

have a copy number change by real-time qPCR. At regions determined to be disomy by SNPacGH, 70.6% were determined using qPCR to be two copies. Previous studies have documented that the most common genomic alteration in papillary carcinoma is amplification on 1p, and deletion or LOH on 16q (2,5,6,8). Our study revealed that chromosomal regions at 3p21.31, 3p14.2, and 20q13.13 were commonly altered ($\geq 4/5$) among five carcinomas in addition to deletion or CNLOH on 16q ($\geq 3/5$) (Fig. 2, Table S2). The significance of 3p and 20q are currently unclear, but require further investigation.

In summary, we have elucidated significant differences in the molecular-cytogenetic profile between papilloma and papillary carcinoma. Thus, papillary carcinoma harbored significantly more genomic alterations than papilloma, even though papilloma had a number of genomic alterations, and the rate of genomic alteration correlated with pathological malignancy classification. These genome-wide findings could not be obtained by conventional cytogenetic study such a fluorescent in situ hybridization or conventional CGH. Our findings may aid clinical management of breast ICPT, and may provide insight into their carcinogenesis.

Masahiro Oikawa, MD^{*,†}

Takeshi Nagayasu, MD, PhD[†]

Hiroshi Yano, MD, PhD[†]

Tomayoshi Hayashi, MD, PhD[‡]

Kuniko Abe, MD, PhD[‡]

Akira Kinoshita, PhD^{*}

Koh-ichiro Yoshiura, MD, PhD^{*}

^{*}Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

[†]Surgical Oncology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan;

[‡]Department of Pathology, Nagasaki University Hospital, Nagasaki, Japan

REFERENCES

1. Fayanju OM, Ritter J, Gillanders WE, *et al.* Therapeutic management of intracystic papillary carcinoma of the breast: the roles of radiation and endocrine therapy. *Am J Surg* 2007;194:497–500.

2. Di Cristofano C, Mrad K, Zavaglia K, *et al.* Papillary lesions of the breast: a molecular progression? *Breast Cancer Res Treat* 2005;90:71–6.

3. Douglas-Jones AG, Verghese A. Diagnostic difficulty arising from displaced epithelium after core biopsy in intracystic papillary lesions of the breast. *J Clin Pathol* 2002;55:780–3.

4. Solorzano CC, Middleton LP, Hunt KK, *et al.* Treatment and outcome of patients with intracystic papillary carcinoma of the breast. *Am J Surg* 2002;184:364–8.

5. Tsuda H, Takarabe T, Susumu N, *et al.* Detection of numerical and structural alterations and fusion of chromosomes 16 and 1 in low-grade papillary breast carcinoma by fluorescence in situ hybridization. *Am J Pathol* 1997;151:1027–34.

6. Tsuda H, Takarabe T, Akashi-Tanaka S, Fukutomi T, Hirohashi S. Pattern of chromosome 16q loss differs between an atypical proliferative lesion and an intraductal or invasive ductal carcinoma occurring subsequently in the same area of the breast. *Mod Pathol* 2001;14:382–8.

7. Boecker W, Buerger H, Schmitz K, *et al.* Ductal epithelial proliferations of the breast: a biological continuum? Comparative genomic hybridization and high-molecular-weight cytokeratin expression patterns. *J Pathol* 2001;195:415–21.

8. Linger RA, Park WS, Man YG, *et al.* LOH at 16p13 is a novel chromosomal alteration detected in benign and malignant microdissected papillary neoplasms of the breast. *Hum Pathol* 1998;29:1113–8.

9. Jacobs S, Thompson ER, Nannya Y, *et al.* Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. *Cancer Res* 2007;67:2544–51.

10. Lyons-Weiler M, Hagenkord J, Sciuilli C, Dhir R, Monzon FA. Optimization of the Affymetrix GeneChip Mapping 10K 2.0 Assay for routine clinical use on formalin-fixed paraffin-embedded tissues. *Diagn Mol Pathol* 2008;17:3–13.

11. Nessling M, Richter K, Schwaenen C, *et al.* Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res* 2005;65:439–47.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. The validation of Array-based comparative genomic hybridization by real time quantitative PCR.

Table S2. Regions of genomic alteration commonly shared in intracystic papillary carcinoma.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Spectrum of *MLL2* (*ALR*) Mutations in 110 Cases of Kabuki Syndrome

Mark C. Hannibal,^{1,2} Kati J. Buckingham,¹ Sarah B. Ng,³ Jeffrey E. Ming,⁴ Anita E. Beck,^{1,2} Margaret J. McMillin,² Heidi I. Gildersleeve,¹ Abigail W. Bigham,¹ Holly K. Tabor,^{1,2} Heather C. Mefford,^{1,2} Joseph Cook,¹ Koh-ichiro Yoshiura,⁵ Tadashi Matsumoto,⁵ Naomichi Matsumoto,⁶ Noriko Miyake,⁶ Hidefumi Tonoki,⁷ Kenji Naritomi,⁸ Tadashi Kaname,⁸ Toshiro Nagai,⁹ Hirofumi Ohashi,¹⁰ Kenji Kurosawa,¹¹ Jia-Woei Hou,¹² Tohru Ohta,¹³ Deshung Liang,¹⁴ Akira Sudo,¹⁵ Colleen A. Morris,¹⁶ Siddharth Banka,¹⁷ Graeme C. Black,¹⁷ Jill Clayton-Smith,¹⁷ Deborah A. Nickerson,³ Elaine H. Zackai,⁴ Tamim H. Shaikh,¹⁸ Dian Donnai,¹⁷ Norio Niikawa,¹³ Jay Shendure,³ and Michael J. Bamshad^{1,2,3*}

¹Department of Pediatrics, University of Washington, Seattle, Washington

²Seattle Children's Hospital, Seattle, Washington

³Department of Genome Sciences, University of Washington, Seattle, Washington

⁴Department of Pediatrics, The Children's Hospital of Philadelphia, The University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

⁵Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

⁶Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan

⁷Department of Pediatrics, Tenshi Hospital, Sapporo, Japan

⁸Department of Medical Genetics, University of the Ryukyus, Okinawa, Japan

⁹Department of Pediatrics, Dokkyo Medical University, Koshigaya Hospital, Saitama, Japan

¹⁰Division of Medical Genetics, Saitama Children's Medical Center, Saitama, Japan

¹¹Division of Clinical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan

¹²Department of Pediatrics, Chang Gung Children's Hospital, Taoyuan, Taiwan, Republic of China

¹³Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan

¹⁴National Laboratory of Medical Genetics, Xiangya Hospital, Central South University, Republic of China

¹⁵Department of Pediatrics, Sapporo City General Hospital, Sapporo, Japan

¹⁶University of Nevada School of Medicine, Las Vegas, Nevada

¹⁷Department of Genetic Medicine, Manchester Academic Health Sciences Centre, University of Manchester, England

¹⁸Department of Pediatrics, University of Colorado, Denver, Colorado

Received 25 February 2011; Accepted 30 March 2011

Additional supporting information may be found in the online version of this article.

Grant sponsor: National Institutes of Health/National Heart Lung and Blood Institute; Grant number: 5R01HL094976; Grant sponsor: National Institutes of Health/National Human Genome Research Institute; Grant numbers: 5R21HG004749, 1RC2HG005608, 5R01HG004316, T32HG00035; Grant sponsor: National Institute of Health/National Institute of Environmental Health Sciences; Grant number: HHSN273200800010C; Grant sponsor: National Institute of Neurological Disorders and Stroke; Grant number: RO1NS35102; Grant sponsor: NIHR Manchester Biomedical Research Centre; Grant sponsor: Ministry of Health, Labour and Welfare of Japan; Grant sponsor: Japan Science and Technology Agency; Grant sponsor: Society for the Promotion of Science; Grant sponsor: Life Sciences Discovery Fund;

Grant numbers: 2065508, 0905001; Grant sponsor: Washington Research Foundation; Grant sponsor: National Institutes of Health/National Institute of Child Health and Human Development; Grant numbers: 1R01HD048895, 5K23HD057331.

Mark C. Hannibal, Kati J. Buckingham, and Sarah B. Ng contributed equally to this work.

*Correspondence to:

Michael J. Bamshad, M.D., Department of Pediatrics, University of Washington School of Medicine, Box 356320, 1959 NE Pacific Street, Seattle, WA 98195. E-mail: mbamshad@u.washington.edu
Published online 10 June 2011 in Wiley Online Library
(wileyonlinelibrary.com).

DOI 10.1002/ajmg.a.34074

Kabuki syndrome is a rare, multiple malformation disorder characterized by a distinctive facial appearance, cardiac anomalies, skeletal abnormalities, and mild to moderate intellectual disability. Simplex cases make up the vast majority of the reported cases with Kabuki syndrome, but parent-to-child transmission in more than a half-dozen instances indicates that it is an autosomal dominant disorder. We recently reported that Kabuki syndrome is caused by mutations in *MLL2*, a gene that encodes a Trithorax-group histone methyltransferase, a protein important in the epigenetic control of active chromatin states. Here, we report on the screening of 110 families with Kabuki syndrome. *MLL2* mutations were found in 81/110 (74%) of families. In simplex cases for which DNA was available from both parents, 25 mutations were confirmed to be de novo, while a transmitted *MLL2* mutation was found in two of three familial cases. The majority of variants found to cause Kabuki syndrome were novel nonsense or frameshift mutations that are predicted to result in haploinsufficiency. The clinical characteristics of *MLL2* mutation-positive cases did not differ significantly from *MLL2* mutation-negative cases with the exception that renal anomalies were more common in *MLL2* mutation-positive cases. These results are important for understanding the phenotypic consequences of *MLL2* mutations for individuals and their families as well as for providing a basis for the identification of additional genes for Kabuki syndrome. © 2011 Wiley-Liss, Inc.

Key words: Kabuki syndrome; *MLL2*; *ALR*; Trithorax group histone methyltransferase

INTRODUCTION

Kabuki syndrome (OMIM#147920) is a rare, multiple malformation disorder characterized by a distinctive facial appearance, cardiac anomalies, skeletal abnormalities, and mild to moderate intellectual disability. It was originally described by Niikawa et al. [1981] and Kuroki et al. [1981] in 1981, and to date, about 400 cases have been reported worldwide [Niikawa et al., 1988; White et al., 2004; Adam and Hudgins, 2005]. The spectrum of abnormalities found in individuals with Kabuki syndrome is diverse, yet virtually all affected persons are reported to have similar facial features consisting of elongated palpebral fissures, eversion of the lateral third of the lower eyelids, and broad, arched eyebrows with lateral sparseness. Additionally, affected individuals commonly have severe feeding problems, failure to thrive in infancy, and height around or below the 3rd centile for age in about half of cases.

We recently reported that a majority of cases of Kabuki syndrome are caused by mutations in *mixed lineage leukemia 2* (*MLL2*; OMIM#602113), also known as either *MLL4* or *ALR* [Ng et al., 2010]. *MLL2* encodes a SET-domain-containing histone methyltransferase important in the epigenetic control of active chromatin states [FitzGerald and Diaz, 1999]. Exome sequencing revealed that 9 of 10 individuals had novel variants in *MLL2* that were predicted to be deleterious. A single individual had no mutation in the protein-coding exons of *MLL2*, though in

How to Cite this Article:

Hannibal MC, Buckingham KJ, Ng SB, Ming JE, Beck AE, McMillin MJ, Gildersleeve HI, Bigham AW, Tabor HK, Mefford HC, Cook J, Yoshiura K-i, Matsumoto T, Matsumoto N, Miyake N, Tonoki H, Naritomi K, Kaname T, Nagai T, Ohashi H, Kurosawa K, Hou J-W, Ohta T, Liang D, Sudo A, Morris CA, Banka S, Black GC, Clayton-Smith J, Nickerson DA, Zackai EH, Shaikh TH, Donnai D, Niikawa N, Shendure J, Bamshad MJ. 2011. Spectrum of *MLL2* (*ALR*) mutations in 110 cases of Kabuki syndrome.

Am J Med Genet Part A 155:1511–1516.

retrospect, his phenotypic features are somewhat atypical of Kabuki syndrome. In a larger validation cohort screened by Sanger sequencing, we found *MLL2* mutations in approximately two-thirds of 43 Kabuki cases, suggesting that Kabuki syndrome is genetically heterogeneous.

Herein we report on the results of screening *MLL2* for mutations in 110 families with one or more individuals affected with Kabuki syndrome in order to: (1) characterize the spectrum of *MLL2* mutations that cause Kabuki syndrome; (2) determine whether *MLL2* genotype is predictive of phenotype; (3) assess whether the clinical characteristics of *MLL2* mutation-positive cases differ from *MLL2* mutation-negative cases; and (4) delineate the subset of Kabuki cases that are *MLL2* mutation-negative for further gene discovery studies.

MATERIALS AND METHODS

Subjects

Referral for inclusion into the study required a diagnosis of Kabuki syndrome made by a clinical geneticist. From these cases, phenotypic data were collected by review of medical records, phone interviews, and photographs. These data were collected from five different clinical genetics centers in three different countries and over a protracted period of time and forwarded for review to two of the authors (M.B. and M.H.). Data on certain phenotypic characteristics including stature, feeding difficulties, and failure to thrive was not uniformly collected or standardized. Therefore, we decided to be conservative in our analysis and use only phenotypic traits that could be represented by discrete variables (i.e., presence or absence) and for which data were available from at least 70% of cases. In addition, these clinical summaries were de-identified and therefore facial photographs were unavailable from most cases studied. Written consent was obtained for all participants who provided identifiable samples. The Institutional Review Boards of Seattle Children's Hospital and the University of Washington approved all studies. A summary of the clinical characteristics of 53 of these individuals diagnosed with Kabuki syndrome has been reported previously [Ng et al., 2010].

Mutation Analysis

Genomic DNA was extracted using standard protocols. Each of the 54 exons of *MLL2* was amplified using Taq DNA polymerase (Invitrogen, Carlsbad, CA) following manufacturer’s recommendations and using primers previously reported [Ng et al., 2010]. PCR products were purified by treatment with exonuclease I (New England Biolabs, Inc., Beverly, MA) and shrimp alkaline phosphatase (USB Corp., Cleveland, OH), and products were sequenced using the dideoxy terminator method on an automated sequencer (ABI 3130xl). The electropherograms of both forward and reverse strands were manually reviewed using CodonCode Aligner (Dedham, MA). Primer sequences and conditions are listed in Supplementary Table I.

