CpG poor and harbors no region that meets the criteria of the 'CpG island'. There are only three CpG dinucleotides in this tested region including the probe CpG. We could demonstrate similar methylation changes at the *LMO2* locus as demonstrated by the array study (Supplementary Figure S4). In conclusion, a consistent methylation pattern was obtained from both BS and HumanMethylation27 analysis.

### Specific histone modifications and repetitive sequences are involved in the differential methylation in CdLS by EpiGRAPH analysis

EpiGRAPH (<http://epigraph.mpi-inf.mpg.de/WebGRAPH/>) is an online software to analyze genomic and epigenomic features that are enriched in given DNA fragments (35–37). As described above, we have identified 924 CpG sites correlated to 902 genes that are differentially methylated in CdLS ( $P < 0.01$ ), out of which the methylation level of 361 CpGs (356 genes) are decreased and the methylation level of 563 CpGs (546 genes) are increased. The EpiGRAPH web service was used to analyze DNA features enriched in regions with differential methylation in CdLS. One kilobase DNA sequence ( $\pm 500$  bp) surrounding each of the 924 CpG dinucleotides was identified and downloaded from the UCSC genome browser (<http://genome.ucsc.edu/>), processed in Galaxy (<http://galaxy.psu.edu/>) and uploaded to EpiGRAPH to identify overrepresented genomic or epigenomic features by comparing the 361 sequences harboring hyper-methylated CpGs and the 563 sequences harboring the hypo-methylated CpGs in CdLS. We chose a size of 1000 bp for the analyses because in normal tissues the extended distance of DNA methylation is generally shorter than 1000 bp (13). Significantly distinct histone modifications were found in chromatin regions with both differential DNA hypo-methylation and hyper-methylation, indicating a tight correlation between the alteration of DNA methylation and the diverse chromatin structure in CdLS (Table 1, Supplementary Table S5). The finding that repetitive sequences are enriched in chromatin regions harboring differentially methylated CpG sites is very interesting because cohesin binding is dysregulated in CdLS and the association between cohesin and repetitive sequences has been suggested to be involved in multiple biological roles (15). ClustalW2 (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) was further used to look for consensus sequences covering the differentially



**Figure 7.** Bisulfite sequencing of a 252-bp fragment that is located at the *CAPN2* promoter region in four control and four CdLS samples. The examined fragment is located on chr1:221,966,278-221,966,529 surrounding probes cg01566170 and cg14972271 on the HumanMethylation27 array, which are 455 bp and 465 bp upstream of the TSS of *CAPN2*, respectively. There are nine CpG dinucleotides in this region. A solid block represents the methylated 'C' allele that remains as 'C' after bisulfite conversion and an empty block represents the unmethylated 'C' allele that was converted to 'T' after bisulfite treatment. Ten to twelve clones were selected for sequencing for each sample. Each row represents one sequenced clone and each column represents one examined 'CpG' site. The arrows point to probes cg01566170 and cg14972271.

**Table 1.** Specific histone modification markers and repetitive genomic sequences are enriched in chromatin regions harboring differentially methylated CpGs in CdLS

	Hyper_M	Hypo_M
Histone modification markers	H2A_H4R3me2	H2A_Z
	H3K27me1, H3K27me2	H3K27me3
	H3K36me1	H3K4me2, H3K4me3
		H3K9me1, H3K9me3
		H3R2me1, H3R2me2
Repetitive sequences	LINE	H4K20me1, H4K20me3
	SINE	H3K79me3
	L2	repClass_Low_complexity
		repFamily_Low_complexity
		L2

EpiGRAPH analysis was performed on 361 1-kb DNA fragments surrounding differentially hyper-methylated CpGs and 563 1-kb fragments surrounding differentially hypo-methylated CpGs in CdLS.

Hyper\_M: chromatin region with differential hyper-methylation in CdLS; Hypo\_M: chromatin region with differential hypo-methylation in CdLS.

methylated CpG sites; however, no consensus sequences could be identified (data not shown).

