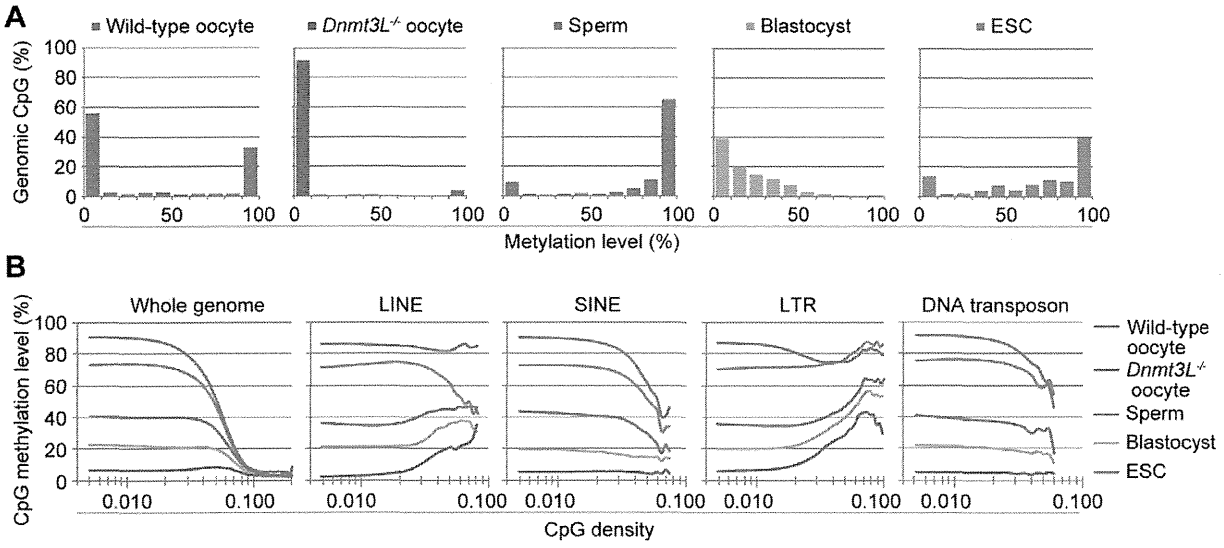


Figure 2. Genome-wide methylation profiling of mouse germ cells. (A) Histograms of methylation levels of genomic CpGs in wild-type oocyte, *Dnmt3L*$^{-/-}$ oocyte, sperm, blastocyst, and embryonic stem cell (ESC) genomes. (B) CpG methylation levels are plotted as a function of CpG density for the whole 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

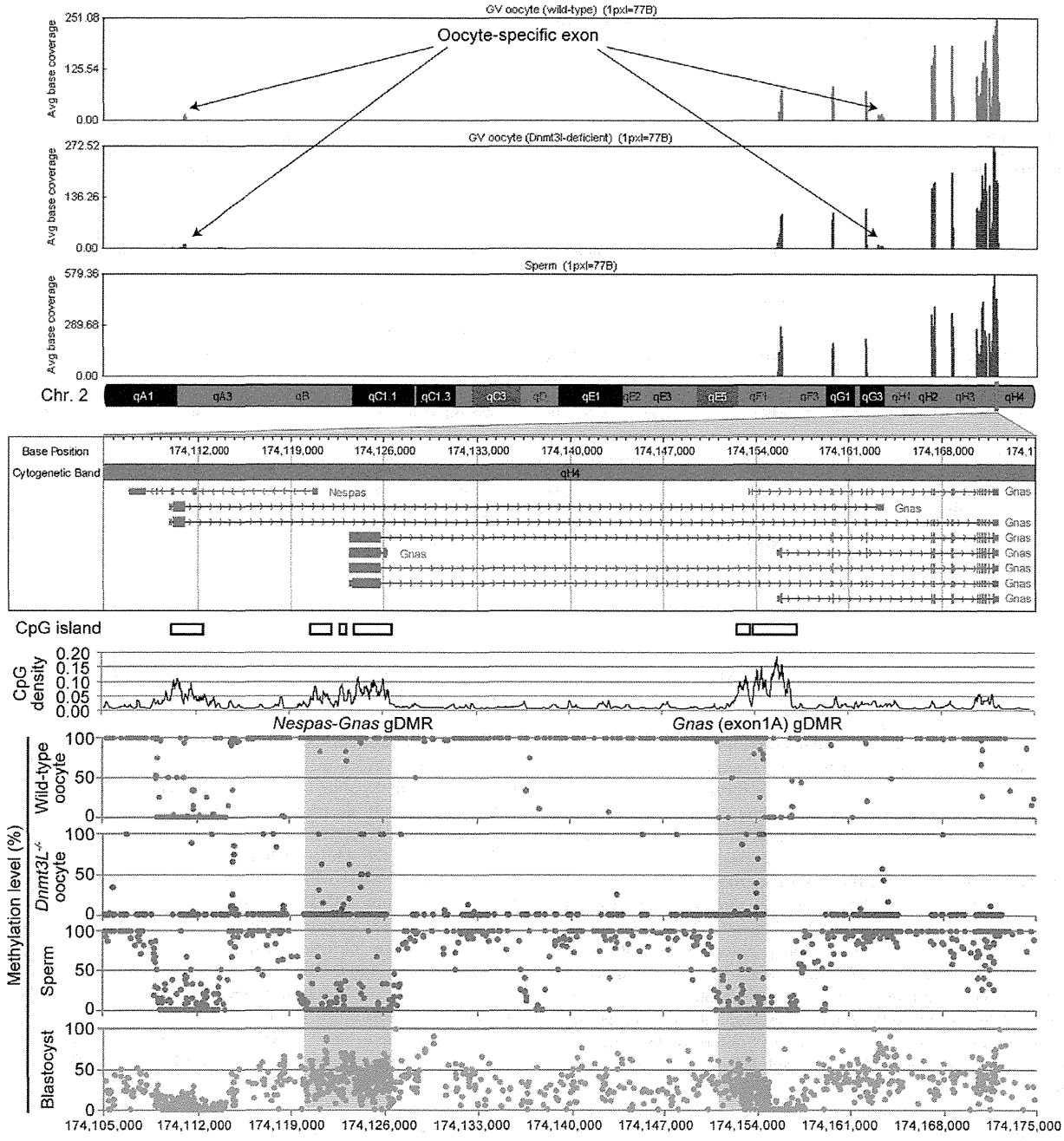


Figure 3. High-resolution genome-wide mRNA expression and CpG methylation profiling. GenomeStudio view of mRNA-seq data and CpG methylation map of the genomic region spanning the *Nespas-Gnas* maternally imprinted locus. (*Top*) Genomic stacked alignment plots of wild-type oocytes, *Dnmt3L*$^{-/-}$ oocytes, 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	*DnmtL*$^{-/-}$ oocyte	Sperm	Blastocyst	ESC
Maternally methylate imprinted genes	*Nespas-Gnas*	2	174,119,863	174,126,564	99.3%	5.6%	3.9%	38.2%	55.9%
	Gnas (exon1A)	2	174,150,877	174,154,638	95.2%	3.5%	4.1%	20.4%	7.8%
	Peg10	6	4,696,743	4,699,483	95.9%	6.7%	5.5%	31.8%	57.1%
	Mest	6	30,684,932	30,689,966	96.5%	2.3%	4.2%	30.7%	52.6%
	Peg3	7	6,679,787	6,684,257	98.1%	3.0%	2.5%	32.1%	42.8%
	Snrpn	7	67,147,381	67,151,583	94.1%	35.7%	4.6%	34.3%	64.9%
	Kcnq1ot1	7	150,480,736	150,482,810	97.9%	2.2%	4.3%	34.1%	52.0%
	Plagl1	10	12,809,697	12,812,131	99.9%	1.3%	7.4%	35.4%	53.0%
	Grb10	11	11,925,127	11,927,100	98.0%	1.2%	5.3%	38.5%	78.7%
	Zrsr1	11	22,871,610	22,874,212	94.1%	5.2%	6.8%	34.8%	47.0%
	Igf2r	17	12,934,169	12,935,816	99.1%	0.9%	3.8%	44.2%	53.2%
	Impact	18	13,130,435	13,133,510	97.2%	2.4%	6.6%	43.1%	38.6%
Paternally methylated imprinted genes	*H19*	7	149,764,673	149,771,930	13.5%	0.6%	96.5%	40.8%	65.5%
	Rasgrf1	9	89,767,090	89,775,128	7.4%	0.7%	92.0%	25.2%	59.4%
	Dlk1-Meg3	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 (≥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 (≥80% methylation in 1 gamete and ≤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: ≥75% methylation in 1 gamete and ≤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 *Zdbf2* 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.