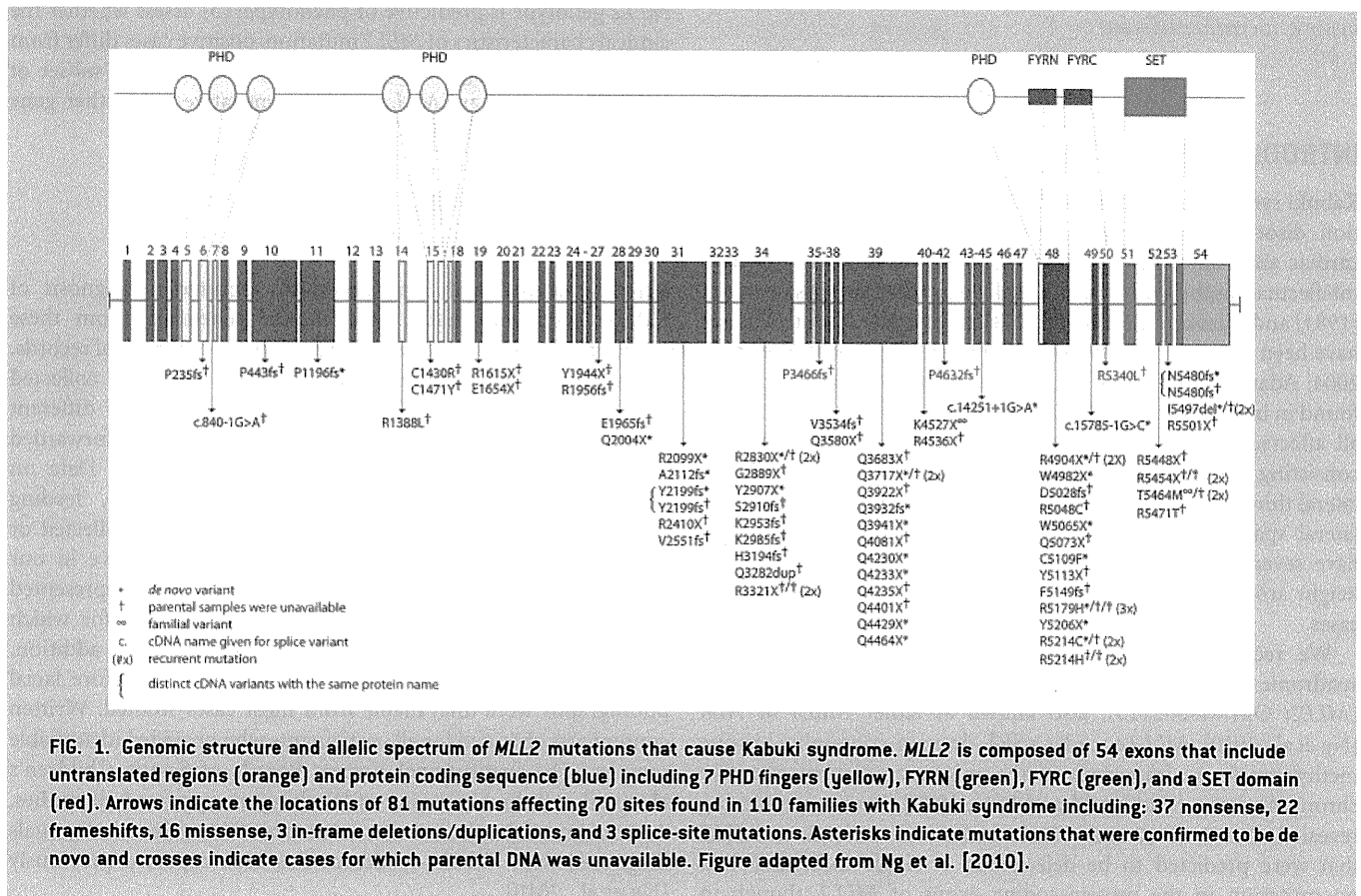
For *MLL2* mutation-negative samples, DNA was hybridized to commercially available whole-genome tiling arrays consisting of one million oligonucleotide probes with an average spacing of 2.6 kb throughout the genome (SurePrint G3 Human CGH Microarray 1 × 1 M, Agilent Technologies, Santa Clara, CA). Twenty-one probes on this array covered *MLL2* specifically. Data were analyzed using Genomics Workbench software according to manufacturer’s instructions.

RESULTS

All 54 protein-coding exons and intron–exon boundaries of *MLL2* were screened by Sanger sequencing in a cohort of 110 kindreds with

Kabuki syndrome. This cohort included 107 simplex cases (including a pair of monozygotic twins) and 3 familial (i.e., parent-offspring) cases putatively diagnosed with Kabuki syndrome. Seventy novel *MLL2* variants that were inferred to be disease-causing were identified in 81/110 (74%) kindreds (Fig. 1 and Supplementary Table II online). These 81 mutations included 37 nonsense mutations (32 different sites and five sites with recurrent mutations), 3 in-frame deletions or duplications (2 different sites and 1 site with a recurrent mutation), 22 frameshifts (22 different sites), 16 missense mutations (11 different sites and 4 sites with recurrent mutations), and 3 splice consensus site (or intron–exon boundary) mutations. None of these variants were found in dbSNP (build 132), the 1000 Genomes Project pilot data, or 190 chromosomes from individuals matched for geographical ancestry. In total, pathogenic variants were found at 70 sites. Additionally, there were 10 sites at which recurrent mutations were observed.

For 25 simplex cases in which we identified *MLL2* mutations, DNA was available from both unaffected parents, and in each case the mutation was confirmed to have arisen de novo (Supplementary Table II online). These included 14 nonsense, 5 frameshift, 3 missense, 2 splice site mutations, and 1 deletion. De novo events were confirmed at 6 of the 10 sites where recurrent mutations were noted. In addition to the 81 kindreds in which we identified causal *MLL2* mutations, we found two *MLL2* variants in each of three simplex cases. In each case, neither *MLL2* mutation could unambiguously



be defined as disease-causing (Supplementary Table II online). In one case, we found both a 21 bp in-frame insertion in exon 39 and a 1 bp insertion in exon 46 predicted to cause a frameshift. However, the unaffected mother also carried the 21 bp insertion suggesting that this is a rare polymorphism, and that the 1 bp deletion is the pathogenic mutation responsible for Kabuki syndrome.

Apparent disease-causing variants were discovered in nearly half (i.e., 22/54) of all protein-coding exons of *MLL2* and in virtually every region known to encode a functional domain (Fig. 1). However, the distribution of variants appeared non-random as 13 and 12 novel variants were identified in exons 48 and 39, respectively. These sites accounted for 25, or more than one-third, of all the novel *MLL2* variants and 31/81 mutations that cause Kabuki syndrome in our cohort. Eleven of the 12 pathogenic variants in exon 39 were nonsense mutations and occurred in regions that encode long polyglutamine tracts.

Four of the families studied herein had two individuals affected with Kabuki syndrome. A pair of monozygous twins with a c.15195G>A nonsense mutation were concordant for mild developmental delay, congenital heart disease, preauricular pits, and palatal abnormalities, but discordant for hearing loss, and a central nervous system malformation. Concordance for mild developmental delay between an affected parent and child was observed in two families with *MLL2* mutations, one with a nonsense mutation, c.13579A>T, p.K4527X, and the other with a missense mutation, c.16391C>T, p.T5464M that was also found in a simplex case. No *MLL2* mutation was found in the remaining affected parent and child pair (Fig. 2).

To examine the relationship between genotype and phenotype, we first compared the frequency of developmental delay, congenital heart disease, cleft lip and/or palate, and structural renal defects between *MLL2* mutation-positive versus *MLL2* mutation-negative cases. No significant difference was observed between groups for three of these four phenotypes (Table Ia). However, renal anomalies were observed in 47% (31/66 cases) of *MLL2* mutation-positive cases compared to 14% (2/14 cases) of *MLL2* mutation-negative cases and this difference was statistically significant ($\chi^2 = 5.1$, $df = 1$, $P = 0.024$). In 35 cases in two clinical cohorts for whom more complete phenotypic data were available, short stature was observed in 54% (14/26) of *MLL2* mutation-positive cases compared to 33% (3/19 cases) of *MLL2* mutation-negative cases. We also divided the *MLL2* mutation-positive cases into those with nonsense and frameshift mutations and those with missense mutations and compared the frequency of developmental delay, congenital heart disease, cleft lip and/or palate, and structural renal defects between groups. No significant differences were observed between groups (Table Ib).

In 26 independent cases of Kabuki syndrome, including one parent-offspring pair, no *MLL2* mutation was identified. Both persons in the mother-child pair had facial characteristics consistent with Kabuki syndrome (Fig. 2), mild developmental delay, and no major malformations. The mother is of Cambodian ancestry and her daughter is of Cambodian and European American ancestry. In general, most of the *MLL2* mutation-negative Kabuki cases had facial characteristics (Fig. 3) similar to those of the *MLL2* mutation-positive Kabuki cases, and a similar pattern of major malformations (Table I) with the exception of fewer renal abnormalities.

TABLE I. Phenotypic Traits Grouped by *MLL2* Mutation Status [a] and Type [b]

Trait	<i>MLL2</i> +	<i>MLL2</i> -
Intellectual disability	74/74 (100%)	19/20 (95%)
Mild	51/74 (69%)	10/20 (50%)
Moderate	18/74 (24%)	4/20 (20%)
Severe	4/74 (5%)	3/20 (15%)
Cleft palate, CL/CP	29/72 (40%)	8/18 (44%)
Congenital heart defect	36/71 (51%)	8/19 (42%)
Renal abnormality	31/66 (47%)	2/14 (14%)
Trait	Truncating (N = 59)	Missense (N = 16)
Intellectual disability	54/54 (100%)	15/15 (100%)
Mild	36/54 (67%)	11/15 (73%)
Moderate	13/54 (24%)	4/15 (27%)
Severe	5/54 (9%)	0/15
Cleft palate, CL/CP	23/54 (43%)	3/14 (21%)
Congenital heart defect	30/54 (55%)	4/13 (30%)
Renal anomaly	9/44 (20%)	2/12 (17%)

We screened the *MLL2* mutation-negative cases by aCGH for large deletions or duplications that encompassed *MLL2*. Abnormalities were found in four cases. In one case, a 1.87 kb deletion of chromosome 5 (hg18, chr5:175,493,803–177,361,744) that included *NSD1* and had breakpoints in flanking segmental duplications identical to the microdeletion commonly found in Sotos syndrome, was found. This suggests that this individual has Sotos syndrome, not Kabuki syndrome [Kurotaki et al., 2002]. A second case had a novel 977-kb deletion of chromosome 19q13 (hg18, chr19:61,365,420–62,342,064) encompassing 20 genes. The majority of genes within the deleted region are zinc finger genes, some of which are known to be imprinted in both human and mouse. A third case had a complex translocation t(8;18)(q22;q21). Finally, a fourth case was found to have extra material for the entire chromosome 12. Average log₂ ratio across chromosome 12 was 0.49, most likely representing mosaic aneuploidy of chromosome 12. No aCGH abnormalities were observed in 21 cases and aCGH failed for one case.

DISCUSSION

We have expanded the spectrum of mutations in *MLL2* that cause Kabuki syndrome and explored the relationship between *MLL2* genotype and some of the major, objective phenotypic characteristics of Kabuki syndrome. The majority of variants found to cause Kabuki syndrome are either novel nonsense or frameshift mutations, and appear to arise de novo. While mutations that cause Kabuki syndrome are found throughout the *MLL2* gene, there appear to be at least two exons (39 and 48) in which mutations are identified with a considerably higher frequency. Mutations in these two exons account for nearly half of all mutations found in *MLL2*, while the length of these exons represents ~24% of the *MLL2* open reading frame (ORF). Furthermore, exon 48, the exon in which mutations are most common, comprises only ~7% of the

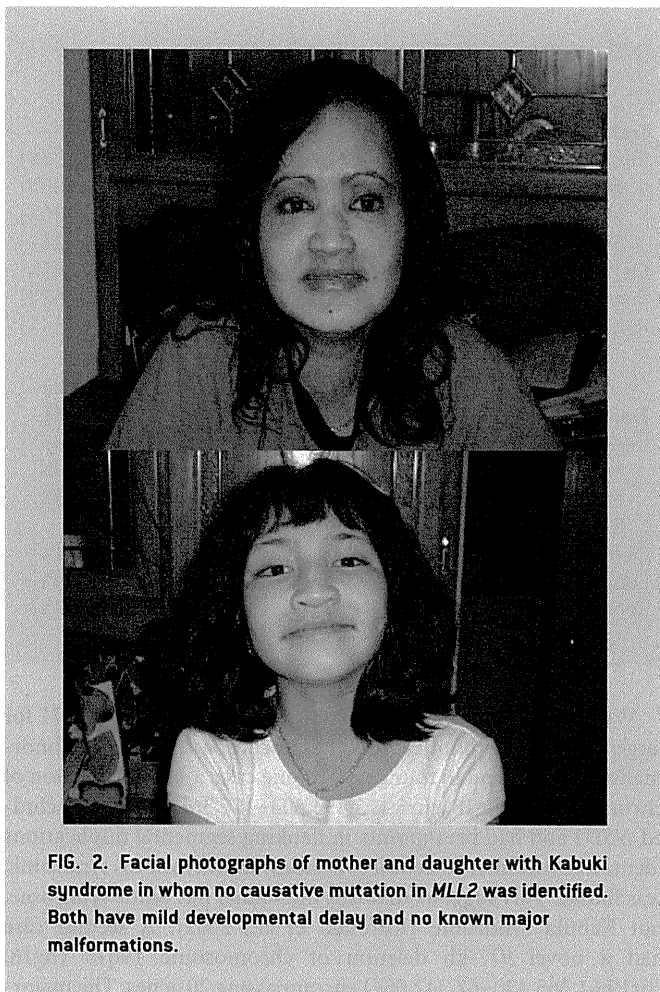


FIG. 2. Facial photographs of mother and daughter with Kabuki syndrome in whom no causative mutation in *MLL2* was identified. Both have mild developmental delay and no known major malformations.

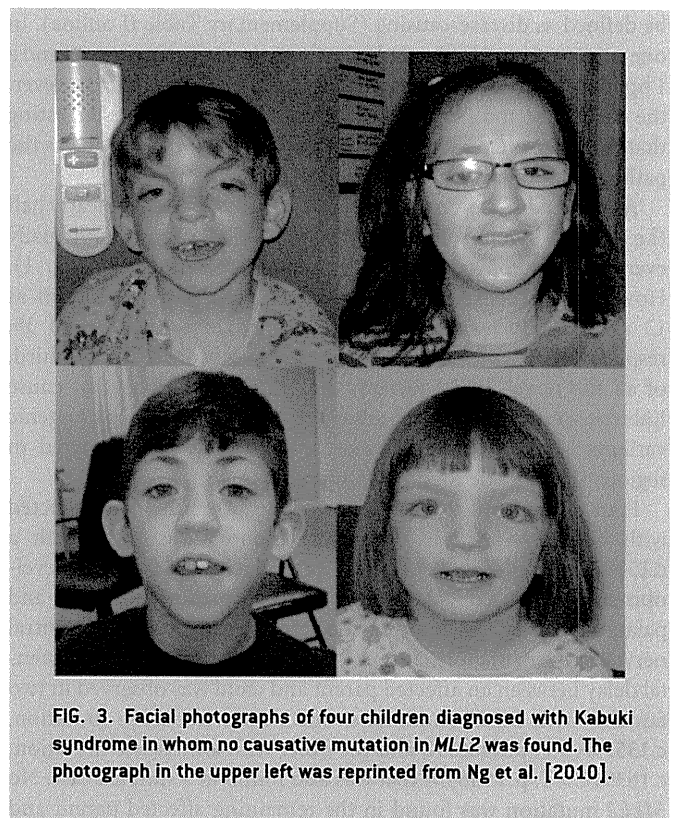


FIG. 3. Facial photographs of four children diagnosed with Kabuki syndrome in whom no causative mutation in *MLL2* was found. The photograph in the upper left was reprinted from Ng et al. [2010].