## DISCUSSION

CdLS is a dominant genetic disorder with multiple-organ system abnormalities, including characteristic facial features, limb defects, mental retardation, developmental delay and gastrointestinal problems. Mutations in the cohesin regulatory protein NIPBL and cohesin subunits SMC1A and SMC3 account for ~65% of confidently diagnosed probands, while the remaining ~35% of probands have no identifiable gene mutations (19). Prior studies (19) have shown that CdLS probands with *SMC3* and *SMC1A* mutations present with a milder clinical picture than the classic form of CdLS typically associated with protein truncating mutations in *NIPBL*. Since NIPBL is a regulator of the cohesin complex and both SMC3 and SMC1A are actual structural components of the cohesin complex, the exact mechanisms by which mutations in these proteins manifest their effects on development and gene regulation are likely to be quite different. The cohesin complex consists of four major subunits, SMC1A, SMC3, RAD21 and STAG1/STAG2, forming a ring structure holding sister chromatid together during mitosis and meiosis (39). Additionally, cohesin has been suggested to play pivotal roles in fundamental biological events in humans such as gene expression (22,26), double-strand DNA repair (40,41), genome instability (42), carcinogenesis (43) and chromatin loop formation (44). The co-localization of cohesin and CTCF in human cells further suggests functional cooperation and overlap may exist between these two proteins (25). Of note, mutations in cohesin accessory factors such as NIPBL, ESCO2 and ATRX have also been identified in human developmental disorders with similar but quite distinct clinical presentations which are currently collectively named 'cohesinopathies' (15). The cohesinopathies provide

valuable experimental models with naturally occurring mutations to study the biological functions of the cohesin pathway in general and the specific proteins in particular in human cells. Recently, the term 'genomic disorder' was proposed to describe groups of developmental disorders or human malignancies in which epigenetic mechanisms including DNA methylation play important roles in the pathogenesis (33). It is not known whether similar mechanisms could also be involved in CdLS.

We chose to use a seemingly phenotypically unrelated tissue type, LCL, in this proof-of-principle study on a human developmental disorder, CdLS, based on two considerations: availability of the sample and goals of the project. As opposed to primary fibroblasts which represent a limited resource, has uncontrollable environmental exposure (such as diet or medications), and has significant variation of growth and survival rate across different samples in culture, patient-derived LCLs for genome-wide DNA methylation studies on human subjects has obvious advantages including (i) easier growth under controlled conditions to minimize the environmental influence on DNA methylation, (ii) ease of identification of sufficient numbers of matched samples (gender, age, clinical manifestations, etc.) for valid statistical analysis that would be much more difficult using fibroblasts (especially for groups engaged in the study of rare human disorders) and (iii) provides continually renewable and stable biomaterials that ensure for sequential integrated genomic analyses. Of note, GM12878, a well-circulated model LCL, has been universally used in the HapMap, ENCODE, and other major projects. Publicized DNA methylation data generated for GM12878 on the same Illumina platform demonstrated good correlation with our LCL samples with  $r = 0.947$  when the low-quality measurements were removed (Zhang *et al.*, manuscript in preparation). In addition, LCLs have been widely used as surrogates or as cellular models to study epigenetic changes in neuropsychiatric illnesses such as autism, bipolar disorder, schizophrenia and in central nervous system disorders such as Parkinson disease (45). Genetic disorders, such as ataxia telangiectasia (46) and Nijmegen breakage syndrome, have also been investigated on patient derived LCLs yielding valuable insights into the pathobiology of these disorders (47).

Methylation and demethylation of regulatory sequences in the genome are known to have profound effects on cellular behavior and fate. In this study, we have performed genome-wide DNA methylation analysis in CdLS using Illumina Methylation27 bead chip carrying 27 578 CpG dinucleotides that represents 14 495 cognate genes in the human genome. The power of this array is that a very large number of sites can be determined simultaneously, allowing highly reproducible global patterns to be discovered. Both unsupervised PCA based on results of all data from available CpG sites and the supervised clustering signatures based on differentially methylated autosomal CpG sites shown in Figure 1 indicate that CdLS has a distinct methylation profile as compared to controls. Of note, *NIPBL* transcription has only dropped 30% in those CdLS cells presenting the unique disease-specific DNA methylation pattern in the current

study (26). The hierarchical clustering also demonstrated that the methylation profile correlates to disease severity and to the specific mutated genes: the mildly affected probands stay midway between severe probands and controls; mildly affected probands with *NIPBL* mutations were clustered as a transient group between the mildly affected individuals with mutations in other genes (*SMC1A* or *SMC3*); and the severe individuals with *NIPBL* mutations suggesting a genotype–phenotype correlation (Figure 1B). However, CdLS DNA methylation profile alone is not sensitive enough to serve as a diagnostic tool (Supplementary Figure S5). Leave-One-Out cross validation using  $P < 0.001$  on 22 healthy controls and 22 severely affected CdLS probands with *NIPBL* mutations was conducted. Controls and probands could be roughly separated; however, 11 samples were mis-classified after 44 rounds with a classification accuracy of only 75%.