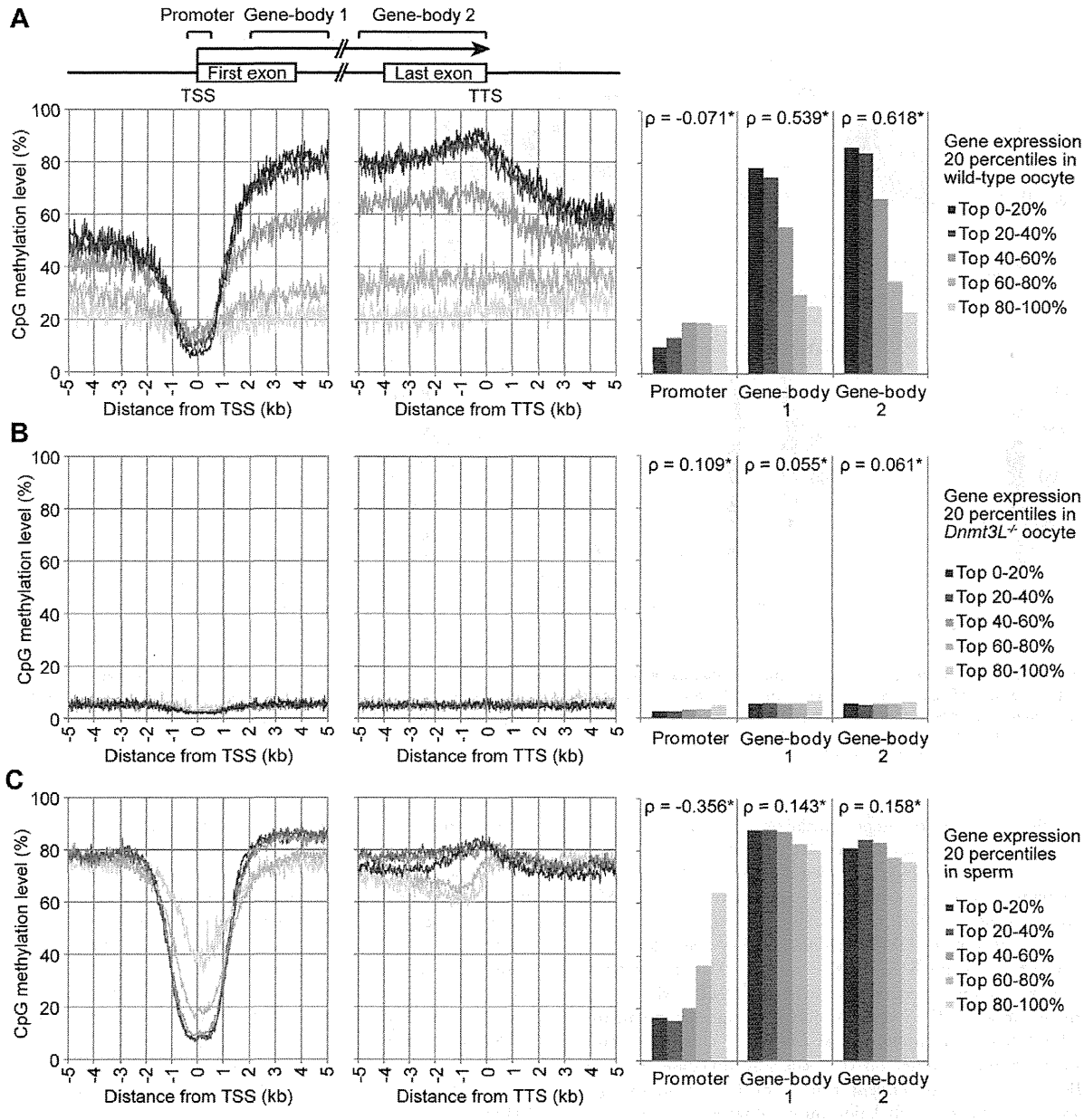


Figure 4. Relationship between gene expression and methylation in promoter and gene-body regions in mouse germ cells. The expression level of genes in wild-type oocytes (A), sperm (B), and *Dnmt3L*$^{-/-}$ oocytes (C) were divided into 5 percentile groups. The distribution of methylation is shown ±5 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 (*: $p<1\times10^{-9}$).
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 (≥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 became less methylated (0–20%) whereas the other gDMRs maintained or increased DNA methylation (Figure S15). Among ICRs, only *Gnas* exon1A ICR was demethylated (7.8%),

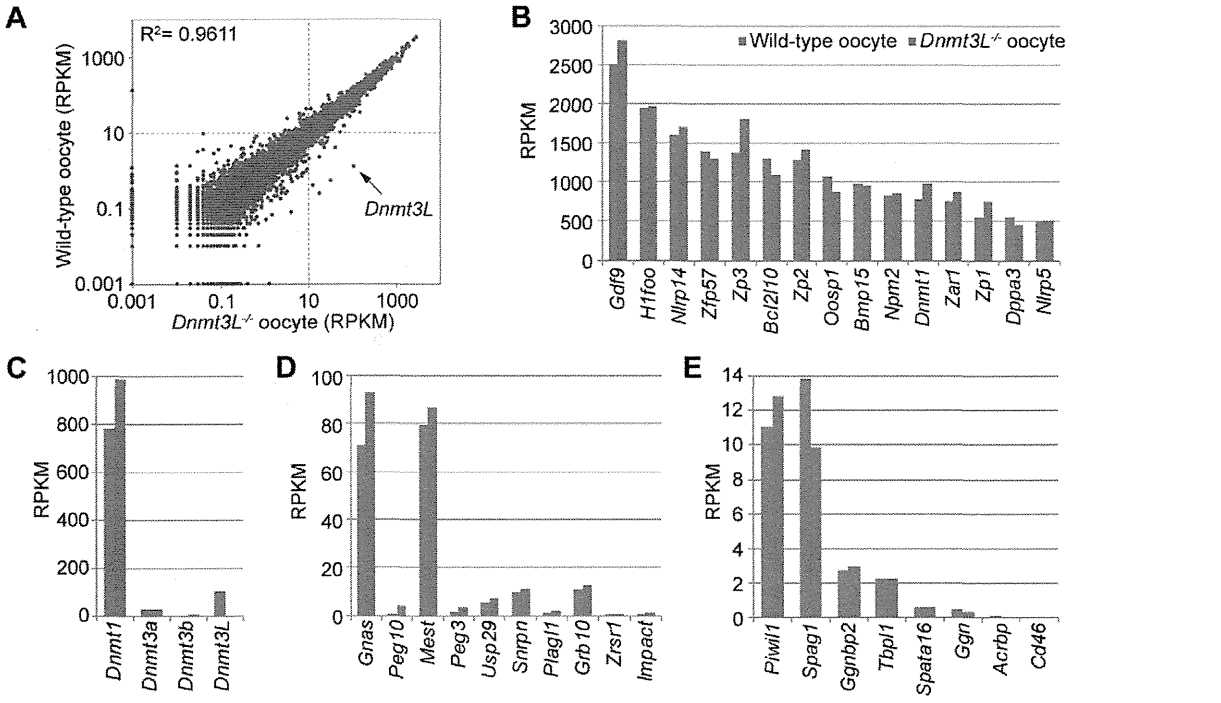


Figure 5. Comparison of gene expression profiles between wild-type and *Dnmt3L*$^{-/-}$ oocytes. (A) Scatter plot and correlation coefficient (R^2) of RPKM values of 20,854 genes in wild-type and *Dnmt3L*$^{-/-}$ oocytes. Expression levels of oocyte-specific genes (B), DNA methyltransferase genes (C), 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 (≥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-