MLL2 ORF. Exon 39 contains several regions that encode long polyglutamine tracts suggesting the presence of a mutational hotspot, although no such explanation is obvious for exon 48. A stepwise approach in which these regions are the first screened might be a reasonable approach to diagnostic testing. However, capture of all introns, exons, and nearby *MLL2* regulatory regions followed by next-generation sequencing would be more comprehensive and likely to be less costly over the long term.

Comparison of four of the objective clinical characteristics of *MLL2* mutation-negative versus *MLL2* mutation-positive cases allowed us to explore both the relationship between *MLL2* genotype and Kabuki phenotype and the phenotype of *MLL2* mutation-negative cases. Overall, the clinical characteristics of *MLL2* mutation-positive cases did not differ significantly from *MLL2* mutation-negative cases with the exception that renal anomalies were more common in *MLL2* mutation-positive cases. Similarly, we observed no significant phenotypic—including the severity of developmental delay—differences between individuals grouped by mutation type. However, the phenotypic data available to us for analysis was limited and, for many cases, we lacked specific information about each malformation present. Furthermore, the most typical phenotypic characteristic, the distinctive facial appearance,

was not compared in detail between cases although it would be of interest to study facial images “blinded” to mutation status to investigate its power to predict genotype. Analysis of genotype–phenotype relationships using both a larger set of Kabuki cases, and with access to more comprehensive phenotypic information would be valuable.

No *MLL2* mutation could be identified in 26 of the cases referred to us with a diagnosis of Kabuki syndrome. In three of these cases, aCGH identified structural variants that could be of clinical significance although additional investigation is required. A fourth case had the classical deletion observed in individuals with Sotos syndrome, and in retrospect it appears that this case was included in the cohort erroneously. The 22 remaining cases, including 1 parent-offspring pair, represent individuals with fairly classic phenotypic features of Kabuki syndrome without a *MLL2* mutation. This observation suggests that Kabuki syndrome is genetically heterogeneous. To this end, in these 22 cases, we sequenced the protein-coding exons of *UTX*, a gene that encodes a protein that directly interacts with *MLL2* but no pathogenic changes were found (data not shown). Exome sequencing of a subset of these *MLL2* mutation-negative cases to identify other candidate genes for Kabuki syndrome is underway.

Whether Kabuki syndrome is the most appropriate diagnosis for the *MLL2* mutation-negative cases is unclear. Some of the *MLL2* mutation-negative cases appear to have a facial phenotype that differs somewhat from that of the *MLL2* mutation-positive cases. Whether these *MLL2* mutation-negative cases diagnosed by expert clinicians should be considered Kabuki syndrome, a variant thereof, or a separate disorder remains to be determined. Our opinion is that

there is simply not yet enough information to make an informed decision about this issue.

Most of the mutations in *MLL2* are predicted to result in haploinsufficiency. However, it is unclear by what mechanism(s) haploinsufficiency of *MLL2* could cause Kabuki syndrome. *MLL2* encodes a histone 3 lysine 4 (H3K4) methyltransferase, one of at least 10 proteins (genes for which have not to our knowledge yet been screened in Kabuki cases in which *MLL2* mutations were not found) that have been identified to specifically modify the lysine residue at the fourth amino acid position of the histone H3 protein [Kouzarides, 2007]. *MLL2* has a SET domain near its C-terminus that is shared by yeast Set1, *Drosophila* Trithorax (TRX) and human MLL1 [FitzGerald and Diaz, 1999]. *MLL2* appears to regulate gene transcription and chromatin structure in early development [Prasad et al., 1997]. In mice, loss of *MLL2* results in embryonic lethality before E10.5, and while *MLL2*^{+/-} mice are viable, they are smaller than wild-type [Ng et al., 2010].

Kabuki syndrome is the most common of a small, but growing group of multiple malformation syndromes accompanied by developmental delay that are caused by mutations in genes that encode proteins involved in histone methylation [De Sario, 2009]. The most notable of these is CHARGE syndrome, which is one of the syndromes often considered in the differential diagnosis of children ultimately diagnosed with Kabuki syndrome. CHARGE syndrome is caused by mutations in *CHD7*, which encodes a chromodomain protein that recognizes the trimethylated H3K4 side chain [Vissers et al., 2004]. Other disorders caused by defects of histone methylation status include several intellectual disability syndromes, some of which are also characterized by malformations (e.g., cleft lip/palate) that overlap with those found in individuals with Kabuki syndrome.

Kabuki syndrome is one of the most common causes of heritable developmental delay. Discovery that mutations in *MLL2* are the most common cause of Kabuki syndrome highlights the role that disrupted regulation of histone methylation plays as a cause of human birth defects. Characterizing the spectrum of mutations in *MLL2* is a small but important first step toward understanding the mechanism(s) that underlies Kabuki syndrome.

ACKNOWLEDGMENTS

We thank the families for their participation and the Kabuki Syndrome Network for their support. Our work was supported in part by grants from the National Institutes of Health/National Heart Lung and Blood Institute (5R01HL094976 to D.A.N. and J.S.), the National Institutes of Health/National Human Genome Research Institute (5R21HG004749 to J.S., 1RC2HG005608 to M.J.B., D.A.N., and J.S.; and 5R01HG004316 to H.K.T.), National Institute of Health/National Institute of Environmental Health Sciences (HHSN273200800010C to D.N.), National Institute of Neurological Disorders and Stroke (RO1NS35102 to C.A.M.), NIHR Manchester Biomedical Research Centre (D. D.), Ministry of Health, Labour and Welfare (K.Y., N.M., T.O., and N.N.), Ministry of Health, Labour and Welfare of Japan (N.M.), Japan Science and Technology Agency (N.M.), Society for the Promotion

of Science (N.M.), the Life Sciences Discovery Fund (2065508 and 0905001), the Washington Research Foundation, and the National Institutes of Health/National Institute of Child Health and Human Development (1R01HD048895 to M.J.B. and 5K23HD057331 to A.E.B.). S.B.N. is supported by the Agency for Science, Technology and Research, Singapore. A.W.B. is supported by a training fellowship from the National Institutes of Health/National Human Genome Research Institute (T32HG00035).

REFERENCES

- Adam MP, Hudgins L. 2005. Kabuki syndrome: A review. *Clin Genet* 67: 209–219.
- De Sario A. 2009. Clinical and molecular overview of inherited disorders resulting from epigenomic dysregulation. *Eur J Med Genet* 52:363–372.
- FitzGerald KT, Diaz MO. 1999. MLL2: A new mammalian member of the trx/MLL family of genes. *Genomics* 59:187–192.
- Kouzarides T. 2007. Chromatin modifications and their function. *Cell* 128:693–705.
- Kuroki Y, Suzuki Y, Chyo H, Hata A, Matsui I. 1981. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. *J Pediatr* 99:570–573.
- Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, Ohashi H, Naritomi K, Tsukahara M, Makita Y, Sugimoto T, Sonoda T, Hasegawa T, Chinen Y, Tomita H, Kinoshita HA, Mizuguchi A, Yoshiura T, Ki K, Ohta T, Kishino T, Fukushima Y, Niikawa N, Matsumoto N. 2002. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat Genet* 30:365–366.
- Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. 2010. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 42:790–793.
- Niikawa N, Matsuura N, Fukushima Y, Ohsawa T, Kajii T. 1981. Kabuki make-up syndrome: A syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *J Pediatr* 99:565–569.
- Niikawa N, Kuroki Y, Kajii T, Matsuura N, Ishikiriyama S, Tonoki H, Ishikawa N, Yamada Y, Fujita M, Umemoto H, et al. 1988. Kabuki make-up (Niikawa-Kuroki) syndrome: A study of 62 patients. *Am J Med Genet* 31:565–589.
- Prasad R, Zhadanov AB, Sedkov Y, Bullrich F, Druck T, Rallapalli R, Yano T, Alder H, Croce CM, Huebner K, Mazo A, Canaani E. 1997. Structure and expression pattern of human ALR, a novel gene with strong homology to ALL-1 involved in acute leukemia and to *Drosophila* trithorax. *Oncogene* 15:549–560.
- Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG. 2004. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 36:955–957.
- White SM, Thompson EM, Kidd A, Savarirayan R, Turner A, Amor D, Delatycki MB, Fahey M, Baxendale A, White S, Haan E, Gibson K, Halliday JL, Bankier A. 2004. Growth, behavior, and clinical findings in 27 patients with Kabuki (Niikawa-Kuroki) syndrome. *Am J Med Genet Part A* 127A:118–127.

Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks

Hisato Kobayashi¹, Takayuki Sakurai¹, Misaki Imai², Nozomi Takahashi¹, Atsushi Fukuda¹, Obata Yayoi¹, Shun Sato³, Kazuhiko Nakabayashi³, Kenichiro Hata³, Yusuke Sotomaru⁴, Yutaka Suzuki⁵, Tomohiro Kono^{1,2*}

1 Department of BioScience, Tokyo University of Agriculture, Tokyo, Japan, **2** Genome Research Center, NODAI Research Institute, Tokyo University of Agriculture, Tokyo, Japan, **3** Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan, **4** Natural Science Center for Basic Research and Development, Hiroshima University, Hiroshima, Japan, **5** Department of Medical Genome Sciences, Graduate School of Frontier, The University of Tokyo, Kashiwa, Japan

Abstract

Genome-wide dynamic changes in DNA methylation are indispensable for germline development and genomic imprinting in mammals. Here, we report single-base resolution DNA methylome and transcriptome maps of mouse germ cells, generated using whole-genome shotgun bisulfite sequencing and cDNA sequencing (mRNA-seq). Oocyte genomes showed a significant positive correlation between mRNA transcript levels and methylation of the transcribed region. Sperm genomes had nearly complete coverage of methylation, except in the CpG-rich regions, and showed a significant negative correlation between gene expression and promoter methylation. Thus, these methylome maps revealed that oocytes and sperms are widely different in the extent and distribution of DNA methylation. Furthermore, a comparison of oocyte and sperm methylomes identified more than 1,600 CpG islands differentially methylated in oocytes and sperm (germline differentially methylated regions, gDMRs), in addition to the known imprinting control regions (ICRs). About half of these differentially methylated DNA sequences appear to be at least partially resistant to the global DNA demethylation that occurs during preimplantation development. In the absence of *Dnmt3L*, neither methylation of most oocyte-methylated gDMRs nor intragenic methylation was observed. There was also genome-wide hypomethylation, and partial methylation at particular retrotransposons, while maintaining global gene expression, in oocytes. Along with the identification of the many *Dnmt3L*-dependent gDMRs at intragenic regions, the present results suggest that oocyte methylation can be divided into 2 types: *Dnmt3L*-dependent methylation, which is required for maternal methylation imprinting, and *Dnmt3L*-independent methylation, which might be essential for endogenous retroviral DNA silencing. The present data provide entirely new perspectives on the evaluation of epigenetic markers in germline cells.

Citation: Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, et al. (2012) Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. *PLoS Genet* 8(1): e1002440. doi:10.1371/journal.pgen.1002440

Editor: Wolf Reik, The Babraham Institute, United Kingdom

Received: July 28, 2011; **Accepted:** November 14, 2011; **Published:** January 5, 2012

Copyright: © 2012 Kobayashi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan (Grant Nos. 222228004, 20062009, 22150002, 50801025). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tomohiro@nodai.ac.jp

Introduction

Throughout mammalian gametogenesis, dynamic DNA methylation changes occur in a sex- and sequence-specific manner. These changes result in the establishment of oocyte- and sperm-specific genomic imprints and unique methylation patterns of repetitive elements via DNA methyltransferase activity [1–4]. This process is indispensable for functional gamete and embryo development. For example, sex-specific methylation imprints are maintained throughout cell division after fertilization, despite genome-wide demethylation and *de novo* methylation during embryogenesis. These imprints control parent-of-origin specific monoallelic expression of a subset of genes, which are known as imprinted genes [5–9]. In addition, DNA methylation during spermatogenesis plays a crucial role in meiotic progression and

retrotransposon silencing [10–14]. However, little is known about the profile and functional role of DNA methylation during oogenesis, except for the establishment of genomic imprints.

Recently, the epigenetic modifications which are responsible for regulating cell differentiation and embryo development have been studied in detail by using high-throughput sequencing: bisulfite sequencing (“BS-seq”); “Methyl-seq” with a methyl-sensitive restriction enzyme; “MeDIP-seq” with methylated DNA immunoprecipitation; and “MBD-seq” with a methyl-DNA binding domain protein antibody [15–26]. However, a major limitation of epigenomic studies is the lack of a standard methodology for DNA methylome analysis. Ideally, the gold standard is high resolution and genome-wide methylome analysis of germ cells. However, genome-wide methylome analysis of female germ cells has almost never been performed due to the limited availability of samples.