We next attempted to correlate the data from the three sets of genome-wide analyses on DNA methylation, gene expression and cohesin binding in CdLS. Venn diagram analysis did not reveal a single gene that had significant alterations in all of the three biological events (Supplementary Figure S6); therefore, no direct correlation between the three dysregulated biological events could be revealed in CdLS, indicating that additional unknown mechanisms might be involved in the pathogenesis of CdLS. That there is no remarkable correlation between gene expression and DNA methylation identified in CdLS probably can be explained by (i) A relatively big variation of DNA methylation level exists in healthy controls as seen on both array and BS (Supplementary Figure S1B and Figure 7) which adds to the difficulty of identifying an unambiguously changed DNA methylation pattern in CdLS. (ii) Global DNA methylation is a more stable type of epigenetic modification modulating the transcriptional plasticity in the human genome than we have realized (48). One study reporting the high-resolution methylation states on 1.9 million CpGs on human chromosomes 6, 20 and 22 from 12 different tissues failed to correlate DNA methylation with mRNA expression levels for 63% of the genes. Therefore, the author suggested that differential promoter methylation might have only a permissive role, such as establishing an open chromatin conformation, for the transcription regulation. In combination with other factors or mechanisms that drive transcription but not alone, DNA methylation could regulate the transcription of the cognate genes (49). To support this argument, in our study we have found keratin genes that are not expressed in the studied LCLs, most promoters of the two keratin gene clusters located on chromosome 12 and 17 are constantly under hyper-methylation in control cells. In CdLS, uniform methylation was found again at the two clusters but with even higher methylation levels (Supplementary Figure S7A and B). Keratin genes remain silent in both control and CdLS cells with differential promoter hyper-methylation levels, suggesting gene expression is not solely controlled by DNA methylation. (iii) The average global DNA methylation level in LCLs is low, and overall methylation changes between control and CdLS are quite subtle, making it more difficult to

identify minor but significant changes. (iv) As seen in *Drosophila*, *NIPBL* or cohesin may regulate transcription mainly by the association with enhancers or other remote regulatory elements first and only subsequently with promoters. The CpG dinucleotides examined on the Methylation27 platform in this study have minimal coverage of the remote cis elements, which could lead to the lack of detection of any DNA methylation alteration on enhancers, silencers, insulators, locus control regions, etc. We therefore propose whole-genome BS to be the next experimental approach.

In humans, most genes are expressed from both alleles in diploid cells; however, more and more genes are being identified that are only expressed from a single allele. Mammalian X inactivation, imprinting (e.g. *IGF2* and *H19*), and allelic exclusion (e.g. olfactory receptor genes, immunoglobulin genes, T cell receptors, interleukins and natural killer cell receptors) are classic examples of monoallelic gene expression. A recent report suggests at least 1000 autosomal human genes are subject to random monoallelic expression (50). Regulation of this group of genes is clearly epigenetic and the role of DNA methylation in imprinting has been well recognized, for example, allele-specific DNA methylation has been observed for the immunoglobulin gene and few other monoallelically expressed genes (50–52). Hence, we combined the data set from this study with our previous genome-wide gene expression study and ChIP–chip assay to test whether differential DNA methylation affects monoallelic expression in CdLS and whether it is also involved in the altered cohesin–chromatin association in CdLS. Information on documented human imprinted genes was obtained from the online database ‘Catalogue of Parent of Origin Effects’ (<http://igc.otago.ac.nz/Summary-table.pdf>), a list of X-linked genes with their X inactivation status and a list of monoallelically expressed autosomal genes were obtained from the literature (50,53). Twenty-nine imprinted genes, 465 X-linked genes and 191 randomly (paternal or maternal) monoallelically expressed autosomal genes have data available from all the three assays (DNA methylation, expression and ChIP–chip). Fisher’s exact tests did not reveal any correlation between differential DNA methylation and the differential expression for monoallelically expressed genes in severely affected CdLS probands with *NIPBL* mutations (Supplementary Table 6). In addition, correlation between the differential DNA methylation and altered cohesin binding (loss or addition) at transcription start sites (TSSs), and correlation between the altered cohesin binding at TSSs and the differential gene expression for the monoallelically expressed genes were not identified either (Supplementary Table S7). In conclusion, the current data are not able to support an impact of CdLS-related disruption of cohesin on monoallelic gene expression. This may be due to the tissue or developmental stage specificity of expression of these genes.