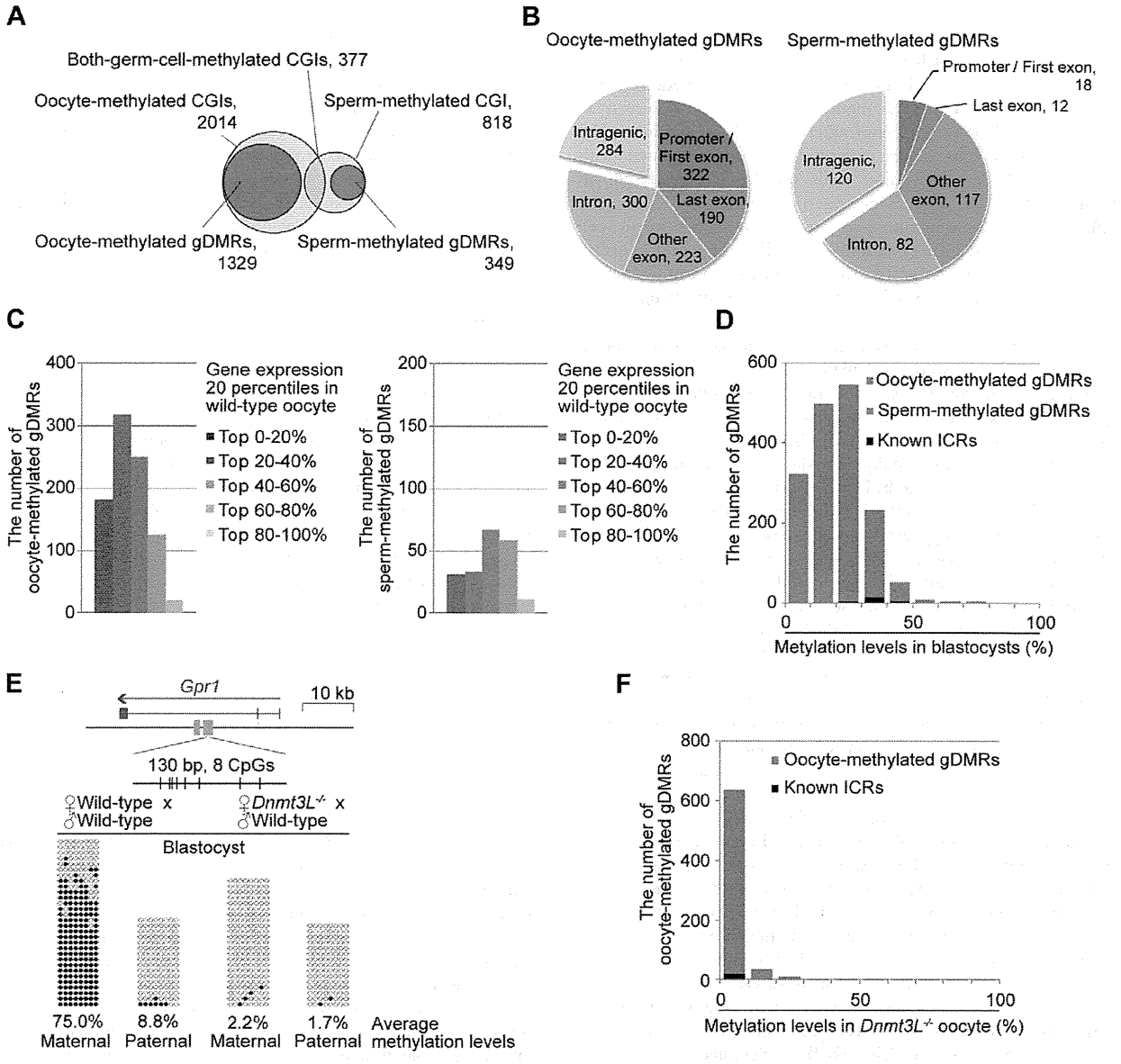


Figure 6. Identification of germline differentially methylated CGIs from DNA methylome profiles. (A) Venn-like diagram of two groups of CGIs, namely, oocyte-methylated CGIs (light pink) and sperm-methylated CGIs (light blue) and two groups of gDMRs, namely, oocyte-methylated gDMRs (red) and sperm-methylated 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-Zdbf2* 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 microograms 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 epididymises 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 μL 10× PCR buffer, 25 μM dNTPs, 1 μL of each PCR Primer PE 1.0 and 2.0 (Illumina) (50 μL 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×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 Bisulfitome 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 Epitect 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′-GATTAGATTAGGTTAG-TTTGGAA-3′) and Gpr1-BSR1 (5′-ACTAAAACACTAAT-CACCAAATA-3′) for the first round; Gpr1-BSF2 (5′-AGAT-TAGGTTAGTTTGGAATT-3′) and Gpr1-BSR2 (5′-AACAC-TAATCACCAAATAATTC-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 Bisulfitome 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)

Figure S12 Relationship between gene expression and intragenic 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 (*: $p<1\times10^{-9}$).
(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.

References

1. Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet 3: 662–673.
2. Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl: 245–254.
3. Lees-Murdock DJ, Walsh CP (2008) DNA methylation reprogramming in the germ line. Epigenetics 3: 5–13.
4. Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat Rev Genet 9: 129–140.
5. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints. Science 294: 2536–2539.
6. 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.
7. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, et al. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429: 900–903.
8. Kono T, Obata Y, Wu Q, Niwa K, Ono Y, et al. (2004) Birth of parthenogenetic mice that can develop to adulthood. Nature 428: 860–864.
9. Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, et al. (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev 22: 1607–1616.
10. Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431: 96–99.
11. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, et al. (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A 102: 4068–4073.

12. Hata K, Kusumi M, Yokomine T, Li E, Sasaki H (2006) Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells. Mol Reprod Dev 73: 116–122.
13. Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, et al. (2009) The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell 17: 775–787.
14. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, et al. (2010) MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev 24: 887–892.
15. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.
16. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, et al. (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26: 779–785.
17. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.
18. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.
19. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, et al. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27: 361–368.
20. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, et al. (2009) Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res 19: 1044–1056.
21. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, et al. (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27: 353–360.
22. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.
23. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20: 320–331.
24. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, et al. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105.
25. Serre D, Lee BH, Ting AH (2010) MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38: 391–399.
26. Li Y, Zhu J, Tian G, Li N, Li Q, et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 8: e1000533. doi:10.1371/journal.pbio.1000533.
27. Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, et al. (2010) Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods 7: 133–136.
28. Howlett SK, Reik W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. Development 113: 119–127.
29. Farthing CR, Ficz G, Ng RK, Chan CF, Andrews S, et al. (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet 4: e1000116. doi:10.1371/journal.pgen.1000116.
30. Barhanin J, Lesage F, Guillemare E, Fink M, Lazdunski M, et al. (1996) K(V)LQT1 and lsK (minK) proteins associate to form the I(Ks) cardiac potassium current. Nature 384: 78–80.
31. Shin J, Bossenz M, Chung Y, Ma H, Byron M, et al. (2010) Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. Nature 467: 977–981.
32. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, et al. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39: 457–466.
33. Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, et al. (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res 20: 972–980.
34. Arnaud P, Hata K, Kaneda M, Li E, Sasaki H, et al. (2006) Stochastic imprinting in the progeny of Dnmt3L−/− females. Hum Mol Genet 15: 589–598.
35. Minami N, Tsukamoto S (2006) Role of oocyte-specific genes in the development of mammalian embryos. Reproductive Medicine and Biology 5: 175–182.
36. Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, et al. (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev 23: 105–117.
37. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, et al. (2010) Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLoS Genet 6: e1001134. doi:10.1371/journal.ppat.1001134.
38. Smallwood SA, Tomizawa SI, Krueger F, Ruf N, Carli N, et al. (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43: 811–814.
39. Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, et al. (2011) Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development 138: 811–820.
40. Hiura H, Sugawara A, Ogawa H, John RM, Miyauchi N, et al. (2010) A tripartite paternally methylated region within the Gpr1-Zdbf2 imprinted domain on mouse chromosome 1 identified by meDIP-on-chip. Nucleic Acids Res 38: 4929–4945.
41. Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. Science 315: 1141–1143.
42. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A 107: 8689–8694.
43. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919.
44. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253–257.
45. Ciccone DN, Su H, Hevi S, Gay F, Lei H, et al. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature 461: 415–418.
46. Ginis I, Luo Y, Miura T, Thies S, Brandenberger R, et al. (2004) Differences between human and mouse embryonic stem cells. Dev Biol 269: 360–380.
47. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, et al. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448: 196–199.
48. Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, et al. (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A 102: 3336–3341.
49. Shiota K, Kogo Y, Ohgane J, Imamura T, Urano A, et al. (2002) Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. Genes Cells 7: 961–969.
50. Kobayashi H, Yamada K, Morita S, Hiura H, Fukuda A, et al. (2009) Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. Genomics 93: 461–472.
51. De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, et al. (2006) Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol 8: 1448–1454.
52. Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, et al. (2010) Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes to Cells 15: 169–179.
53. Shock LS, Thakkar PV, Peterson EJ, Moran RG, Taylor SM (2011) DNA methyltransferase 1, cytosine methylation, and cytosine hydroxymethylation in mammalian mitochondria. Proc Natl Acad Sci U S A 108: 3630–3635.
54. Iqbal K, Jin SG, Pfeifer GP, Szabo PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A 108: 3642–3647.
55. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, et al. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2: 241.
56. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, et al. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471: 68–73.
57. Koide T, Moriwaki K, Uchida K, Mita A, Sagai T, et al. (1998) A new inbred strain JF1 established from Japanese fancy mouse carrying the classic piebald allele. Mamm Genome 9: 15–19.
58. Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. Nucleic Acids Res 36: W170–175.

CHEMICAL REVIEWS

Review

pubs.acs.org/CR

Biomimetic Cell Culture Proteins as Extracellular Matrices for Stem Cell Differentiation

Akon Higuchi,*[,†,‡,§] Qing-Dong Ling,[§,∥] Shih-Tien Hsu,[⊥] and Akihiro Umezawa[‡]

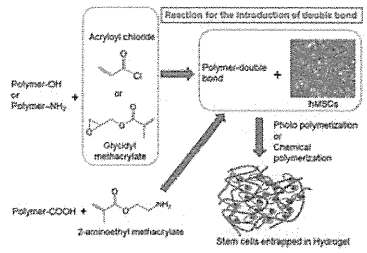
[†]Department of Chemical and Materials Engineering, National Central University, Jhongli, Taoyuan, 32001 Taiwan

[‡]Department of Reproductive Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan

[§]Cathay Medical Research Institute, Cathay General Hospital, No. 32, Ln 160, Jian-Cheng Road, Hsi-Chi City, Taipei, 221 Taiwan

[∥]Institute of Systems Biology and Bioinformatics, National Central University, No. 300, Jhongda RD., Jhongli, Taoyuan, 32001 Taiwan

[⊥]Taiwan Landseed Hospital, 77 Kuangtai Road, Pingjen City, Tao-Yuan County, 32405 Taiwan