Author Summary

In mammals, germ-cell-specific methylation patterns and genomic imprints are established throughout large-scale de novo DNA methylation in oogenesis and spermatogenesis. These steps are required for normal germline differentiation and embryonic development; however, current DNA methylation analyses only provide us a partial picture of germ cell methylome. To the best of our knowledge, this is the first study to generate comprehensive maps of DNA methylomes and transcriptomes at single base resolution for mouse germ cells. These methylome maps revealed genome-wide opposing DNA methylation patterns and differential correlation between methylation and gene expression levels in oocyte and sperm genomes. In addition, our results indicate the presence of 2 types of methylation patterns in the oocytes: (i) methylation across the transcribed regions, which might be required for the establishment of maternal methylation imprints and normal embryogenesis, and (ii) retroviral methylation, which might be essential for silencing of retrotransposons and normal oogenesis. We believe that an extension of this work would lead to a better understanding of the epigenetic reprogramming in germline cells and of the role for gene regulations.

of the 21 million cytosines of CpGs in the mouse genome were covered by at least 1 sequence read from GV oocytes and sperm, respectively; whereas the average read depth (*i.e.*, the number of hits of reads that were mapped to a given position) was over 10× for both germ cells (Figure S2). The WBA-seq method generated 307 and 397 million tags from GV oocytes obtained from wild-type and *Dnmt3L*-deficient (*Dnmt3L*^{-/-}) mice, respectively. WBA-seq libraries for GV oocytes showed higher genome coverage (60% of genomic CpGs were covered by at least 1 read) but with smaller average read depth (7.4×) than MethylC-seq library. Some reads from the oocyte libraries strongly matched mitochondrial DNA (mtDNA), satellite, low complexity, or simple repeat sequences (Figure S3), which might have been due to a distinct genomic copy number bias in the mitochondria of germ cells or an over-amplification bias. Thus, SBS results were simplified by removing the redundancy information (only mtDNA was separately examined for DNA methylation) and combining MethylC-seq and WBA-seq results for wild-type oocytes. Consequently, the average read depth was 18.8×, 4.4×, and 12.5× for wild-type and *Dnmt3L*^{-/-} oocytes, and sperm, respectively, and 70.8%, 45.6%, and 79.9% of genomic CpGs were covered by at least 1 sequence read from each cell type (Table 1 and Figure S3). Furthermore, the average read depths of MethylC-seq of mouse blastocysts and embryonic stem cells (ESCs), which served as zygote and stem cell controls, were 12.8× and 6.1×, respectively (Table 1).

Shotgun bisulfite sequencing (SBS) may be able to overcome this limitation and enable the determination of the cytosine methylation status of individual CpG sites at a whole-genome level without a bias toward CpG-rich regions [22,23,26] and with only relatively small-scale DNA samples [24,27]. As a result, in this study, an improved SBS method for small-scale DNA samples was used to analyze the DNA methylome of mouse germ cells. In addition, the mouse germ cell transcriptome was investigated using high-throughput cDNA sequencing (mRNA-seq) to reveal relationships between DNA methylation and gene transcription in both male and female germ cells.

Results

Genome sequencing

We performed SBS analysis by using MethylC-seq [22] and a new SBS method called “whole bisulfite-amplified DNA sequencing” (WBA-seq). The MethylC-seq and WBA-seq libraries were generated as shown in Figure S1. The MethylC-seq method generated 1010 and 1085 million tags (reads) from germinal vesicle (GV) stage oocytes and epididymal sperm, respectively. Oocyte DNA libraries generated by MethylC-seq showed higher redundancies than sperm DNA libraries. For example, 33.0% and 81.7%

Methylome of mouse germ cells

The average methylation level of wild-type oocytes (40.0%) was less than half that of sperm (89.4%) (Figure S4). This difference in global DNA methylation between male and female germ cells was consistent with results from the previous studies [28,29]. The *Dnmt3L*^{-/-} oocyte genome was observed to be hypomethylated, exhibiting a methylation level of only 5.5%. Furthermore, blastocysts showed a lesser extent of methylation (21.3%) than did wild-type oocytes; ESCs, on the other hand, showed relatively high levels of methylation (70.6%). To elucidate the distribution of methylation levels on CpG sites, on regional and genome-wide scales, we created dot plots of CpG methylation for individual chromosomes and histograms of the methylation levels for all CpGs. These graphs revealed that hypermethylated CpGs in oocytes tended to cluster in transcribed regions of particular genes (*e.g.*, *Kcnq1* or *Rlim* genes, known to be expressed in oocytes [30,31]); the sperm genome was almost entirely hypermethylated, except at most CpG-rich regions (Figure 1 and Figure S5). Specifically, 55.7% of the CpGs in the oocyte genome exhibited <10% methylation, whereas another 32.0% of CpGs exhibited ≥90% methylation (Figure 2A). The *Dnmt3L*^{-/-} oocyte genome was also hypomethylated in almost all chromosomal regions (Figure S6). The methylation level of the mtDNA genome in

Table 1. Summary of shotgun bisulfite sequencing data.

Sample	Method	Aligned tags (base)	Genome coverage		Read depth
			(>x1)	(>x5)	
Wild-type oocyte	MethylC-seq & WBA-seq	51,166,451,066	70.8%	39.4%	18.8
<i>Dnmt3L</i> ^{-/-} oocyte	WBA-seq	11,872,662,647	45.6%	19.6%	4.4
Sperm	MethylC-seq	34,153,237,944	79.9%	63.4%	12.5
Blastocyst	MethylC-seq	34,857,014,339	86.2%	79.4%	12.8
ESC	MethylC-seq	16,691,289,063	73.0%	38.9%	6.1

doi:10.1371/journal.pgen.1002440.t001

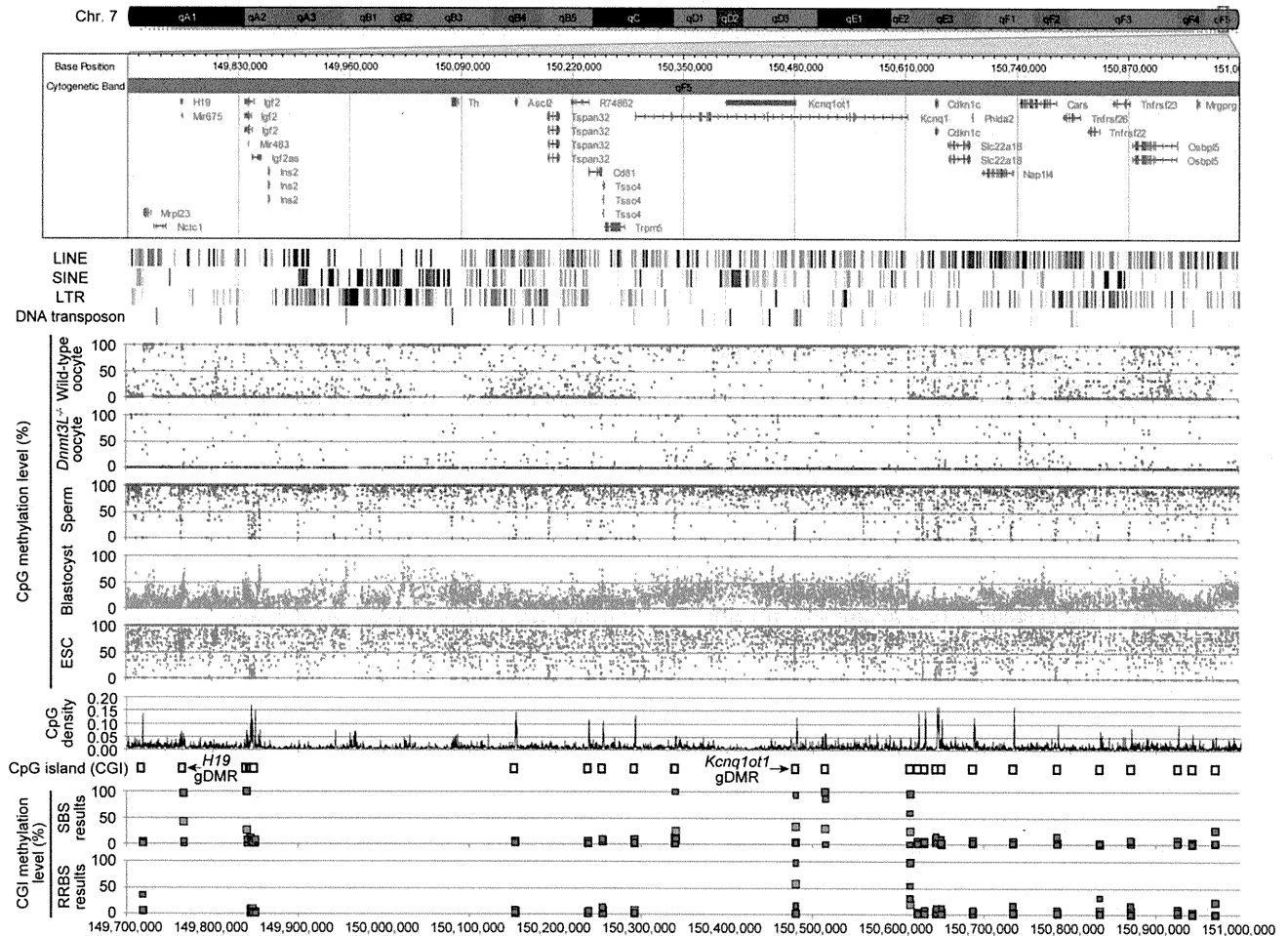


Figure 1. High-resolution DNA methylome map of mouse distal chromosome 7 imprinting cluster. Illumina GenomeStudio viewer displays the locations of genes in distal chromosome 7 (149,700,000–151,000,000). Black vertical bars represent the location of 4 repetitive elements: LINE, SINE, LTR, and DNA transposons. Red, purple, blue, green, and khaki dots represent the methylation levels at individual CpGs in wild-type oocyte, *Dnmt3L*^{-/-} oocyte, sperm, blastocyst, and ESC genomes, respectively. Black line plots depict the distribution of CpG densities (number of CpG per 200 nt) of individual CpGs. Open boxes represent the location of CpG islands (CGIs). Red, purple, blue, and green boxes represent the methylation levels at individual CGIs in wild-type oocyte, *Dnmt3L*^{-/-} oocyte, sperm, and blastocyst genomes, respectively, determined by our results from shotgun bisulfite sequencing (SBS) method and Smallwood's results from reduced representation bisulfite sequencing (RRBS) method [38]. doi:10.1371/journal.pgen.1002440.g001

Dnmt3L^{-/-} oocytes (4.4%) was lower than that observed in wild-type oocytes (6.6%). Sperm methylation levels, by comparison, were relatively high (14.7%), whereas those of the blastocysts and ESCs were quite low (1.3% and 2.1%, respectively) (Figure S4).

Since previous studies revealed a significant correlation between CpG frequency and methylation within intra- and intergenic regions in somatic cells [32,33], the CpG density and methylation levels were compared to identify genome-wide differential methylation patterns in germ cells. CpG density was defined as the number of CpG dinucleotides in 200 nucleotide (nt) windows (e.g., 1 CpG dinucleotide per 200 nt corresponds to a density of 0.005). At low CpG densities (range, 0.005–0.05), the oocyte genome was about 50% methylated, whereas the sperm genome was 80–90% methylated. At moderate to high CpG densities (range, 0.05–0.2), both male and female germ cells were hypomethylated (Figure 2B). Furthermore, 4 families of transposable elements (long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), long terminal repeats (LTRs), and DNA transposons) were moderately methylated in oocyte genomes but were hypermethylated in sperm. In addition, a general trend towards higher methylation levels at higher CpG densities in the oocyte genome

occurred in LTRs. Conversely, a trend toward lower CpG methylation levels at higher CpG densities in the wild-type oocyte and sperm genomes was observed in SINEs and DNA transposons. In contrast, all of these transposable elements were hypomethylated in *Dnmt3L*^{-/-} oocytes. Interestingly, however, there was partial CpG methylation in LINEs and LTRs at relatively high CpG densities (range, 0.03–0.1). These complete or partial undermethylations were confirmed by bisulfite sequencing in L1 LINEs, B1/Alu SINEs, and intracisternal A particle (IAP) LTRs (Figure S7). These results suggested that each germ cell has a unique sequence- and CpG-density-dependent methylation pattern. In addition, oocyte CpG methylation, except in a subset of retrotransposons, appears to be *Dnmt3L* dependent.

We also characterized the methylation patterns of 15 germline-differentially methylated regions (gDMRs). The differential (between oocyte and sperm) methylation occurs at imprinted gene loci (also called imprinting control regions (ICRs)). The ICRs of maternally methylated imprinted genes (e.g., *Nespa5-Gnas*) were shown to be hypermethylated in oocytes but hypomethylated in sperm, while the converse was true in ICRs of paternally-methylated imprinted genes (e.g., *H19*) (Figure 3 and Figure S8). Interestingly, only the *Snrpn*

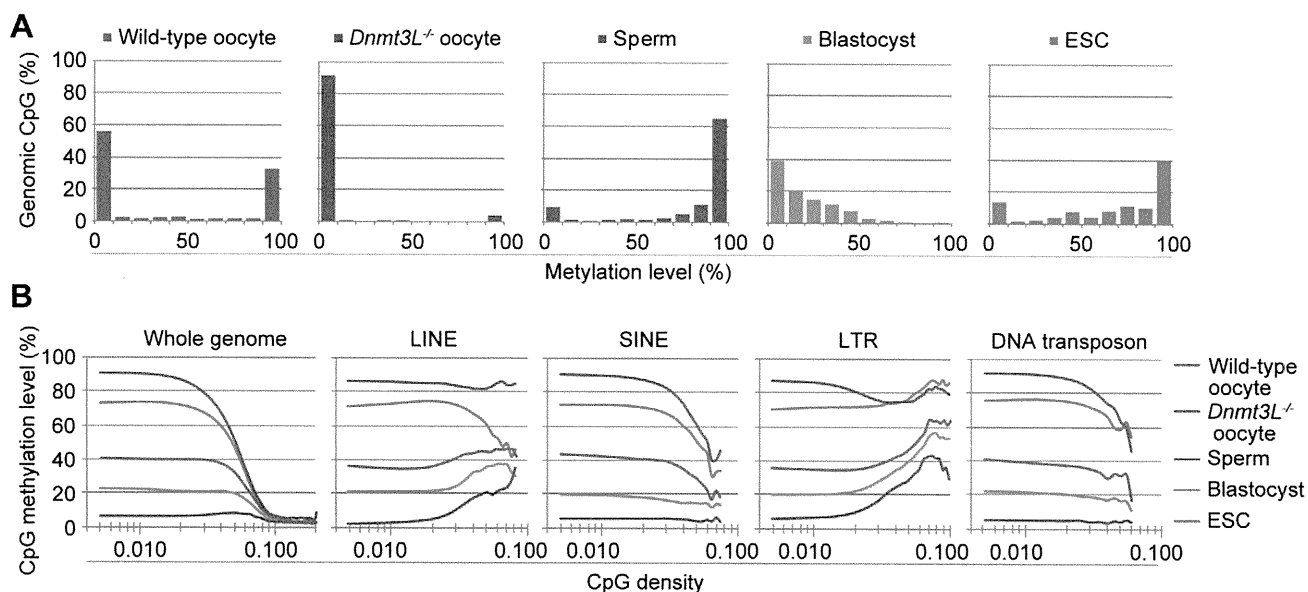


Figure 2. Genome-wide methylation profiling of mouse germ cells. (A) Histograms of methylation levels of genomic CpGs in wild-type oocyte, *Dnmt3L*^{-/-} oocyte, sperm, blastocyst, and embryonic stem cell (ESC) genomes. (B) CpG methylation levels are plotted as a function of CpG density for the whole genome and 4 families of transposable elements (long interspersed nuclear element (LINE), short interspersed nuclear element (SINE), long terminal repeat (LTR), and DNA transposon). doi:10.1371/journal.pgen.1002440.g002

gDMR was partially methylated (35.7%), whereas all other maternal ICRs were hypomethylated in *Dnmt3L*^{-/-} oocytes (Table 2). This residual methylation might result in the stochastic acquisition of the maternal imprint in the progeny of *Dnmt3L*^{-/-} females [34]. These results strongly suggested that the methylation level of individual CpGs can be determined from DNA methylome maps with a high degree of accuracy.