DNA methylation state is influenced by a number of endogenous and exogenous parameters such as gender, age, tissue type or passage of cell cultures (54). To be consistent, all samples for expression and methylation studies were from Caucasians, all three genome-wide

assays used the same tissue type (LCLs), and the same set of samples were used for all the studies whenever possible. Gender was strictly matched for the differential expression and DNA methylation analyses. Only two factors, the age and the culture passage, could not be controlled for stringently due to the limitation of resources. Most of our CdLS probands are children and some of the healthy controls are adults. The LCLs we have used are stock cell lines that have been continuously collected in our laboratory for decades. However, in one report, no detectable global changes in average DNA methylation levels can be recognized between a group of 26-year-olds and the second group of 68-year-olds (49), and, in another report, the methylation change during prolonged passage in culture is suggested to be insignificant (55).

At present, genome-wide information of gene expression, DNA methylation and cohesin binding in CdLS is available from our integrative studies. The tip of the iceberg revealed in these studies will help us to design new assays to further understand how cohesin regulates critical biological pathways in humans. For example, '6C' could be conducted to identify epigenetic chromatin looping structure mediated by cohesin at specific loci that associate with a particular gene transcription pattern. Allele-specific DNA methylation cannot be studied on the current array platform but could be possible if combined with the SNP information. The combined data will help us to illustrate NIPBL and cohesin's role in monoallelic expression in the human genome, knowledge of which is currently lacking. ChIP-chip or ChIP-Seq studies to identify specific histone modifications in CdLS are also a needed next step to understand how chromatin structure is involved in human developmental disorders. Finally, animal models with complete or partial NIPBL or cohesin knockdown may help us to interpret cohesin-dependent cellular functions *in vivo*.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

We are grateful for the participation of the children and families with CdLS and to the CdLS Foundation for their support.

## FUNDING

The CdLS Foundation (a Fellowship Grant to J.L.); National Institutes of Health P01 HD052860 and R21 HD050538 (to IDK); KO8 HD055488 (to M.A.D.); Genome Network Project and Grant-in-Aid for Scientific Research (S) from the MEXT, Japan (to K.S.); Pennsylvania Department of Health (to N.B.S.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding for open access charge: National

Institutes of Health (P01 HD052860 and R21 HD050538 to I.D.K.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Chen, K. and Rajewsky, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
- Geiman, T.M. and Robertson, K.D. (2002) Chromatin remodeling, histone modifications, and DNA methylation—how does it all fit together? *J. Cell. Biochem.*, **87**, 117–125.
- Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.
- Niehrs, C. (2009) Active DNA demethylation and DNA repair. *Diff. Res. Biol. Diver.*, **77**, 1–11.
- Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y. and Ito, T. (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.*, **14**, 247–266.
- Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. and Issa, J.P. (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.*, **3**, 2023–2036.
- Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA*, **99**, 3740–3745.
- Zhang, Y., Rohde, C., Tierling, S., Jurkowski, T.P., Bock, C., Santacruz, D., Ragozin, S., Reinhardt, R., Groth, M., Walter, J. *et al.* (2009) DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet.*, **5**, e1000438.
- Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, **449**, 248–251.
- Ehrlich, M., Turner, J., Gibbs, P., Lipton, L., Giovannetti, M., Cantor, C. and van den Boom, D. (2008) Cytosine methylation profiling of cancer cell lines. *Proc. Natl Acad. Sci. USA*, **105**, 4844–4849.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
- Brena, R.M., Huang, T.H. and Plass, C. (2006) Toward a human epigenome. *Nat. Genet.*, **38**, 1359–1360.
- Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
- Liu, J. and Krantz, I.D. (2008) Cohesin and human disease. *Annu. Rev. Genomics Hum. Genet.*, **9**, 303–320.
- Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A. *et al.* (2004) Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat. Genet.*, **36**, 631–635.
- Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J. and Strachan, T. (2004) NIPBL, encoding a homolog of fungal Sec2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.*, **36**, 636–641.
- Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P. and Larizza, L. (2006) X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat. Genet.*, **38**, 528–530.
- Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A.D. *et al.* (2007) Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.*, **80**, 485–494.
- Nasmyth, K. and Haering, C.H. (2005) The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.*, **74**, 595–648.