CONTENTS

1. INTRODUCTION

Each year, millions of people suffer loss or damage to organs and tissues due to accidents, birth defects, and disease. Stem cells are an attractive prospect for tissue engineering and regenerative medicine because of their unique biological properties. Embryonic stem cells (ESCs) derived from

Received: January 14, 2012
Published: May 23, 2012

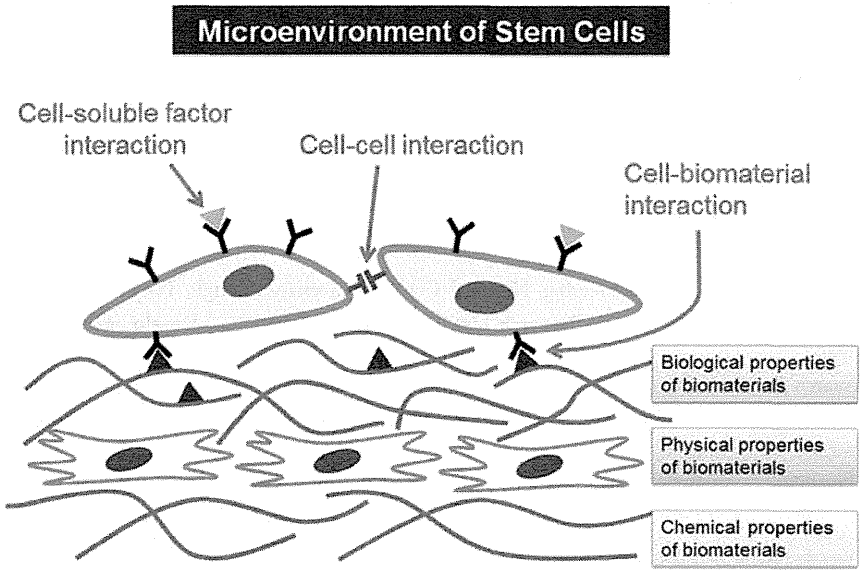


Figure 1. Schematic representation of the microenvironment and niches of stem cells and their regulation by the following factors: (a) soluble factors, such as growth factors or cytokines, nutrients, and bioactive molecules; (b) cell–cell interactions; (c) cell–biomaterial interactions. Biological, physical, and chemical properties of biomaterials also regulate stem cell fate.

preimplantation embryos have the potential to differentiate into any cell type derived from the three germ layers—the ectoderm (epidermal tissues and nerves), mesoderm (muscle, bone, and blood), and endoderm (liver, pancreas, gastrointestinal tract, and lungs).[1] The basis of pluripotency lies in conserved regulatory networks composed of numerous transcription factors and multiple signaling cascades. Together, these regulatory networks maintain human ESCs (hESCs) in a pluripotent and undifferentiated state, and alterations in the stoichiometry of these signals promote differentiation. hESCs have been shown to generate multipotent stem and progenitor cells in vitro and are capable of differentiating into a limited number of cell fates, and thus they have great potential for use in transplantation of cells and tissues into patients.[2]

Although hESCs are promising donor sources for cell transplantation therapies,[1] they face immune rejection after transplantation. Furthermore, ethical issues regarding human embryos hinder their widespread usage. These concerns can be circumvented if pluripotent stem cells can be derived directly from patients' own somatic cells.[3] Recently, pluripotent stem cells similar to ESCs, known as induced pluripotent stem cells (iPSC's), were derived from adult somatic cells by inducing a "forced" expression of certain pluripotent (stem cell) genes[4−6] such as Oct3/4, Sox2, (c-myc), and klf-4, or certain miRNAs[7] or proteins (piPS).[8] iPSC's are believed to be similar to ESCs in many respects, including the expression of certain stem cell genes and proteins, chromatin methylation patterns, doubling time, embryoid body formation, teratoma formation, viable chimera formation, pluripotency, and differentiability.

The pluripotent nature of iPSC's opens many avenues for potential stem cell-based regenerative therapies and for development of drug-discovery platforms.[9,10] The nearest-term therapeutic uses of iPSC's may exist in the transplantation of differentiated nerve cells or β-cells for treatment of Parkinson's Disease and diabetes, respectively, which arise from disorders of single cell types. However, there are several barriers to the clinical application of iPSC's, such as the use of viral vectors, cultivation using xeno-derived materials [e.g., mouse embryonic fibroblasts (MEFs)], and the extremely low efficiency of iPSC generation.[11]

Stem cells have also been isolated from a variety of somatic tissues, including hematopoietic stem cells (HSCs) derived from umbilical cord blood and mesenchymal stem cells (MSCs) derived from bone marrow, umbilical cord blood, umbilical cord, dental pulp, and tissues such as fat. There have been no reports to date of MSCs or fetal stem cells differentiating into tumors, unlike ESCs and iPSC's. Consequently, HSCs, MSCs, and fetal stem cells are the most promising sources of cells for tissue engineering and cell therapies. Currently, MSCs are thought to be the most widely available autologous source of stem cells for practical and clinical applications. Fetal stem cells derived from amniotic fluid are pluripotent cells capable of differentiating into multiple lineages, including cell types of the three embryonic germ layers. Bone marrow MSCs, adipose-derived stem cells (ADSCs), and amniotic fluid stem cells may be more suitable sources of stem cells in regenerative medicine and tissue engineering than ESCs and iPSC's because of ethical concerns regarding their use and concerns about xenogenic contamination arising from the use of mouse embryonic fibroblasts (MEFs) as a feeder layer for ESC and iPSC culture.[11]

Stem cell characteristics, such as proper differentiation and maintenance of pluripotency, are regulated not only by the stem cells themselves but also by the microenvironment. Therefore, mimicking stem cell microenvironments and niches using biopolymers will facilitate the production of large numbers of stem cells and specifically differentiated cells needed for in vitro regenerative medicine. Several factors in the microenvironment and niches of stem cells influence their fate: (i) soluble factors, such as growth factors or cytokines, nutrients, and bioactive molecules; (ii) cell–cell interactions; (iii) cell–biomacromolecule (or biomaterial) interactions; and (iv) physical factors, such as the rigidity of the environment (Figure 1). Some excellent review articles addressing the

engineering of stem cell microenvironments and niches using natural and synthetic biopolymers are listed in Table 1.[11−22]

Table 1. Key Review and Articles Dealing with Biopolymers for Culture and Differentiation of Stem and Progenitor Cells

author	contents	ref (year)
Lee and Mooney	hydrogels for tissue engineering	12 (2001)
Little et al.	biomaterials for neural stem cell microenvironments	13 (2008)
Higuchi et al.	polymeric materials for ex vivo expansion of HSCs	16 (2009)
Mei et al.	combinatorial development of biomaterials for clonal growth of human pluripotent stem cells	17 (2010)
Melkoumian et al.	synthetic peptide-acrylate surfaces for long-term self-renewal of hESCs	18 (2010)
G. J. Delcroix et al.	adult cell therapy for brain neuronal damages and the role of tissue engineering	22 (2010)
Higuchi et al.	biomaterials for the feeder-free culture of hESCs and human iPSC's	11 (2011)
Balakrishnam and Banerjee	biopolymer-based hydrogels for cartilage tissue engineering	14 (2011)
Kim et al.	design of artificial extracellular matrices for tissue engineering	15 (2011)
Engler et al.	matrix elasticity directs stem cell lineage	19 (2006)
Gilbert et al.	substrate elasticity regulates skeletal muscle stem cell self-renewal	20 (2010)
Huebsch et al.	harnessing traction-mediated manipulation of the cell/matrix interface to control stem-cell fate	21 (2010)

These articles focus on biopolymers employed for maintenance of pluripotency of hESCs, iPSC's, or hematopoietic stem cells (HSCs),[16−18] and for specific differentiation lineages such as chondrocytes (cartilage), muscle cells, and neural cells.[13,14,20] There have been no review articles specifically describing extracellular matrix (ECM) scaffolds (ECM in 3D) or ECM-immobilized dish coatings (ECM in 2D) that guide stem cell fates and differentiation. Therefore, this review focuses on the chemical, physical, and biological characteristics of natural biopolymers, especially ECM proteins, which are the major functional biopolymers, and deals with the ability of these biopolymers to guide differentiation of MSCs into osteogenic, chondrogenic, adipogenic, cardiomyogenic, and neural cell lineages.

2. CELL SOURCES AND ANALYSIS OF DIFFERENTIATION LINEAGES OF MSCS

2.1. Cell Sources

Human MSCs (hMSCs), including fetal stem cells, are one of the most widely available autologous sources of stem cells for clinical applications. hMSCs can be obtained from bone marrow,[23,24] adipose tissue,[25,26] dental pulp,[27] and urine,[28] among other sources. Fetal stem cells can be obtained from amniotic fluid,[29−31] umbilical cord,[32−34] menstrual blood,[35,36] umbilical cord blood,[25,34,37] and placenta.[38,39] hMSCs derived from bone marrow and fat are primarily used for biomaterials research on stem cell culture and differentiation because bone marrow MSCs and ADSCs are easily accessible and can be obtained in large quantities. Bone marrow MSCs (BMSCs) are now commercially available from several companies. Stem cell research is facilitated with these stem cell sources because it is not necessary to obtain permission from ethics committees of the Institutional Review Board (IRB) for use of commercially available MSCs. Otherwise, informed consent from donors and permission from the IRB must be obtained.