The study of mammalian DNA methylation patterns has previously suggested that methylation predominantly occurs at CpG sites; however, more recent studies, based on SBS methods, have indicated that methylation at non-CpG sites also occurs in human ESCs [22,23]. Detection of non-CpG methylation is one of the applications of the bisulfite-based methylation analysis but is problematic due to the incomplete conversion of cytosine, and overestimates of such cytosine by PCR amplification, which cannot be discriminated from true methylation. In order to evaluate the methylation status of non-CpG sites and avoid these problems, additional SBS analysis of mouse GV oocytes, sperm, blastocysts, and ESCs was performed by a non-amplification technique, termed Post-Bisulfite Adapter Tagging (PBAT) [Miura F. & Ito T, personal communication]. All C (originally methylated cytosine) and T (originally unmethylated cytosine) that mapped to genomic CpG and CpH sites (H = A, T, or C) were counted. The PBAT results showed CpG methylation ratios (C ratios = 0.395, 0.748, 0.137, 0.615 in oocytes, sperm, blastocysts, and ESCs) which are similar to the average methylation levels of individual DNA methylome maps obtained by MethylC-seq and WBA-seq among all examined cells. Interestingly, a relatively high fold enrichment of non-CpG methylation was observed in GV oocytes (C ratio = 0.034–0.038), but not in the other cell types, including mouse ESCs (C ratio < 0.01) (Figure S11).

Relationship between the DNA methylome and transcriptome of mouse germ cells

To elucidate the interaction between intragenic DNA methylation and gene transcription, the correlation between promoter

and gene-body methylation and expression levels for 20,854 different genes was examined. The mRNA-seq profiles for germ cells and ESCs are shown in Table S1. The results showed that mRNA transcript levels in oocytes were strongly correlated to gene-body methylation levels (Spearman's $\rho > 0.5$, $p < 1 \times 10^{-9}$) but were not significantly correlated to promoter methylation levels ($|\rho| < 0.1$) (Figure 4A). For example, the regions +2 to +5 kb from the transcription start site (TSS) and 0 to -5 kb from the transcription termination site (TTS) were hypermethylated (60–90% methylation) for the top 20% of expressed genes but were hypomethylated (10–30% methylation) for the bottom 20% of expressed genes. However, areas near the TSS (± 500 base pairs (bp)) were hypomethylated (10–20% methylation) in all genes, regardless of their expression level. In contrast, in the *Dnmt3L*^{-/-} oocyte genome, the correlation between gene expression and gene-body methylation was very weak ($|\rho| < 0.1$) (Figure 4B). In the sperm genome, promoter methylation was negatively correlated (Spearman's $\rho = -0.36$, $p < 1 \times 10^{-9}$) with gene expression, whereas gene-body methylation was positively correlated (Spearman's $\rho = 0.14$ – 0.16 , $p < 1 \times 10^{-9}$) to gene expression; the latter correlation was weaker than that observed in the oocyte genome (Figure 4C).

Role of *Dnmt3L* in the DNA methylome/transcriptome relationship

Further investigation of gene expression patterns in oocyte genomes revealed that the mRNA transcript levels between wild-type and *Dnmt3L*^{-/-} oocytes were very highly correlated ($R^2 = 0.9611$) (Figure 5A). In fact, there were no significant differences in the expression levels of representative oocyte-specific genes (e.g., *Gdf9*, *Bmp15*, *Bcl2l10*, *Zp1*, *Zp2*, *Zp3*, *Zar1*, *Npm2*, *Nlrp5*, and *Dppa3*, which are responsible for ovarian follicle formation, reproduction, and early development [35]) and DNA methyltransferase genes (e.g., *Dnmt1*, a maintenance methyltransferase, and *Dnmt3a* and *Dnmt3b de novo* methyltransferases); the expected difference in the expression level of *Dnmt3L* between wild-type and

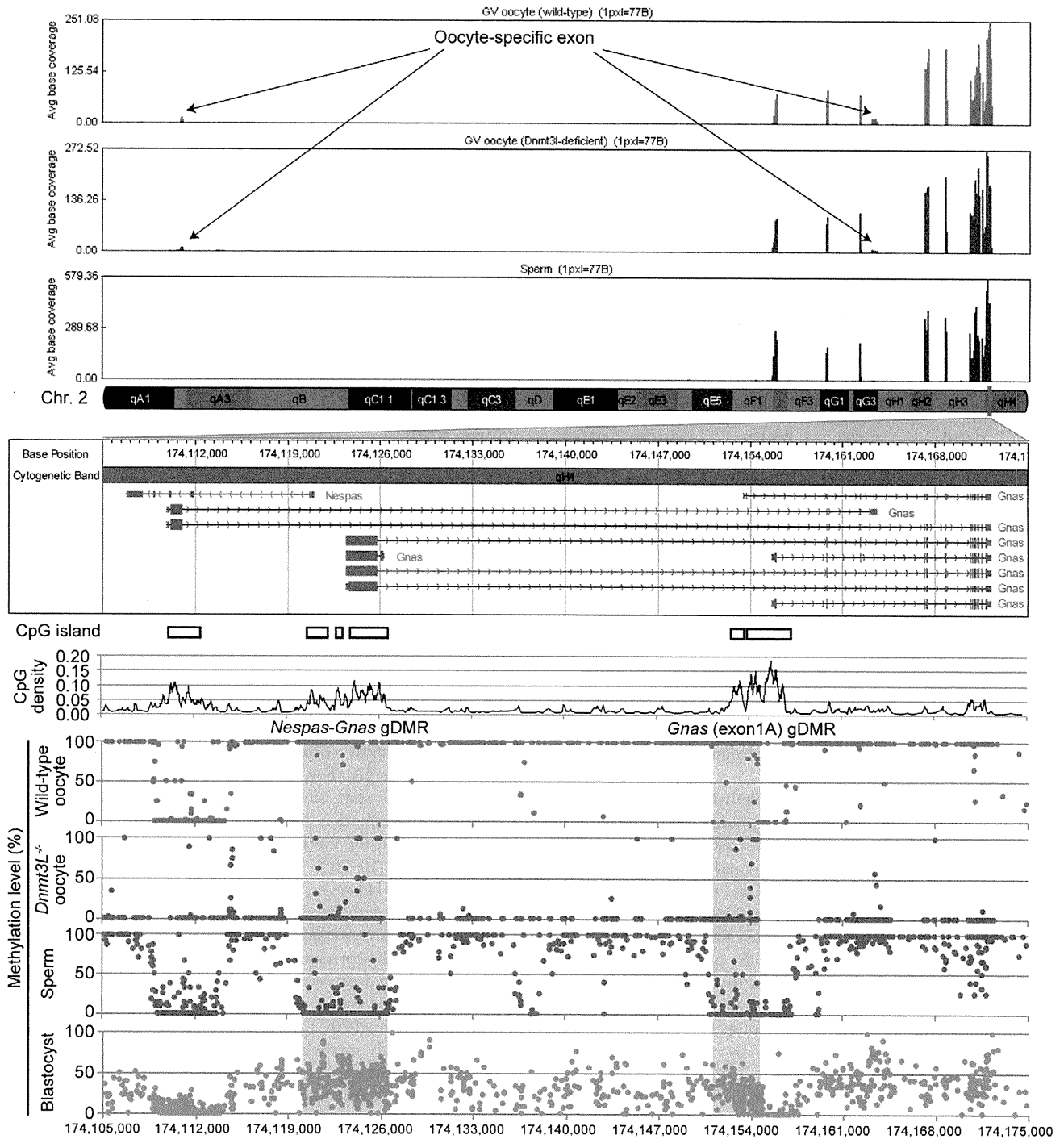


Figure 3. High-resolution genome-wide mRNA expression and CpG methylation profiling. GenomeStudio view of mRNA-seq data and CpG methylation map of the genomic region spanning the *Nespas-Gnas* maternally imprinted locus. (Top) Genomic stacked alignment plots of wild-type oocytes, *Dnmt3L*^{-/-} oocytes, and sperm. (Middle) Open boxes and black line plots represent the location of CGIs and the distribution of CpG densities of individual CpGs, respectively. (Bottom) Red, purple, blue, and green dots represent the methylation levels at individual CpGs in wild-type oocyte, *Dnmt3L*^{-/-} oocyte, sperm, and blastocyst genomes, respectively. The red shaded areas show the extent of two maternal imprinting control regions (ICRs). doi:10.1371/journal.pgen.1002440.g003

Dnmt3L^{-/-} oocytes was observed (Figure 5B, 5C). These results suggested that changes in gene expression did not occur during oogenesis, despite global intragenic hypomethylation in *Dnmt3L*^{-/-} oocytes. Furthermore, the expression levels and exon patterns of maternally-methylated imprinted genes across each ICR were

not altered in *Dnmt3L*^{-/-} oocytes (Figure 3 and Figure 5D). This result suggested that the disruption of maternal methylation imprints in the *Dnmt3L*^{-/-} oocyte genome was not due to the lack of their transcription [36]. On the other hand, maternal methylation imprints at ICRs (and many other hypermethyla-

Table 2. CpG methylation profiling of 12 maternal and 3 paternal imprinting control regions.

	Gene locus	Chr.	Extents of the ICRs†		Average methylation levels				
			Start	End	Wild-type oocyte	<i>Dnmt1</i> ^{-/-} oocyte	Sperm	Blastocyst	ESC
Maternally methylate imprinted genes	<i>Nespas-Gnas</i>	2	174,119,863	174,126,564	99.3%	5.6%	3.9%	38.2%	55.9%
	<i>Gnas</i> (exon1A)	2	174,150,877	174,154,638	95.2%	3.5%	4.1%	20.4%	7.8%
	<i>Peg10</i>	6	4,696,743	4,699,483	95.9%	6.7%	5.5%	31.8%	57.1%
	<i>Mest</i>	6	30,684,932	30,689,966	96.5%	2.3%	4.2%	30.7%	52.6%
	<i>Peg3</i>	7	6,679,787	6,684,257	98.1%	3.0%	2.5%	32.1%	42.8%
	<i>Snrpn</i>	7	67,147,381	67,151,583	94.1%	35.7%	4.6%	34.3%	64.9%
	<i>Kcnq1ot1</i>	7	150,480,736	150,482,810	97.9%	2.2%	4.3%	34.1%	52.0%
	<i>Plagl1</i>	10	12,809,697	12,812,131	99.9%	1.3%	7.4%	35.4%	53.0%
	<i>Grb10</i>	11	11,925,127	11,927,100	98.0%	1.2%	5.3%	38.5%	78.7%
	<i>Zrsr1</i>	11	22,871,610	22,874,212	94.1%	5.2%	6.8%	34.8%	47.0%
	<i>Igf2r</i>	17	12,934,169	12,935,816	99.1%	0.9%	3.8%	44.2%	53.2%
	<i>Impact</i>	18	13,130,435	13,133,510	97.2%	2.4%	6.6%	43.1%	38.6%
Paternally methylated imprinted genes	<i>H19</i>	7	149,764,673	149,771,930	13.5%	0.6%	96.5%	40.8%	65.5%
	<i>Rasgrf1</i>	9	89,767,090	89,775,128	7.4%	0.7%	92.0%	25.2%	59.4%
	<i>Dlk1-Meg3</i>	12	110,762,703	110,773,093	18.9%	0.9%	96.8%	32.4%	83.1%

†: The extents of each region in germ cells were determined by bisulfite sequencing study [39].
doi:10.1371/journal.pgen.1002440.t002

tions at transcribed regions) in wild-type oocyte genomes might be the result of gene transcription via *Dnmt3L*-mediated intragenic methylation.

Surprisingly, gene expression in ESC genomes was negatively correlated with promoter methylation and was not positively correlated with gene-body methylation (Figure S12). Meanwhile, these ESCs showed the apparent expression of all DNA methyltransferase gene families including *Dnmt3L* (Figure S13). Previous studies indicated that the zygotic and somatic functioning of *Dnmt3L* is not essential for global methylation in ESCs in mice [6]. Thus, unlike oocytes, the functional role of *Dnmt3L* in gene-body methylation after fertilization is unclear. However, the expression of pluripotency-associated genes, *Pou5f1*, *Klf4*, *Sox2*, *Myc*, *Nanog*, and *Lin28a*, was clearly observed in ESCs. The expression of *Pou5f1*, *Lin28a*, and *Glis1*, recently identified as maternal reprogramming factors, were also observed in oocytes (Figure S14). While differential expression of the pluripotency genes among germ and stem cells was observed, the promoter regions of these genes demonstrated low-level methylation in almost all of the examined cells. In sperm cells, only the *Nanog* promoter was hypermethylated (this result was similar to a previous study [29]).

Identification and characterization of germline differentially methylated regions

To identify gDMRs, the average CpG methylation levels of individual CpG islands (CGIs), which are CpG-rich genomic regions often lacking DNA methylation, were calculated. Recently, Illingworth et al. determined the number of CGIs by deep sequencing of isolated, unmethylated DNA clusters [37]. Among the 23,021 mouse CGIs (22,974 CGIs were informative in both oocytes and sperm), 2014 were highly methylated ($\geq 80\%$ methylation) in oocytes, 818 were highly methylated in sperm, and 377 were highly methylated in both germ cells (Figure 6A). Furthermore, we also identified 1678 gDMRs ($\geq 80\%$ methylation

in 1 gamete and $\leq 20\%$ in the other), 1329 of which were oocyte-specific methylated CGIs, while the remaining 349 were sperm-specific methylated CGIs (Figure 6A, Figure S6, and Table S2). Among these gDMRs, 646 gDMRs were confirmed to show a differential methylation status between GV oocytes and sperm (by similar criteria: $\geq 75\%$ methylation in 1 gamete and $\leq 25\%$ in the other); the methylation status was previously examined by performing large-scale bisulfite sequencing of CpG-rich regions of the genome (reduced representation bisulfite sequencing: RRBS) (Table S3) [38]. Additionally, almost all known ICRs except *Zdby2* DMRs (which do not have any CGIs) were re-identified from our gDMR list (Table S2).