21. Hirano, T. (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev.*, **16**, 399–414.
22. Rollins, R.A., Morcillo, P. and Dorsett, D. (1999) Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics*, **152**, 577–593.
23. Rollins, R.A., Korom, M., Aulner, N., Martens, A. and Dorsett, D. (2004) Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Mol. Cell. Biol.*, **24**, 3100–3111.
24. Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D. *et al.* (2008) Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. *Chromosoma*, **117**, 89–102.
25. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T. *et al.* (2008) Cohesin mediates transcriptional insulation by CCCT C-binding factor. *Nature*, **451**, 796–801.
26. Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E. *et al.* (2009) Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.*, **7**, e1000119.
27. Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G. and Rauscher, F.J. III (2005) The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem. Biophys. Res. Commun.*, **331**, 929–937.
28. Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell*, **99**, 451–454.
29. Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M. *et al.* (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*, **439**, 871–874.
30. Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F. *et al.* (2007) Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature*, **445**, 671–675.
31. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N. *et al.* (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.*, **40**, 663–669.
32. Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M. and Matzke, A.J. (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.*, **40**, 670–675.
33. Elizondo, L.I., Jafar-Nejad, P., Clewing, J.M. and Boerkoel, C.F. (2009) Gene clusters, molecular evolution and disease: a speculation. *Curr. Genomics*, **10**, 64–75.
34. Oliphant, A., Barker, D.L., Stuelpnagel, J.R. and Chee, M.S. (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *BioTechniques*, (Suppl), 56–58, 60–61.
35. Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T. and Walter, J. (2006) CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. *PLoS Genet.*, **2**, e26.
36. Bock, C., Walter, J., Paulsen, M. and Lengauer, T. (2007) CpG island mapping by epigenome prediction. *PLoS Comput. Biol.*, **3**, e110.
37. Bock, C., Halachev, K., Buch, J. and Lengauer, T. (2009) EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. *Genome Biol.*, **10**, R14.
38. Clark, S.J., Harrison, J. and Molloy, P.L. (1997) Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene*, **195**, 67–71.
39. Hirano, T. (2006) At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell. Biol.*, **7**, 311–322.
40. Vrouwe, M.G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B.C., Bhuiyan, Z.A., Redeker, E.J., Mannens, M.M., Mullenders, L.H., Pastink, A. *et al.* (2007) Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Hum. Mol. Genet.*, **16**, 1478–1487.
41. Revenkova, E., Focarelli, M.L., Susani, L., Paulis, M., Bassi, M.T., Mannini, L., Frattini, A., Delia, D., Krantz, I., Vezzoni, P. *et al.* (2009) Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. *Hum. Mol. Genet.*, **18**, 418–427.
42. Musio, A., Montagna, C., Mariani, T., Tilenni, M., Focarelli, M.L., Brait, L., Indino, E., Benedetti, P.A., Chessa, L., Albertini, A. *et al.* (2005) SMC1 involvement in fragile site expression. *Hum. Mol. Genet.*, **14**, 525–533.
43. Barber, T.D., McManus, K., Yuen, K.W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S. *et al.* (2008) Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl Acad. Sci. USA*, **105**, 3443–3448.
44. Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G. and Merkenschlager, M. (2009) Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, **460**, 410–413.
45. Sie, L., Loong, S. and Tan, E.K. (2009) Utility of lymphoblastoid cell lines. *J. Neurosci. Res.*, **87**, 1953–1959.
46. Watts, J.A., Morley, M., Burdick, J.T., Fiori, J.L., Ewens, W.J., Spielman, R.S. and Cheung, V.G. (2002) Gene expression phenotype in heterozygous carriers of ataxia telangiectasia. *Am. J. Hum. Genet.*, **71**, 791–800.
47. Cheung, V.G. and Ewens, W.J. (2006) Heterozygous carriers of Nijmegen Breakage Syndrome have a distinct gene expression phenotype. *Genome Res.*, **16**, 973–979.
48. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
49. Eckhardt, F., Lewin, J., Cortese, R., Rakan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.*, **38**, 1378–1385.
50. Gimelbrant, A., Hutchinson, J.N., Thompson, B.R. and Chess, A. (2007) Widespread monoallelic expression on human autosomes. *Science*, **318**, 1136–1140.
51. Rouhi, A., Gagnier, L., Takei, F. and Mager, D.L. (2006) Evidence for epigenetic maintenance of Ly49a monoallelic gene expression. *J. Immunol.*, **176**, 2991–2999.
52. Mostoslavsky, R., Singh, N., Kirillov, A., Pelanda, R., Cedar, H., Chess, A. and Bergman, Y. (1998) Kappa chain monoallelic demethylation and the establishment of allelic exclusion. *Genes Dev.*, **12**, 1801–1811.
53. Carrel, L. and Willard, H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
54. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**(Suppl), 245–254.
55. Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E. *et al.* (2006) Human embryonic stem cells have a unique epigenetic signature. *Genome Res.*, **16**, 1075–1083.