2.2. Analysis of Differentiation Lineages

MSCs are multipotent stem cells that can be differentiated into various mesodermal lineages, including osteoblasts, chondrocytes (cartilages), adipocytes, myocytes, and cardiomyocytes.[19,40,41] MSCs are also reported to be able to differentiate into ectodermal lineages (e.g., neuron, oligodendrocyte, astrocyte, neural stem cells, and dopamine-secreting cells)[22,42−45] and endodermal lineages (hepatocytes and β-cells),[31,46−52] although with lower probability than mesoderm lineages. Table 2 summarizes methods for characterizing specific differentiated cells from MSCs.[11,34,46,48,51−87]

MSCs differentiate into an osteogenic phenotype in vitro when supplements such as ascorbic acid, β-glycerophosphate, dexamethasone, and/or bone morphogenic protein 2 (BMP-2) are added to the culture medium. Figure 2 shows the expression of several genes and proteins, as well as mineral deposition, by MSCs upon osteogenic differentiation. Runt-related transcription factor 2 (Runx2, also known as Cbfa1, Pebp2αA, and AML3) is a master regulator of osteogenic gene expression and osteoblast differentiation, and it is an early marker of osteogenesis.[88−90] Runx2 activity is stimulated by mitogen-activated protein kinase (MAPK) signaling and is negatively regulated by thrombin-like enzyme 2 (TLE2). Alkaline phosphatase (ALP) activity is an early osteogenic marker, and osteopontin and osteocalcin are late osteogenic markers.[88] Mineral deposition is generated in the late stage of osteogenic differentiation and is detected by Alizarin Red staining (calcium deposition) and von Kossa staining (calcium phosphate deposition).[57,60,62]

MSCs commit to a chondrogenic phenotype when supplied with transforming growth factor-β1 (TGF-β1). Chondrogenic differentiation of MSCs is typically determined by immunostaining for specific proteins, such as collagen type II and Sox9, dye labeling of glycosamino glycans, and evaluation of expression of chondrogenic proteins or transcription factors (such as collagen type II and type X, cartilage oligomeric protein, aggrecan, and Sox9) (Table 2).[63,64,67,70,91] Sulfated glycosaminoglycans (sGAG's) are visualized by staining with Alcian blue.[91] Accumulation of sulfated proteoglycans are also visualized by Safranin O staining.[72]

Only a few groups have investigated adipogenic differentiation of MSCs cultured on natural and artificial biomaterials[53,62,70, 74,75,92] because adipose tissue is in less demand in clinical usage than osteoblasts and cartilage cells. Adipogenic differentiation is also analyzed by immunostaining for specific proteins (vimentin), dye staining of oil droplets, and measuring expression of transcription factors or other marker proteins, such as peroxisome proliferator-activated receptorγ [PPARγ] and adipocyte Protein 2 (aP-2).[53,61,62, 74,75,92] aP-2 is a carrier protein for fatty acids that is primarily expressed in adipocytes.[93] Preadipocytes and mature adipocytes contain multiple or single lipids in cell bodies, respectively. Therefore, Oil Red O or Nile red staining of preadipocytes and mature adipocytes is frequently used for the detection of lipids.

Neural differentiation of MSCs is primarily analyzed by observing characteristic morphologies of neurons, astrocytes, oligodendrocytes, and microglia. Neuronal progenitor cells and early-stage neurons are also identified by Sox1, Sox2, and CD133 gene expression and by nestin and β-tubulin-III

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507−4540

Table 2. Characterization of Differentiation of MSCs into Specific Lineages [Osteoblasts and Chondrocyte (Cartilages)]

differentiation lineage	characterization	specification	ref (example)
1. Osteoblast	morphology	spread shape tends to differentiate into osteoblasts, bonelike nodule formation	53−55
	protein level (immunostaining)	**collagen I**, osteocalcin, osteonectin	56, 57
	surface marker analysis and immunostaining	osteopontin, bisphosphonate [2-(2-pyridinyl)ethylidene-BP] (PEBP), alkaline phosphatase (ALP)	34, 58
	enzyme activity	**alkaline phosphatase**	
	gene level	**runt-related transcription factor 2** [Runx2 or core binding protein A-1 (CBFA-1)], **osterix** (OSX), **osteocalcin** (OCN), **osteopontin** (OPN), **bone sialoprotein** (BSP), alkaline phosphatase, integrin-binding sialoprotein (IBSP), bone γ-carboxyglutamate protein (BGLAP)	34, 58−61
	dye staining	**Alizarin Red staining** (calcium)	62
	mineral deposition	**von Kossa staining** (calcium phosphate)	57, 60
2. Chondrocytes	protein level (immunostaining)	**collagen type II** (Col II), **collagen type X** (Col X), **aggrecan** (AGN), Sox-9, chondroitin-4-sulfate, chondroitin-6-sulfate, sulphated glycosaminoglycans	56, 57, 63−68
	glycosaminoglycan assay	glycosaminoglycan content	
	dimethylmethylene blue (DMMB) assay	proteoglycan (PG) content	69
	hydroxyproline assay	collagen content	65
	gene level	**collagen II**, collagen IX (Col IX), **collagen X**, collagen XI (Col XI), **aggrecan**, Sox 5, Sox 6, **Sox 9**, **cartilage oligomeric protein** (COMP), xylosyltransferase I (XT-1), α-4-*N*-acetylhexosaminyltransferase (EXTL2), β-1,4-*N*-acetylgalactosaminyltransferase (GalNAcT), glucuronyl C5 epimerase (GlcAC5E)	63, 64, 67, 70−73
	dye staining	**Safanin O staining (proteoglycan)**, **Alcian blue staining (proteoglycan)**, EVG-staining, Masson's trichrome staining	34, 62, 64, 67, 70, 72
3. Adipocytes	morphology	round shape cells tends to differentiated into adipocytes	53, 54
	protein level	vimentin, adipocyte lipid-binding protein (ALBP)	53, 74
	enzyme activity	glycerol-3-phosphate dehydrogenase activity	75
	gene level	PPARγ, aP-2	61
	staining	Oil red O and Nile red staining for lipid droplet	62
4. Neural cells	morphology	neuronal-like cells having long neurites	76
	protein level	nestin, neuron-specific class III β-tubulin (TuJ1), galactosylceramidase (GalC), glial fibrillary acidic protein (GFAP), β-tubulin-III, microtubule-associated protein 2 (MAP2), O4, tyrosine hydroxylase (TH), neurofibromatosis (NFM), neurone-specific enolase (NSE)	76−81
	gene level	nestin, Musashi 1, neuron-specific class III β-tubulin (TuJ1), glial fibrillary acidic protein, microtubule-associated protein 2, Sox1, Sox2, CD133, tyrosine hydroxylase, neurofibromatosis, Nurr1, dopamine transporter (DAT), dihydropyrimidinase-related protein 2 (DRP-2), purine-sensitive aminopeptidase (PSA)	11, 61, 76, 81, 82
5. Cardiomyocytes	morphology	contractile cells	
	protein level	**cardiac troponin T (cTnT)**, **desmin**, myosin light chain (MLC), myosin heavy chain (MHC)	81
	gene level	Nkx2.5, GATA-4, MYH-6, TNNT2, TBX-5, myosin light chain (Mlc2a, MLC-2 V), tropomyosin, cTnI, ANP, **desmin**, **myosin heavy chain** (α-MHC, **β-MHC**), **cardiac troponin T**, Isl-1, and Mef2c	11
	electrocardiogram	electrocardiogram	
6. Smooth muscle cells	protein level	α-smooth muscle actin (ASMA), h1-calponin (CALP), SM2	83
	gene level	α-smooth muscle actin, h1-calponin, caldesmon, Smemb, SM22α, SM1, SM2	83
7. Epidermis	protein level	keratin 10 (early marker), filaggrin (intermediate marker), involucrin (late marker)	84
	gene level	keratin 10 (early marker), filaggrin (intermediate marker), involucrin (late marker)	84
8. Hepatocyte	morphology	oval cell morphology, small round cell morphology	46

dx.doi.org/10.1021/cr3000169 | *Chem. Rev.* 2012, 112, 4507−4540

Table 2. continued

differentiation lineage	characterization	specification	ref (example)
	protein level	CXCR4 (endoderm), α-fetoprotein (AFP), albumin (ALB), asialoglycoprotein receptor (ASGPR), cytochrome P450 (CYPA_1), hepatocyte nuclear factor-1α (HNF-1α), hepatocyte nuclear factor-3β (HNF-3β), hepatocyte nuclear factor-4α (HNF-4α), CCAAT-enhancer binding protein α (C/EBPα), cytokeratin-18 (CK18), cytokeratin-19 (CK19), low-density lipoprotein (LDL), GATA4	46, 51, 52, 86, 87, 113
	gene level	Sox17 (endoderm), Foxa2 (endoderm), Gata6 (endoderm), α-fetoprotein, albumin, hepatocyte nuclear factor-1α, hepatocyte nuclear factor-3β, hepatocyte nuclear factor-4α, cytokeratin 18, cytokeratin-19, asialoglycoprotein receptor, tryptophan oxygenase (TO), cytochrome P450 (CYP1A1, CYP2B6), CCAAT-enhancer binding protein α, glucose 6-phosphate (G6P), GATA4	46, 51, 52, 86, 87, 113
	urea assay	urea production	46, 51, 113
	albumin assay	albumin production	52, 86, 113
	glycogen assay	glycogen production	46, 52, 113
	α-fetoprotein assay	α-fetoprotein production	52, 86
	pentoxyresorufin (PROD) assay	cytochrome P450 activity	113
	staining	periodic acid–Schiff (PAS) staining for glycogen storage	46, 113

immunostaining. Mature neurons express neuron-specific class III β-tubulin (Tuj1), microtubule-associated protein 2 (MAP2), neuron-specific enolase (NSE), and purine-sensitive aminopeptidase (PSA). Oligodendrocytes express galactosylceramidase (GalC) and O4. Dopaminergic neurons express tyrosine hydroxylase (TH), neurofibromatosis (NFM), and dopamine transporter (DAT). Nerve cells are electrically excitable cells that transmit information by electrical and chemical signaling. Therefore, electrical and action potentials in nerve cells can be monitored using electrodes.