A total of 78% oocyte-methylated gDMRs ($n = 1045$) were located within the intragenic regions. Approximately 25% of the oocyte-methylated gDMRs ($n = 322$) overlap with either the first exon or the proximal promoter regions of the genes, as has been observed with most of the described maternal ICRs [39]; only 5% of the sperm-methylated gDMR ($n = 18$) showed such overlap. Alternatively, 34% of sperm-methylated gDMRs ($n = 120$) overlap with intergenic regions, as in all known paternal ICRs (Figure 6B). Interestingly, oocyte-methylated gDMRs in transcribed regions tended to be more abundant within highly expressed genes, but such a trend was not observed in the sperm genome (Figure 6C). Oocyte-methylated gDMRs were also identified in non-imprinted genes, such as the DNA methyltransferase genes (e.g., *Dnmt1* and *Dnmt3b*) and some male germline-specific genes (e.g., *Piwil1*, *Spag1*, *Ggnbp2*, *Tbpl1*, *Spata16*, *Ggn*, *Acrbp*, and *Cd46*). The oocyte-methylated gDMR in *Dnmt1* was located in spermatocyte- and somatic-specific exons, while oocyte-specific exons were hypomethylated in oocytes (Figure S9). *Dnmt3L*^{-/-} oocytes also showed hypomethylation in most of these gDMRs. Significant changes in the expression levels of genes with alternative splicing patterns were not observed in the *Dnmt3L*^{-/-} oocyte genome (Figure 3, Figure 5E, and Figure S9). These results indicate that these oocyte-specific methylated gDMRs do not regulate gene expression or alternative splicing during the oocyte stage.

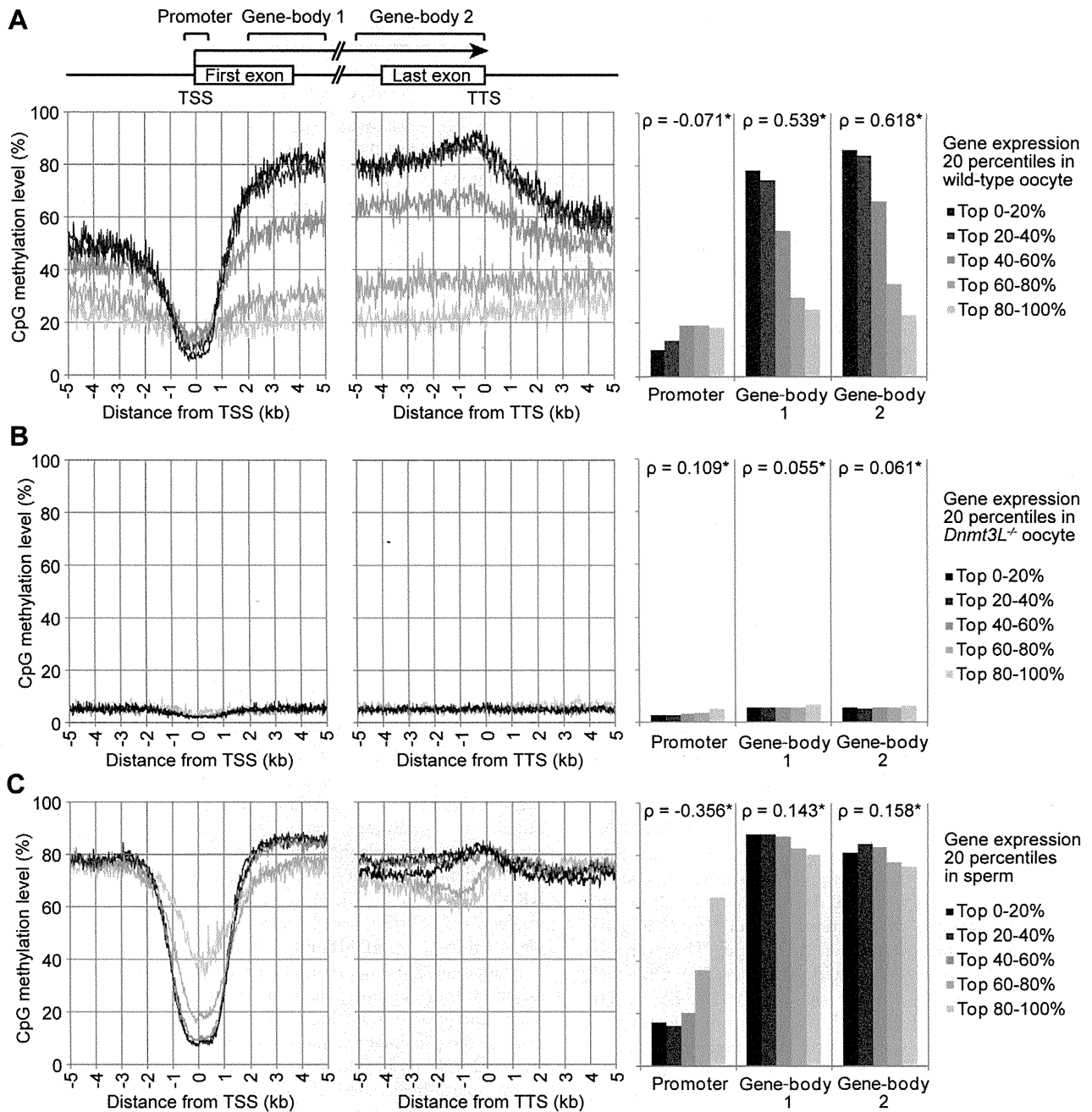


Figure 4. Relationship between gene expression and methylation in promoter and gene-body regions in mouse germ cells. The expression level of genes in wild-type oocytes (A), sperm (B), and *Dnmt3L*^{-/-} oocytes (C) were divided into 5 percentile groups. The distribution of methylation is shown ± 5 kb from the transcription termination site (TTS; left) and transcription start site (TSS; middle). The graphs on the right show the average methylation levels in the promoter and gene-body regions. Spearman's rank correlation coefficient (ρ) was used to test the statistical significance of the correlation between gene expression and DNA methylation levels (*: $p < 1 \times 10^{-9}$). doi:10.1371/journal.pgen.1002440.g004

To determine whether or not these germ cell-specific methylations are maintained after fertilization, when the genomes undergo global demethylation, the individual CGI methylation levels in blastocyst genomes were calculated. In blastocysts, all ICRs demonstrated low to moderate methylation (25.1–64.3%), whereas many gDMRs were demethylated (0–20%) (Figure 6D). Furthermore, 817 oocyte-methylated gDMRs (including *Piwil1*, despite being a non-imprinted gene locus) and 34 sperm-specific gDMRs were resistant to demethylation during early embryogenesis ($\geq 20\%$ methylation in blastocysts) (Figure 6D and Table S2).

Among the demethylation-resistant gDMRs, a novel gDMR in the intron of *Gpr1* (Figure S10) was found to be a tissue-specific, paternally-expressed imprinted gene [40]. Bisulfite sequencing analysis showed that this gDMR was hypomethylated in *Dnmt3L*^{-/-} oocytes and maternal allele-specific methylation was detected in this region in blastocysts (Figure 6E). Methylation profiles in ESCs showed that 26% ($n = 213$) of the demethylation-resistant gDMRs maintained or increased DNA methylation (Figure S15). Among ICRs, only *Gnas* exon1A ICR was demethylated (7.8%),

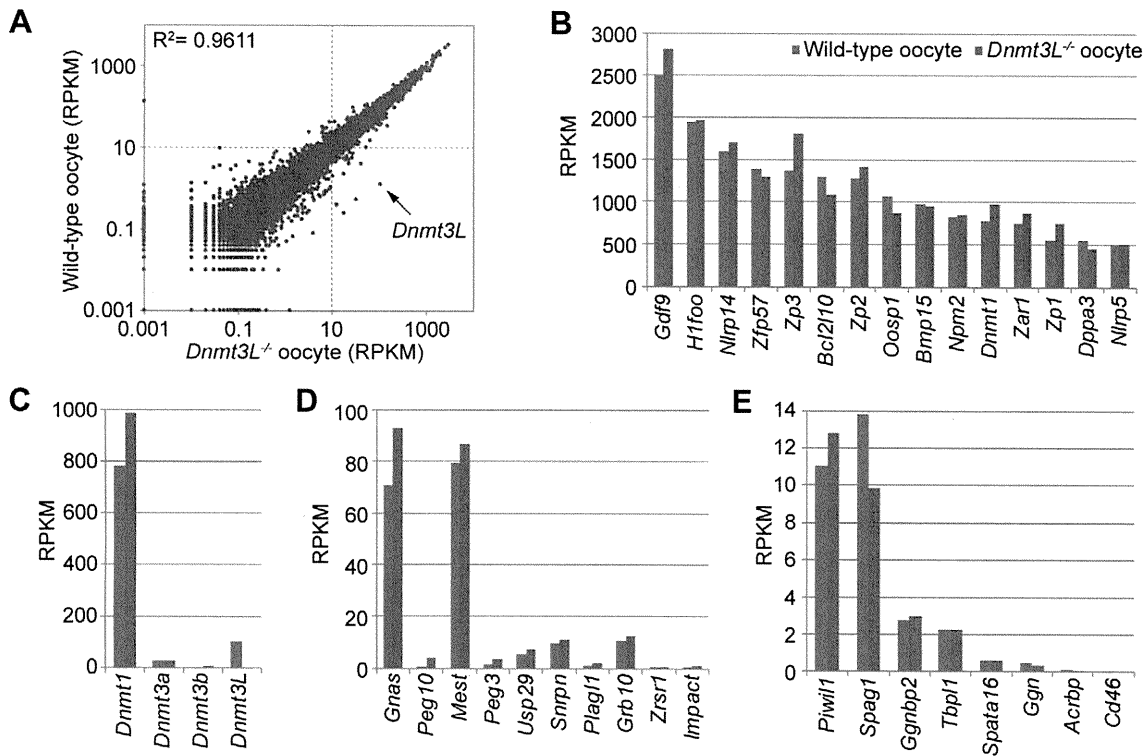


Figure 5. Comparison of gene expression profiles between wild-type and *Dnmt3L*^{-/-} oocytes. (A) Scatter plot and correlation coefficient (R^2) of RPKM values of 20,854 genes in wild-type and *Dnmt3L*^{-/-} oocytes. Expression levels of oocyte-specific genes (B), DNA methyltransferase genes (C), maternally-imprinted genes that are potentially necessary to establish methylation imprints (D), and male germline-specific genes that contain oocyte-specific methylated CpG islands (CGIs) (E). doi:10.1371/journal.pgen.1002440.g005

whereas the other ICRs developed partial or high methylation levels (range, 38.6–83.1%) in ESCs (Table 2). Among other demethylation-sensitive gDMRs, which were demethylated (<20% methylation) in blastocysts, many (76%, $n = 264$) sperm-methylated gDMRs were re-methylated ($\geq 20\%$ methylation); most (81%, $n = 416$) of the oocyte-methylated gDMRs maintained low methylation (0–20%) in ESCs (Figure S15). Finally, out of 704 demethylation-resistant (in blastocysts) oocyte-methylated gDMRs which were informative in *Dnmt3L*^{-/-} oocytes, only 4 remained hypermethylated (80–100% methylation) in the *Dnmt3L*^{-/-} oocyte genome. However, almost all other oocyte-specific methylation marks at gDMRs were *Dnmt3L*-dependent (Figure 6F). These results suggest that *Dnmt3L*-mediated methylation during oogenesis regulates the establishment of most heritable oocyte-specific marks, including genomic imprints.

Discussion

To the best of our knowledge, this is the first study to generate single-base resolution maps of DNA methylomes spanning the entire genome of mouse germ cells. The oocyte maps are particularly valuable and informative because, in the past, such an analysis was prohibitive due to the need for large quantities of DNA. Recently, Smallwood et al. [38] reported large-scale DNA methylation patterns in mouse germ cells by using the RRBS method, which targets only CpG-rich regions. However, our more comprehensive results provide strong evidence that gene expression was positively correlated to *Dnmt3L*-dependent intragenic methylation in oocytes, and that methylation patterns in oocytes differed from those in sperm and non-germline cells.

The functional role of gene-body methylation has been an enigma despite its conservation in plants and animals [41–43]. Maunakea et al. [44] suggested that gene-body methylation is involved in the regulation of alternative splicing events. Although methylated gDMRs were detected in the alternative exons of *Dnmt1* and *Gnas* in mouse oocytes, loss of oocyte-specific methylation marks in the *Dnmt3L*^{-/-} oocytes did not affect the expression patterns of alternatively spliced transcripts. Therefore, our results indicate that gene-body methylation is not involved in alternative splicing in oocytes.

Previously, Chotalia et al. [36] showed that transcription during the oocyte stage is required for the establishment of maternal methylation marks on an imprinted gene. The present results show that *Dnmt3L*^{-/-} oocytes lost almost all of their maternal methylation imprints while maintaining a constant amount of mRNA through each ICR despite the global loss of intragenic methylation. Thus, these results strongly suggest that the establishment of genomic imprints via transcription is mediated by *Dnmt3L*-dependent intragenic methylation.

A possible mechanism for gene-body methylation involves the exposure of intragenic regions to DNA methyltransferases, considering that RNA polymerase disrupts the chromatin structure during transcription. However, not all transcripts across gDMRs corresponded to highly expressed genes in oocytes (Figure 6C). Therefore, other epigenetic marks with an open chromatin structure might also be important for DNA methylation in oocytes. For instance, a recent knockout study showed that *Kdm1b*, which encodes histone H3K4 demethylase, is required for the establishment of some maternal methylation imprints [45]. Thus, several factors, including transcriptional and epigenetic modifica-

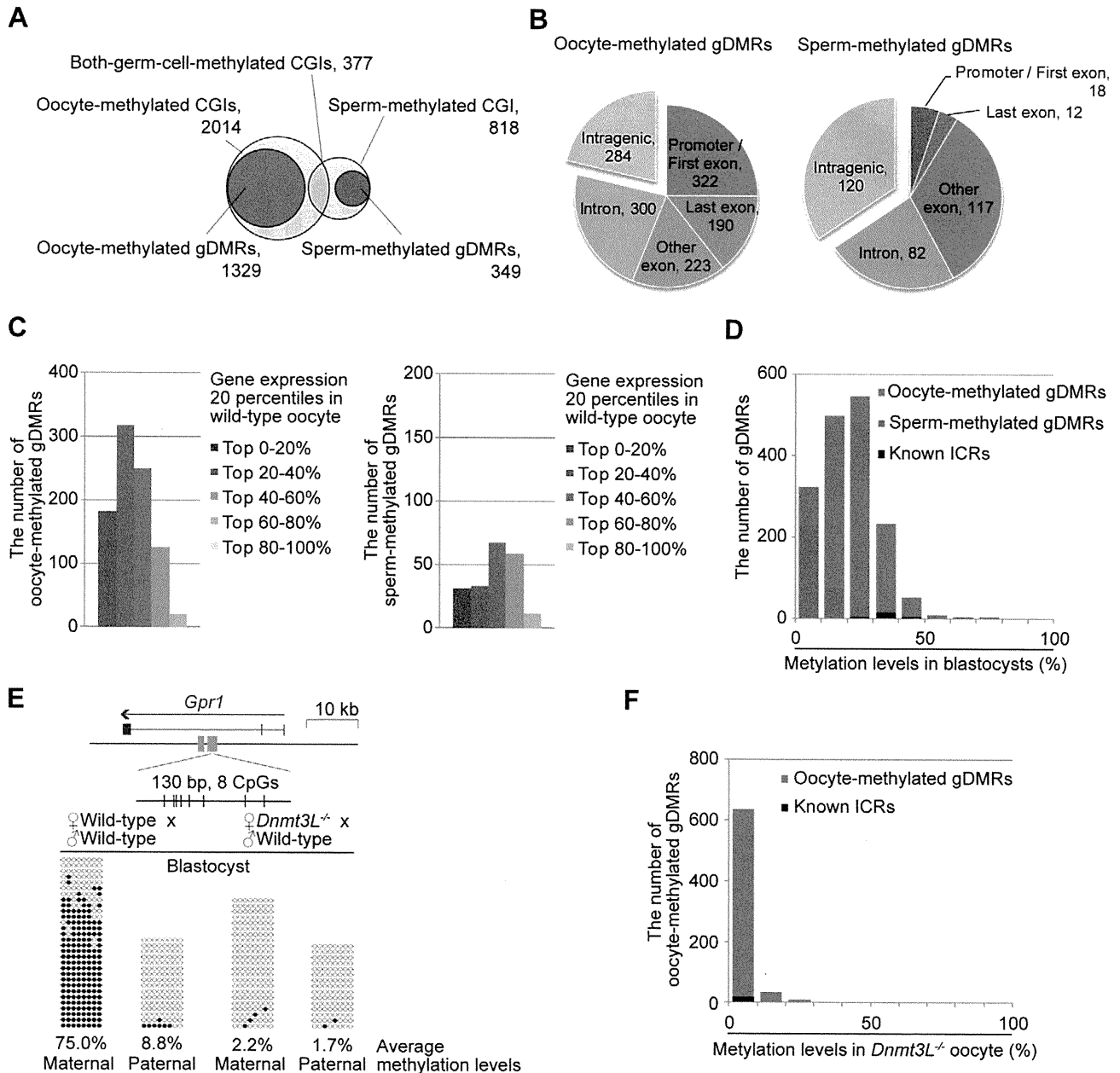


Figure 6. Identification of germline differentially methylated CGIs from DNA methylome profiles. (A) Venn-like diagram of two groups of CGIs, namely, oocyte-methylated CGIs (light pink) and sperm-methylated CGIs (light blue) and two groups of gDMRs, namely, oocyte-methylated gDMRs (red) and sperm-methylated gDMRs (blue). (B) The genomic distribution of 1329 oocyte-methylated (left) and 349 sperm-methylated gDMRs (right). The gDMRs were classified into 5 genomic locations; promoter (within 500-bp upstream from the first exon) or first exon, last exon, other exon, intron, and intergenic region. (C) The locations of the intragenic 1045 oocyte-methylated (left) and 229 sperm-methylated gDMRs (right). The gDMRs were classified into 5 gene group locations; the genes were divided into 5 percentile groups according to their expression levels in wild-type oocytes and sperm, as shown in Figure 3. (D) Histograms of the methylation levels of the gDMRs in blastocysts. The number of newly identified oocyte-specific, sperm-specific methylated gDMRs, and known ICRs are shown in black, red, and blue, respectively. (E) Bisulfite sequencing at the *Gpr1* gDMR in mouse blastocysts. (Top) Schematic representation of paternally-expressed *Gpr1*. The gene and gDMRs are shown in blue and green, respectively, and CpG sites are represented by vertical bars. (Bottom) Methylated and unmethylated CpGs are indicated by open and closed circles, respectively. The maternal and paternal alleles were distinguished by three polymorphisms between C57BL/6N and JF1 mice (G/A at 63,247,064; T/A at 63,247,072; and TA/AG at 63,247,089–63,247,090 on chromosome 1). (F) Histograms of the methylation levels of the demethylation-resistant oocyte-methylated gDMRs in *Dnmt3L*^{-/-} oocytes. The number of newly identified oocyte-specific, sperm-specific methylated gDMRs, and known ICRs are shown in black, red, and blue, respectively.
 doi:10.1371/journal.pgen.1002440.g006

tions, might be involved in *Dnmt3L*-mediated intragenic methylation.

The results of this study show that gene-body methylation was correlated to gene expression in sperm. However, the extent of

that correlation is much less than in oocytes due to genome-wide hypermethylation, including in low-CpG-density regions. In male germline cells, global methylation acquisition begins during late embryonic development and before birth [3]. To more clearly

show this correlation, analysis of early-stage germ cells in fetal or neonatal animals might be required. Surprisingly, a positive correlation between mRNA expression and gene-body methylation was not observed in mouse ESCs. In addition, the accumulation of non-CpG methylation was not observed in mouse ESCs. These results contradict the results of another study, which showed that active transcription was associated with intragenic DNA methylation with non-CpG methylation in human ESCs [22,23]. This discrepancy might reflect the differences between human and mouse ESCs, the precise cell derivations or culture conditions [46,47]. However, further comparative studies on germ cell epigenomes from other species are required to further elucidate the functional role of epigenetic marking systems.

In this study, a large number of heritable oocyte-specific methylation marks were identified within a set of novel CpG islands [37]. The difference in the number of oocyte- and sperm-specific gDMRs reflects the fact that only 3 or 4 paternally-methylated imprinted loci were observed, as compared to approximately 20 maternally-methylated imprinted loci. The reason for the relative abundance of oocyte-specific methylated CGIs might be related to the intragenic methylation of CpG-rich regions, which are hypomethylated in sperm. The results show that most of the oocyte-specific marks are *Dnmt3L*-dependent, similar to results recently obtained by RRBS-based analysis [38]. However, whether all of these CpG-rich regions serve as imprinting methylation marks is unclear. For instance, although many genes with oocyte-specific methylation marks were identified (Figure 6B), the evidence that these genes were imprinted was lacking (e.g., *Piwil1* and *Dnmt1*). These methylation marks might not be involved in the formation of a fertile oocyte but might play crucial roles in gene expression after fertilization. Furthermore, ESC methylomes showed that many gDMRs, especially sperm-specific gDMRs, acquired new methylation patterns after implantation. Methylation of these CGIs might control tissue-specific gene expression [48,49]. Partial alternation of imprinted methylation patterns in ESCs were observed in the present study, potentially caused by significant differences in the extent of the ICRs during embryo development [39]. A fuller understanding of epigenetic stability will require further methylome profiling during early embryogenesis and stem cell differentiation. The present study also identified a gDMR as a novel ICR candidate in the intron of the imprinted *Gpr1* gene. Thus, traditional promoter arrays may not identify all ICRs. However, further analyses are needed to determine which gDMRs, identified in the CpG methylome maps, are true ICRs at the imprinted *Gpr1- ζ dbf2* locus [40,50].

mRNA-seq results showed that the expression levels of most genes in the wild-type and *Dnmt3L*^{-/-} oocytes were similar. For instance, the expression level of almost all oocyte-specific genes, which regulate ovarian follicle formation, reproduction, and early development, were not significantly altered (Figure 5B and Table S1). These results are consistent with the findings of previous studies, which showed that *Dnmt3L*^{-/-} female mice were capable of producing fertile oocytes (however, their offspring were not viable due to the lack of imprinting) [5,6]. Thus, regulation of oocyte-specific genes must be beyond the control of *Dnmt3L*-dependent cytosine methylation.

Although *Dnmt3L*^{-/-} oocytes showed global hypomethylation at low to high CpG densities, some families of retrotransposons, such as LINEs and LTRs, were partially methylated at moderate to high CpG densities. Therefore, *Dnmt3L*-independent methylation might be involved in the silencing of retrotransposons and completion of oocyte meiosis. Previously, De La Fuente et al. [51]

showed that *Hells* (also known as *Lsh*), which encodes a member of the sucrose non-fermenter 2 (SNF2) family of chromatin remodeling proteins, is required for DNA methylation of IAP and pericentromeric satellite repeats as well as repression of IAP retrotransposition in pachytene oocytes. Unfortunately, measurement of the methylation levels of satellite DNA, which is abundant in the pericentromeric regions, was not possible because these sequences were excluded from our analysis. However, a previous sequencing study showed that methylation levels of satellite DNA did not differ between the wild-type and *Dnmt3L*^{-/-} oocytes [52]. Combined, these results suggest the presence of 2 types of oocyte methylation patterns: (i) *Dnmt3L*-mediated intragenic methylation that is essential for early embryogenesis and (ii) *Dnmt3L*-independent retroviral and pericentromeric methylation, which may be mediated by *Hells* activity, is crucial for oocyte meiosis [51]. Further studies on *Hells*-mediated oocyte methylation are required to elucidate the details of this mechanism.

Previous studies on the cytosine methylation of mtDNA have been highly controversial. A recent study by Shock et al. [53] reported cytosine methylation and hydroxymethylation in mammalian mitochondria. Our results indicated that mtDNA is unmethylated in blastocysts and ESCs, but is partially methylated in germ cells. Whether or not 5-hydroxymethylcytosine (5-hmC) exists in mitochondrial or genomic chromosomes of germ cells remains unclear. Meanwhile, rapid hydroxylation of 5-methylcytosine (5-mC) in the paternal pronucleus during zygotic development was also recently reported [54,55]. Currently, it is difficult to assess hydroxymethylation profiles in oocyte genomes due to the limited DNA recovery. Further investigation of cytosine modification during germ cell and zygote development will be required in the future to better understand this process.

The DNA methylome maps of mouse germ cells, in this study, were derived from SBS data and, therefore, accurately represent methylation levels of individual CpGs on a whole-genome level. The adaptation of the SBS method for small-scale DNA analysis, described in the present report, has the potential to enable further analyses of germline lineages. The current work examined SBS library construction using 3 methods, MethylC-seq, WBA-seq, and PBAT. MethylC-seq basically required only micrograms of DNA [22,23,56], thus over amplification might cause redundancy in oocyte libraries. The latter methods allow comprehensive methylome analysis in samples with low amounts of starting DNA by avoiding DNA damage due to sodium bisulfite treatment (after adapter ligation, in the case of MethylC-Seq). Recent studies using BS sequencing have shown that methylated cytosine is abundant in the non-CpG regions of human pluripotent stem cells and mouse oocytes [22,23,39,56]; however, the function of non-CpG methylation in mammalian genomes remains unclear. The PBAT results also showed an abundance of non-CpG methylation in oocytes, with results similar to a previous sequencing study on imprinted loci [39]. However, accurate assessment of non-CpG methylation is required using increased sequencing depths because methylation levels of the non-CpG sites were much lower than those of the CpG sites. SBS library construction was conducted by WBA-seq from 2000 fully matured (metaphase II stage) oocytes; sufficient quantities for sequencing were not obtained. During oogenesis, most of the oocyte specific imprinted methylation marks were established during the GV stage. This contrasted to a previous study where a continuous increase in methylation levels was observed [38]. Further improvement of SBS methods, requiring smaller amounts of DNA, is needed to provide complete germ cell methylome maps and to elucidate the exact function of non-CpG methylation in germ cells.

In conclusion, we constructed the first extensive, high-resolution maps of DNA methylomes of mouse oocytes and sperm. These maps described the epigenetic properties of these DNA methylomes. Our data could serve as a platform for future studies to elucidate the role of epigenetic modifications in the development and functioning of germ and stem cells. Such studies are anticipated to improve our understanding of epigenetic reprogramming.

Materials and Methods

Preparation of MethylC-seq libraries

Five thousand germinal vesicle (GV)-stage oocytes were collected from the ovarian follicles of adult (7- to 9-week-old) female C57BL/6N mice (Clea Japan, Tokyo, Japan) 44–48 h after they were injected with equine chorionic gonadotropin. Three hundred blastocysts at embryonic day 3.5 were obtained from superovulated adult female C57BL/6N mice by flushing the uterus. Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Sperm were released from the cauda epididymides of adult male C57BL/6N mice. Sperm DNA was isolated by a standard phenol-chloroform extraction procedure with dithiothreitol (DTT). Genomic DNA from 2 lines of ESCs derived from C57BL/6J mice (Clea Japan) was extracted using the DNeasy Blood & Tissue Kit (Qiagen). DNA samples were sheared into 100-bp fragments in oocytes and 200-bp fragments in other samples using the Covaris S2 focused acoustic system (Covaris, Woburn, MA). Cytosine-methylated adapters (Illumina, San Diego, CA) were ligated to DNA by using the Paired-End DNA Sample Prep Kit or ChIP-Seq DNA Sample Prep Kit (Illumina). DNA fragments were isolated by 2–3% agarose gel electrophoresis and purified using the QIAquick Gel Extraction Kit (Qiagen). Sodium bisulfite conversion was performed using the Epitect Bisulfite Kit (Qiagen).

All bisulfite-converted DNA molecules were polymerase chain reaction (PCR)-amplified as follows: 2.5 U of Hot Start Taq polymerase (TaKaRa, Tokyo, Japan), 5 μ L 10 \times PCR buffer, 25 μ M dNTPs, 1 μ L of each PCR Primer PE 1.0 and 2.0 (Illumina) (50 μ L final). Thermocycling parameters were: initial denaturation at 94°C for 1 min, 15–25 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 30 s, and extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min. PCR reaction products were purified using the QIAquick kit (Qiagen).

Preparation of whole WBA-seq libraries

Two thousand GV-stage oocytes were collected from 7- to 9-week-old female C57BL/6N mice (Clea Japan) and, 2300 GV-stage oocytes were collected from 7–15-week-old *Dnmt3L*^{-/-} female mice (129SvJae \times C57BL/6N hybrid genetic background) [6,57]. Genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen), and then bisulfite-treated with Epitect Bisulfite Kit (Qiagen). Subsequently, the bisulfite-converted DNA was amplified using Epitect Whole Bisulfite Kit (Qiagen). The collected DNA was sheared into 200-bp fragments using Covaris S2. Unmodified Paired-End adapters (Illumina) were ligated to the DNA by using the Paired-End DNA Sample Prep Kit (Illumina). DNA fragments were isolated by 2% agarose gel electrophoresis and purified using the QIAquick Kit (Qiagen). All DNA was PCR amplified and purified in the same manner as the MethylC-seq method, except the number of PCR cycles was reduced to 7.