3. PREPARATION OF CULTURE MATRIX

Biomimetic stem cell cultures can be categorized as two-dimensional (2D) or three-dimensional (3D). 2D culture is useful for basic research to investigate the fundamental interactions between cells and immobilized nanosegments on dishes, but 3D culture of stem cells in biomaterials is essential for clinical applications. Figure 3 shows some examples of biomaterial designs for carrying stem cells, as well as direct injection of biomaterials without cells. The injection of hydrogels or scaffolds containing stem cells is categorized as 3D cultures. Cell sheets prepared on a surface-grafting polymer having low critical solution temperature (LCST), such as poly(*N*-isopropylacrylamide) (poly(NIPAM)), can be prepared on 2D dishes.[94,95] Recently, patch sheets of immobilized antibodies or ligands targeting specific stem cells, which recruit the stem cells from the patient's body, are reported to be effective in gathering autologous stem cells at sites of injury.[40] The following sections describe methods for (a) surface immobilization of ECM proteins and ECM-mimicking peptides on 2D culture dishes and (b) preparing hydrogels or scaffolds containing ECM proteins and ECM-mimicking peptides for 3D culture of stem cells.

3.1. ECM Immobilization on 2D Dishes

Typically, 2D cell culture dishes are coated with ECM proteins or ECM-mimicking peptides. Tables 3 and 4 show examples of the ECM proteins and ECM-mimicking peptides used to coat culture dishes and their binding sites on stem cells.[16,18,53,58,71,83,91,96–118] Collagen types I, II, and IV, gelatin, laminin, laminin-1, laminin-5, vitronectin, and fibronectin are typically used as coating materials.[58,71,83,91,96–98,100–102] ECM-mimicking peptides (e.g., RGD, DGEA, YIGSR, IKVAV, KRSR, P15, and GFOGER) are commonly used as coating or grafting materials.[16,18,53,97,103–118] Covalent binding is preferable for long-term effects in culture, but noncovalent coating is the simplest method for the preparation of dishes with immobilized ECM proteins or ECM-mimicking peptides. Figure 4 summarizes typical surface reactions for the covalent immobilization of ECM proteins and peptides on dishes. Proteins and ECM-mimicking peptides should be used in aqueous solution, as they are unstable biomolecules. Reactions between amino groups and between amino groups and carboxylic acids can be used to bind ECM proteins and ECM-mimicking peptides to plastic dishes. These plastic surfaces should therefore have amino groups, carboxylic acid groups, or hydroxyl groups to bind and immobilize ECM proteins or peptides. For dishes made of polyesters, such as poly(ε-caprolactone) (PCL), poly(glycolic acid) (PGA), poly(lactic acid) (PLA), or poly(lactic acid-*co*-glycolic acid) (PLGA), treatment with a diamine, such as hexamethylene diamine, generates amino groups on the surface by an aminolysis reaction. Then, ECM proteins and ECM-mimicking

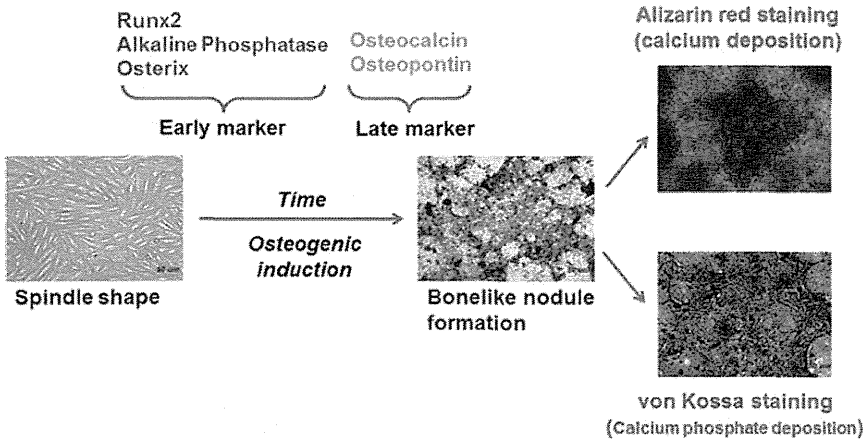


Figure 2. Osteogenic differentiation of MSCs, gene expression, and mineral deposition at early and late stages.

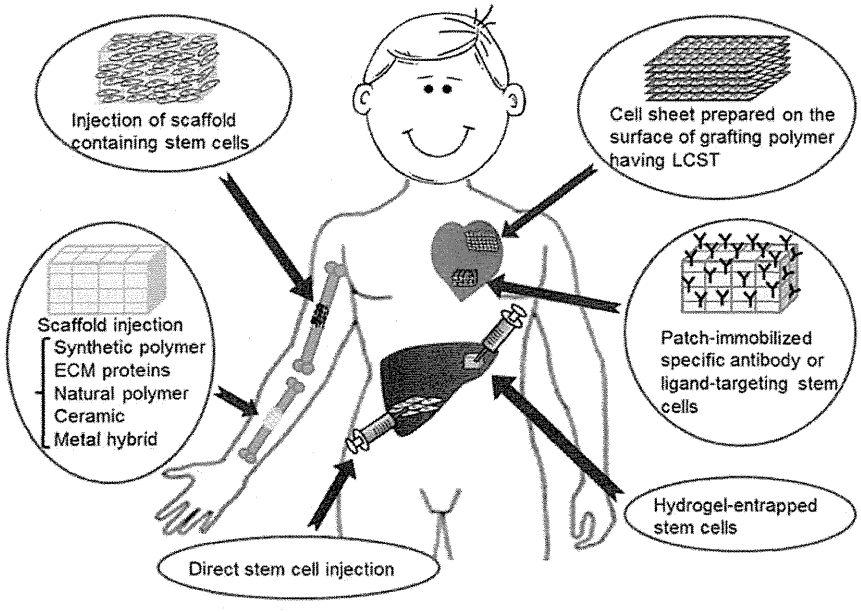


Figure 3. Some examples of biomaterial designs with and without stem cells for the injection of biomaterials in clinical applications: (a) injection of scaffold containing stem cells, (b) injection of scaffold without cells, (c) direct stem cell injection, (d) injection of cell sheets, (e) injection of patch-immobilized specific antibody or ligand-targeting stem cells, and (f) injection of hydrogel-entrapped stem cells.

peptides can be covalently immobilized using hexamethylene diisocyanate (HMDIC), 1,6-dimethyl suberimidate dihydrochloride (DMS),[119] or NHS/EDC reagent,[18] where NHS is *N*-hydroxysuccinimide and EDC is *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (Figure 4). EDC is a water-soluble carbodiimide that is generally used in the 4.0–6.0 pH range. Therefore, it is possible to immobilize ECM proteins and ECM-mimicking peptides in aqueous solution using NHS/EDC reagents. The covalent bonding between amino groups can be reacted with aqueous DMS.[119]

Genipin is generally used to cross-link proteins, such as collagen and gelatin, and chitosan via amino groups.[120,121] Genipin can also be used for the immobilization of ECM proteins and peptides on the surface of culture dishes with amino groups (Figure 4). NHS/EDC, DMS, and genipin are the recommended reagents to covalently immobilize ECM proteins and ECM-mimicking peptides on culture dishes.