Preparation of PBAT libraries

GV-stage oocytes (400) and blastocysts (100) were obtained from 7- to 9-week-old female C57BL/6N mice (Clea Japan), and

genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen). The isolated oocyte and blastocyst genomic DNA and 100 ng of genomic DNA from sperm, blastocysts, and ESCs containing 1:200 amount of unmethylated lambda DNA (Invitrogen, Carlsbad, CA) were bisulfite-treated using the MethylCode Bisulfite Conversion Kit (Invitrogen). Details of the PBAT method are unpublished [Miura F & Ito T, personal communication]. Briefly, bisulfite-treated DNA were double-stranded using Klenow Fragments (3'-5' exo-) (New England Biolabs, Ipswich, MA) with random primers containing 5' biotin tags and Illumina PE adaptors. The biotinylated molecules (first strand) were captured using Dynabeads M280 Streptavidin (Invitrogen) and double-stranded using Klenow Fragments (3'-5' exo-) with random primers containing Illumina PE adaptors (second strand). Finally, template DNA strands were synthesized as complementary DNA with a second strand (unmethylated C is converted to T) using Phusion Hot Start High-Fidelity DNA Polymerase (New England Biolabs) with PCR Primer PE 1.0 (Illumina).

Preparation of mRNA sequencing libraries

Total RNA from 1000 wild-type GV oocytes, 500 *Dnmt3L*^{-/-} GV oocytes, sperm, and ESCs was extracted using the RNeasy Mini Kit (Qiagen) and treated with DNase I (Promega, Madison, WI). RNA-Seq libraries were constructed using the mRNA-Seq Sample Preparation Kit (Illumina).

Sequencing

The MethylC-seq for blastocysts, WBA-seq, and PBAT libraries were sequenced on a HiSeq 2000 sequencing system (Illumina); the other MethylC-seq and mRNA-seq libraries were sequenced on a Genome Analyzer II (Illumina). Sample preparation, cluster generation, and sequencing were performed using the Paired-End Cluster Generation Kit-HS and the TruSeq SBS Kit-HS for the HiSeq 2000. Similarly, the Paired-End Cluster Generation Kits v2 and v4 and 18- and 36-Cycle Sequencing Kits v3 and v4 were used for the Genome Analyzer II. All kits were from Illumina.

Gene mapping

All sequenced reads were processed using the standard Illumina base-calling pipeline (v1.4–1.7). Generated sequence tags were mapped onto the mouse genome (mm9, UCSC Genome Browser, July 2007, Build 37.1) by using the Illumina ELAND program.

MethylC-seq tags (36 or 76 nt) were mapped with a custom Perl program, as described previously [17,22]. Briefly, all cytosines in the tags were replaced by thymines. Next, these tags were aligned to 2 mouse genome reference sequences (mm9), such that the antisense strand had cytosines replaced by thymines and the sense strand had guanines replaced by adenines. Finally, all tags (32–76 nt) that mapped uniquely without any mismatches to both strands were compiled and used for further analyses.

The 76 nt WBA-seq tags were mapped as follows. All tags were converted to 2 types of reads; in 1 read (“For” read), cytosines were replaced by thymines and in the other read (“Rev” read), guanines were replaced by adenines. Both “For” and “Rev” reads were aligned to sense and antisense mm9 strands. A total of 793, 397, 948, 480, and 238 million tags were aligned in wild-type oocytes, *Dnmt3L*^{-/-} oocytes, sperm, blastocysts, and ESC genomes, respectively. To avoid bias, tags mapped with multiple hits or matched chromosome M (mitochondria), chromosome Y, or 3 types of repetitive sequences (simple repeat, low complexity repeat, and satellite DNA sequences) were omitted from further analyses.

The 47 nt PBAT tags (trimmed first 4 nt and last 1 nt) were mapped as follows. All guanidines in the tags were replaced by

adenines, and these tags were aligned to sense and antisense strands mm9.

For gene-level analysis, the concentrations of the perfectly matching 35 nt (trimmed first nt) mRNA-seq tags from wild-type oocytes, *Dnmt3L*^{-/-} oocytes, sperm, and ESCs were calculated for the genomic regions corresponding to those covered by the RefSeq transcript models. The expression level of 20,854 unique genes was ranked by expression levels (calculated as RPKM values) in each library (Table S1). A total of 33, 28, 23, and 25 tags were aligned in 4 mRNA-seq libraries, respectively. mRNA-seq data analysis was performed and visualized using GenomeStudio Data Analysis software (Illumina).

Methylation analysis

The percentage of individual cytosines methylated at all CpG sites covered by at least 1 read was calculated as 100×(number of aligned cytosines (methylated cytosines))/(total number of aligned cytosines and thymines (originally unmethylated cytosines)). All genomic CpG methylation data are available on our website (http://www.nodai-genome.org/mouse_en.html). The CpG and non-CpG (CpH) methylation levels determined by PBAT results were calculated as the ratio between the total read C and the total read T mapped to genomic cytosines. Bisulfite conversion failure rates were calculated by read C:T ratios from lambda DNA mapping data. The failure rates were as follows: GV oocyte, 0.009; sperm, 0.008; blastocysts, 0.011; and ESCs, 0.006. Locations of transposable elements in the mouse genome (mm9) were obtained from the UCSC Genome Browser, and the average methylation levels of the whole genome and each transposable element were recalculated from the ratio of the aligned cytosines and thymines in each sequence. Lists of 23,021 CGIs were obtained from a previous report [37]. Around the TSS and TTS (±5 kb), genomic regions were divided into 20-bp bins. For each bin, the average methylation value was calculated for each gene. The expression level of 20,854 genes was divided into 5 percentile groups ranked by RPKM values, and the average methylation level for each group was mapped onto the gene structure model. These computational analyses were performed using a custom Perl program. Supercomputing resources were provided by the Human Genome Center, Institute of Medical Science, University of Tokyo.

Statistical analysis

Correlations between gene expression ranks and average methylation levels in the promoter (±500 bp from the TTS) or gene-body regions (gene-body 1: +2 to +5 kb from the TSS; gene-body 2: 0 to -5 kb from the TTS) were calculated using Spearman's rank correlation coefficient (ρ). An R-squared value (R^2) was calculated to evaluate the correlation of RPKM values between wild-type and *Dnmt3L*^{-/-} oocytes. Statistical analysis was performed using the R statistical package.

Bisulfite sequencing

To analyze the methylation of the three transposable elements (L1 LINE, B1/Alu SINE, and IAP LTR), 20 wild-type GV oocytes were obtained from adult female C57BL/6N mice. Bisulfite sequencing conditions and primer sets for the three transposable elements were described, previously [52]. To analyze the methylation of the *Gpr1* locus, 10 blastocysts were obtained from BJF1 (C57BL/6N×JF1) and *Dnmt3L*^{mat-/-} (*Dnmt3L*^{-/-}×JF1) mice [6,57]. Genomic DNA from blastocysts was isolated using the QIAamp DNA Mini Kit (Qiagen) and treated with sodium bisulfite with the Epiect Bisulfite Kit (Qiagen). The *Gpr1* gDMR sequence was amplified with 2 rounds of nested PCR. The

first-round PCR reaction contained 1 U of Hot Start Taq polymerase (TaKaRa), 1× PCR buffer, 200 μM dNTPs, 1 μM forward primer, and 1 μM reverse primer (20 μL final). Thermocycling parameters were as follows: initial denaturation at 94°C for 1 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min. Subsequently, 2 μL of the product was used as the input for the second-round PCR, which was performed in the same manner. Primer sets for the nested PCR were as follows: Gpr1-BSF1 (5'-GATTAGATTAGGTTAGTTTGGA-3') and Gpr1-BSR1 (5'-ACTAAAACACTAATCACCAAATA-3') for the first round; Gpr1-BSF2 (5'-AGATTAGGTTAGTTTGGAATT-3') and Gpr1-BSR2 (5'-AACACTAATCACCAAATAATTC-3') for the second round. The second-round PCR product was subcloned and sequenced, as described previously [50]. The percentage methylation was calculated as 100×(number of methylated CpG dinucleotides)/(total number of CpGs). At least 10 clones from each parental allele were sequenced. Sequence data were analyzed using the QUMA quantification tool for methylation analysis [58].

Accession number

The MethylC-seq, WBA-seq, PBAT, and mRNA-seq data in this study have been deposited in the DNA Data Bank of Japan (DDBJ) under accession number DRA000484.

Supporting Information

Figure S1 Schematic of the SBS library construction procedure. MethylC-Seq libraries were generated by ligation of methylated sequencing adapters to fragmented genomic DNA followed by gel purification, sodium bisulfite conversion, and PCR amplification (*left*). WBA-seq libraries were generated by ligation of unmodified sequencing adapters to bisulfite-modified (amplified using EpiTect Whole Bisulfite Kits) and fragmented genomic DNA followed by gel purification and PCR amplification (*middle*). PBAT libraries were generated by double-stranded DNA synthesis from bisulfite-treated (single-stranded) DNA with random primers containing sequencing adapters (*right*). (TIF)

Figure S2 The percent of the oocyte and sperm genomes covered by differing minimum numbers of MethylC-seq and WBA-seq reads. (TIF)

Figure S3 Sequencing bias towards mitochondrial and repetitive DNA sequences. (A) Average read depths for autosomal chromosomes and chromosome M (mitochondria) of mouse oocyte and sperm genomes. Occupancy of transposable elements in reads from SBS libraries before (B) and after (C) filtering the biased reads. (D) Genomic CpG coverage of SBS reads for each chromosome of mouse oocyte (orange: MethylC-seq, red: combined between MethylC-seq and WBA-seq) and sperm genomes (blue). (TIF)

Figure S4 Average CpG methylation levels in genomic chromosomal DNA and mitochondrial DNA. (TIF)

Figure S5 High-resolution DNA methylome map on mouse X inactivation center region in chromosome X (100,200,000–101,200,000). GenomeStudio view of Refseq's positions, repetitive element, CpG methylation map, CpG densities, CGI positions, and CGI methylation map were shown. Red, purple, blue, green,

and khaki dots and boxes represent the methylation levels at individual CpGs and CGIs in wild-type oocyte, *Dnmt3L*^{-/-} oocyte, sperm, blastocyst, and ESC genomes, respectively, as shown in Figure 1.

(TIF)

Figure S6 DNA methylome maps of each chromosome of mouse germ cells. The methylation levels of each chromosome in wild-type oocytes, *Dnmt3L*^{-/-} oocytes, and sperm in 10 kb windows (excluding mitochondrial chromosome, chromosome Y, and unplaced contigs). Red, purple, and blue lines represent the methylation levels in wild-type oocytes, *Dnmt3L*^{-/-} oocytes, and sperm, respectively. Red and blue boxes represent oocyte-methylated and sperm-methylated gDMRs, and red and blue pins indicate maternal and maternal ICRs, respectively.

(TIF)

Figure S7 Methylation profiling of transposable elements in mouse germ cells. (A) CpG methylation levels are plotted as a function of CpG densities for L1 LINE, B1/Alu SINE, and LTR/ERVK retrotransposons (approximately 10% of the latter are intracisternal A particle (IAP) LTRs). Data for high CpG densities including less than 100 genomic CpGs were not plotted. (B) Bisulfite sequencing of L1 LINE, B1/Alu SINE, and IAP LTR retrotransposons. Methylated and unmethylated CpGs are indicated by open and closed circles, respectively.

(TIF)

Figure S8 Transcriptome and DNA methylome profiling at *H19-Igf2*. GenomeStudio view of mRNA-seq data (*top*) and CpG methylation map (*bottom*) of the genomic region spanning each locus. The blue shaded areas show the extent of the paternally-methylated gDMR.

(TIF)

Figure S9 Transcriptome and DNA methylome profiling at *Dnmt1*. The red shaded areas show the extent of the maternally-methylated gDMR.

(TIF)

Figure S10 Transcriptome and DNA methylome profiling at *Gpr1-Zdbf2*. The blue and red shaded areas show the extent of the paternally- and maternally-methylated gDMRs, respectively.

(TIF)

Figure S11 Quantification of the ratio of methylated (total number of read C) versus unmethylated cytosines (total number of read T) by PBAT results. Bar charts represent cytosine methylation ratio (A) at CpG (*left*), CpHpG (*middle*), and CpHpH (*right*) contexts and bisulfite-conversion failure rate (B) calculated by C:T ratio from lambda DNA mapping data. Total number of mapped reads is shown on these charts (*Top*).

(TIF)

References

- Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3: 662–673.
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33 Suppl: 245–254.
- Lees-Murdock DJ, Walsh CP (2008) DNA methylation reprogramming in the germ line. *Epigenetics* 3: 5–13.
- Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 9: 129–140.
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science* 294: 2536–2539.
- Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129: 1983–1993.

Figure S12 Relationship between gene expression and intra-genic methylation in ESCs. (A) The expression level of genes in ESCs was divided into 5 percentile groups. The distribution of methylation is shown ± 5 kb from the transcription termination site (TTS; *left*) and transcription start site (TSS; *middle*). The graphs on the right show the average methylation levels in the promoter and gene-body regions. Spearman's rank correlation coefficient (ρ) was used to test the statistical significance of the correlation between gene expression and DNA methylation levels (*: $p < 1 \times 10^{-9}$).

(TIF)

Figure S13 Expression profiles of DNA methyltransferase gene families. Red, purple, blue, and khaki bars represent RPKM values of individual genes in wild-type oocytes, *Dnmt3L*^{-/-} oocytes, sperm, and ESCs.

(TIF)

Figure S14 Expression profiles of pluripotency-associated genes among wild-type oocytes, *Dnmt3L*^{-/-} oocytes, sperm, and ESCs.

(TIF)

Figure S15 Histograms of the methylation levels of the demethylation-resistant (*left*) and demethylation-sensitive gDMRs (*right*) in ESCs. The number of oocyte-specific and sperm-specific methylated gDMRs is shown in red and blue, respectively.

(TIF)

Table S1 Gene transcript profiling for germ cells, blastocysts, and embryonic stem cells by mRNA-seq.

(XLSX)

Table S2 DNA methylation profiles of 23,021 CGIs.

(XLSX)

Table S3 Average DNA methylation profiles of 646 gDMRs determined by SBS and RRBS methods.

(XLSX)

Acknowledgments

We thank Terumi Horiuchi, Etsuko Sekimori, and Satoshi Sano for their assistance with analysis of the Illumina data. We are grateful to Fumihito Miura and Takashi Ito for their technical advice with the PBAT method and Takahiro Arima, Hitoshi Hiura, and Tom Moore for their helpful discussions and generous support.

Author Contributions

Conceived and designed the experiments: H Kobayashi, Y Suzuki, T Kono. Performed the experiments: H Kobayashi, T Sakurai. Analyzed the data: H Kobayashi, M Imai, Y Suzuki. Contributed reagents/materials/analysis tools: H Kobayashi, T Sakurai, M Imai, N Takahashi, A Fukuda, O Yayoi, S Sato, K Nakabayashi, K Hata, Y Sotomaru, Y Suzuki. Wrote the paper: H Kobayashi, T Kono.