3.2. 3D Culture in Hydrogels

Hydrogels are physically or chemically cross-linked polymer networks that are able to absorb large amounts of water. Injectable hydrogels containing stem cells can be delivered to sites of damage in patients with minimal invasiveness, and the hydrogels ensure that stem cells remain localized to the damaged sites more effectively than injected cells alone. Physical cross-linking is performed on ECM proteins with thermosensitive properties of lower critical solution temperature (LCST) or upper critical solution temperature (UCST), such as collagen and gelatin. Collagen can be dissolved in aqueous solutions at low temperature and forms gels at ~37 °C because of its LCST characteristics, and gelatin can be dissolved in aqueous solution at high temperatures and forms gels at room temperature because of its UCST. Therefore, stem cells can be dissolved in ECM protein solutions and efficiently entrapped in ECM gels at 20–37 °C. However, most ECM

Table 3. ECM Immobilized on Dishes for Adhesion, Differentiation, And Proliferation of Stem Cells and Some Examples of the Literature

ECM	binding site of cells	ref
collagen I	integrin ($\alpha V\beta 3$, $\alpha 2\beta 1$)	58, 96
collagen I	integrin ($\alpha 1\beta 1$)	97
collagen I	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 3\beta 1$)	71
collagen II	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$)	71, 91
collagen IV	integrin ($\alpha 2\beta 1$, CD44)	98
gelatin		99
fibronectin	integrin ($\alpha 4\beta 1$, $\alpha 5\beta 1$, $\alpha V\beta 3$, $\alpha IIb\beta 3$, $\alpha V\beta 6$, $\alpha V\beta 5$)	58, 96
laminin	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	100
laminin-1 (laminin 111)	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 6\beta 1$, $\alpha 7\beta 1$, $\alpha 9\beta 1$), α-dystroglycan, suifade, and heparan sulfate proteoglycan	83, 101
laminin-5 (laminin 332)	integrin ($\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	102
laminin-10/11	integrin ($\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	100
vitronectin	integrin ($\alpha V\beta 3$, $\alpha V\beta 5$)	58, 96

proteins and ECM-derived oligopeptides (ECM peptides) need other forms of cross-linking to trap stem cells and generate hydrogels. Typically, photocross-linking and chemical cross-linking of ECM proteins and ECM peptides are used. There are several excellent reviews that discuss hydrogel preparation and reaction in detail.[12,14] Therefore, this section deals briefly with the preparation of ECM hydrogels using photocross-linking and chemical cross-linking with cross-linking agents. The application of ECM hydrogels containing stem cells is discussed in section 5 for specific ECM proteins and ECM peptides.

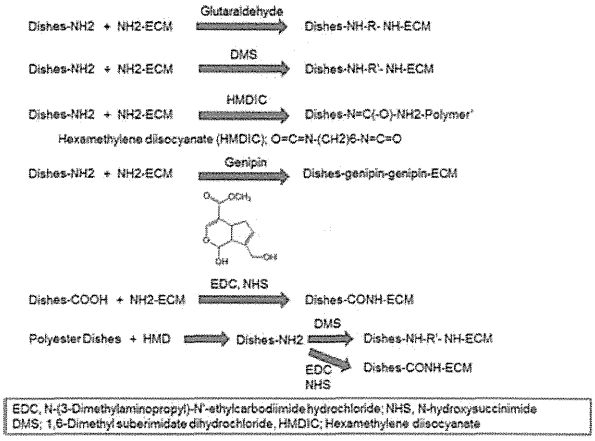


Figure 4. Surface reactions of covalent immobilization of ECM proteins and ECM-mimicking peptides on dishes.

3.2.1. Photocross-Linking of ECM Proteins and ECM Peptides. Hydrogels containing stem cells can be easily prepared by UV irradiation of ECM proteins and ECM–peptide solutions. These preparations can be used as injectable hydrogels via photocross-linking. However, it is first necessary to introduce double bonds into ECM proteins and ECM peptides for photocross-linking. ECM proteins and ECM peptides have $-OH$, $-NH_2$, and $-COOH$ functional groups. Double bonds can be introduced into ECM proteins and ECM

Table 4. ECM-Mimicking Peptides Immobilized on Dishes for Adhesion, Differentiation, And Proliferation of Stem Cells

ECM-mimicking peptide	ECM proteins for mimicking	binding site of cells	ref
DGEA	collagen I	integrin ($\alpha 2\beta 1$)	103–105
GTPGPQGIAGQRGVV (P15)	collagen I	integrin ($\alpha 2\beta 1$)	103, 106
$(RADA)_4$GG**DGEA**	collagen I	integrin ($\alpha 2\beta 1$)	116
$(RADA)_4$GGFPGERGVEGPGP	collagen I		116
GFOGER	collagen	integrin ($\alpha 2\beta 1$)	103, 107, 108
MNYYSNS	collagen IV		109
RGD	collagen I	integrin ($\alpha V\beta 3$)	97, 110
ELIDVPST (CS-1)	fibronectin	integrin ($\alpha 4\beta 1$); VLA-4	16, 111
FN-40	fibronectin	integrin ($\alpha 4\beta 1$, VLA-4)	16, 112
FN-120	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	16, 112
FN–CH296	fibronectin	integrin ($\alpha 4\beta 1$, $\alpha 5\beta 1$)	16, 112
KGGAVTG**RGD**SPASS	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	18, 113
GRGDSPK	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	18, 113
KNNQKSEPLIGRKKT	fibronectin	heparin-binding domain	53
RGDS	fibronectin		109
PHSRN	fibronectin		109
KYGAASIKVAVSADR	laminin		18, 114
YIGSR	laminin		109
IKVAV	laminin		115
PPFLMLLKGSTR	laminin-5 (laminin332)	integrin ($\alpha 3\beta 1$)	
(RADA)4-GGP**DSGR**	laminin		116
(RADA)4-GGSDPG**YIGSR**	laminin		116
(RADA)4-GGIKVAV	laminin		116
KGGPQVT**RGD**VFTMP	vitronectin	integrin ($\alpha V\beta 5$)	18, 117
KGGNGEP**RGD**TYRAY	bone sialoprotein (BSP)		18, 118
PEO4-NGEP**RGD**TYRAY	BSP-linker		18, 118
RGD	osteopontin	integrin ($\alpha V\beta 3$)	97

peptides by the reactions of acryloyl chloride,[122] glycidyl methacrylate,[12,123] and 2-aminoethylmethacrylate[12,124] (Figure 5). Figure 5 also shows a schematic for preparation method of

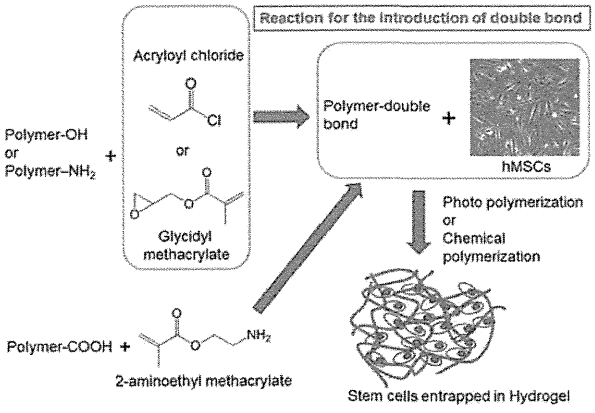


Figure 5. Schematic of the preparation method of hydrogels with entrapped stem cells by photopolymerization.

hydrogels with entrapped stem cells by photopolymerization. Aqueous solutions containing stem cells and macromers of ECM proteins and ECM peptides are irradiated with UV light to generate hydrogels with entrapped stem cells.

Poly(ethylene glycol)diacrylate (PEODA) is typically added to the reaction solution to generate optimal hydrogels.[65,125–129] Yang et al. prepared PEODA hydrogels incorporating RGD adhesive peptides and goat BMSCs by photopolymerization. They found that RGD-conjugated PEODA hydrogels promoted the osteogenic differentiation of BMSCs, and RGD enhanced differentiation in a dosage-dependent manner, with the highest concentration (2.5 mM) in the reaction solution being optimal in their study.[125]

3.2.2. Chemical Cross-Linking of Hydrogels. Hydrogels of ECM proteins can also be prepared by chemical cross-linking. Similar to ECM protein immobilization on 2D dishes, as discussed in section 3.1, NHS/EDC, DMS, HMDIC, and genipin are typically used as cross-linking agents. Glutaraldehyde is not commonly used for the preparation of hydrogels in tissue engineering because it is relatively toxic to stem cells. DMS, HMDIC, and genipin allow cross-linking between amino groups, whereas NHS/EDC leads to cross-linking between carboxylic acids and amino groups in ECM proteins.

Chang et al. compared gelatin hydrogels cross-linked with genipin and gelatin hydrogels cross-linked with glutaraldehyde.[120] They found that the degree of inflammatory reaction in wounds treated with the genipin-cross-linked gelatin was significantly less severe than those covered with the glutaraldehyde-cross-linked gelatin in vivo.[120] In addition, the healing rates of wounds treated with the genipin-cross-linked gelatin were notably faster than those with glutaraldehyde-cross-linked hydrogels.[120]

3.3. 3D Culture in Scaffolds

Scaffolds seeded with stem cells can support 3D tissue formation artificially. It is optimal for scaffolds (a) to allow cell attachment and migration, (b) to allow diffusion of nutrients, growth factors, and waste secreted by cells, and (c) to have mechanical properties similar to the natural tissue. Most of the scaffolds have high porosity (>80%) and large pore size (200–800 μm), which allow diffusion of nutrients, growth factors, and waste, but these properties also lead to weak mechanical properties. Biodegradability of scaffolds is often required because scaffolds should be absorbed by the surrounding tissues without the necessity of surgical removal. It is preferable that the degradation rate of scaffolds should be matched to the speed of tissue formation. The degradation speed of scaffolds can be regulated by the degree of cross-linking. Scaffolds prepared from ECM proteins and ECM peptides are desirable because of their biodegradable characteristics. ECM proteins used for the preparation of scaffolds are typically collagen type I, collagen type II, gelatin, fibronectin, laminin, and vitronectin. ECM proteins can be used as (a) coating materials, (b) blending materials, and (c) main materials of scaffolds.

3.3.1. Preparation of Scaffolds. There are several methods used to prepare scaffolds for tissue engineering and 3D culture of stem cells, including (a) freeze-drying, (b) salt leaching, (c) porogen leaching, (d) use of nonwoven fabric or mesh, (e) nanotopography, and (f) electrospinning. In the freeze-drying method, ECM proteins are dissolved in a buffer solution. The ECM solution is frozen at −20 or −80 °C and then lyophilized in a freeze-dryer before being washed and stored (Figure 6). If necessary, the scaffolds are also cross-linked.

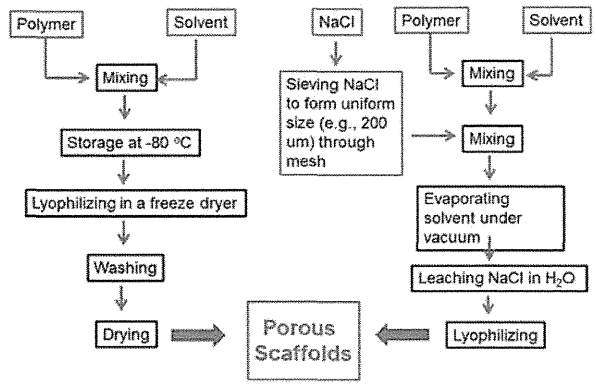


Figure 6. Typical preparation method of porous scaffolds by freeze-drying (a) and salt leaching (b).

The salt-leaching method is as follows. Biopolymers and/or ECM proteins are dissolved in a solvent. Salt, typically NaCl, is sieved to generate a uniform distribution of size using filtration through mesh and added into the solution. The solvent of the solution is vaporized under vacuum to generate dry scaffolds. Salt is then leached from the scaffolds by immersion in water after drying the scaffolds (Figure 6). The porogen-leaching method is a similar method to the salt-leaching method, but other uniformly sized particles, such as polymeric particles, are used instead of salt.

3.4. 3D Culture in Nanofibers

Peptide amphiphiles (PAs), which have a hydrophilic domain and a hydrophobic domain, are known to spontaneously generate self-assembled nanofibers above critical micelle concentrations.[109,116,130] MSC differentiation on self-assembled nanofibers using ECM peptides is discussed in section 5.8.1.

A typical method to create nanofibers is electrospinning. Electrospun scaffolds can support cell adhesion and growth and

promote differentiation of stem cells.[131] Nanofibers can be generated from a spinning nozzle when high voltage is applied between the spinning nozzle and a flat metal collector. Typical electrospinning products are flat and highly interconnected scaffolds with a nonwoven fabric sheetlike morphology. These characteristics hinder cell infiltration and growth throughout the scaffolds. Blakeney et al. have developed a three-dimensional cotton ball-like electrospun scaffold that consists of low-density, uncompressed nanofibers.[131] A grounded spherical dish and an array of needle-like probes were used instead of a traditional flat-plate collector to create a cotton ball-like scaffold. Scanning electron microscopy showed that the cotton ball-like scaffold consisted of electrospun nanofibers with a similar diameter, but with larger pores and less dense structures than traditional electrospun scaffolds.[131] The cotton-ball like scaffolds prepared from ECM proteins by electrospinning will be interesting for use as scaffolds for guiding specific lineages of stem cell differentiation.

4. PHYSICAL PROPERTIES OF BIOPOLYMERS (BIOMATERIALS) GUIDE STEM CELL DIFFERENTIATION FATE (LINEAGE)

The interactions between MSCs and ECM proteins are classified as physical, chemical, and biological. It has recently been recognized that stem cell differentiation is directed by physical properties of culture materials as well as by biochemical responses to growth factors and ECM proteins.[19,20,132] Cells in bone, muscle, liver, and brain tissues reside in different environments that have diverse physical properties.[133] The matrix stiffness for differentiated cells is known to influence focal-adhesion structure and the cytoskeleton.[134−139] Engler et al. reported that soft materials, with similar stiffness to the brain, tend to differentiate MSCs into neurogenic cells, whereas stiffer materials that mimic muscle guide MSCs into myogenic cells and rigid materials similar to collagenous bone induce osteogenic differentiation (Figure 7).[19] However, this work was performed on a 2D surface of hydrogels coated with collagen. The effect of stiffness in 3D culture may produce different results than in 2D culture.

Gilbert et al. also reported that the elasticity of culture materials regulates self-renewal of skeletal muscle stem cells.[20]

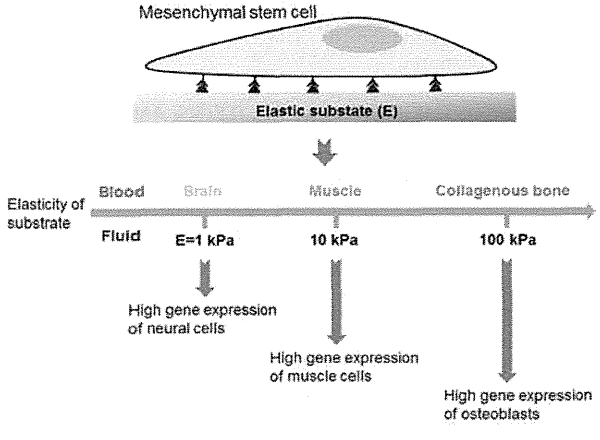


Figure 7. Physical properties decide the fate of stem cell cultured on biomaterials with different elasticity. Modified with permission from ref 19. Copyright 2006 Elsevier Inc.

Muscle stem cells (MuSC's) exhibit robust regenerative capacity in vivo, but this capacity is rapidly lost in culture. They showed that the elasticity of culture materials was a potent regulator of MuSC fate. MuSC's cultured on soft hydrogel substrates that mimicked the elasticity of muscle (12 kPa) self-renew in vitro and contributed extensively to muscle regeneration when transplanted into mice, unlike MuSC's grown on rigid plastic dishes (~106 kPa), as shown by histology and bioluminescence imaging. These studies provide evidence that propagation of adult muscle stem cells is possible by recapitulating physiological tissue rigidity.[20] This finding may contribute to future cell-based therapies for muscle-wasting diseases.

The effect of physical interactions between MSCs and culture materials on stem cell fate is discussed in several articles.[19,20,61,133,140−154] Some landmark findings are summarized in Table 5, and some examples of physical effects on differentiation of MSCs cultured on ECM proteins are reviewed here.

Table 5. Some Articles Discussing Physical Effect of Substrates on Differentiation of MSCs Cultured on the Substrates

authors	contents	ref (year)
J. R. Mauney et al.	**mechanical stimulation** promotes osteogenic differentiation of hBMSCs	140 (2004)
J. S. Park et al.	differential effects of **equiaxial and uniaxial strain** on MSCs	141 (2004)
V. E. Meyers et al.	**microgravity** disrupts collagen I/integrin signaling during osteogenic differentiation of hMSCs	142 (2004)
V. I. Sikavitsas et al.	**flow perfusion** enhances the calcified matrix deposition of marrow stromal cells in scaffolds	143 (2005)
H. Hosseinkhani et al.	**perfusion culture** enhances osteogenic differentiation of MSCs	144 (2005)
A. J. Engler et al.	**matrix elasticity** directs stem cell lineage specification	19 (2006)
R. D. Sumanasinghe et al.	osteogenic differentiation of hMSCs in collagen matrices: effect of **uniaxial cyclic tensile strain**	145 (2006)
D. F. Ward et al.	**mechanical strain** promotes osteogenic differentiation of hMSCs	61 (2007)
E. K. F. Yim et al.	**nanostructures** inducing differentiation of hMSCs into neurinal lihneage	154 (2007)
B. Lanfer et al.	growth and differentiation of MSCs on **aligned collagen** matrices	146 (2009)
Q. Li et al.	ECM with the **rigidity** of adipose tissue helps adipocytes maintain insulin responsiveness	147 (2009)
M. Zscharnack et al.	**low O_2** expansion improves subsequent chondrogenesis of BMSCs in hydrogel	148 (2009)
C. H. Huang et al.	interactive effects of **mechanical stretching** and ECM proteins on initiating osteogenic differentiation of hMSCs	149 (2009)
P. M. Gilbert et al.	substrate **elasticity** regulates skeletal muscle stem cell self-renewal in culture	20 (2010)
G. C. Reilly and A. J. Engler	intrinsic ECM properties regulate stem cell differentiation (**mechanobiology**)	150 (2010)
J. M. Kemppainen and S. J. Hollister	differential effects of **designed scaffold** permeability on chondrogenesis by BMSCs	151 (2010)
E. K. F. Yim et al.	**nanotopography**-induced changes in focal adhesions, cytoskeletal organization, and mechanical properties of hMSCs	152 (2010)
J. Tang et al.	regulation of stem cell differentiation by cell−cell contact on **micropatterned** material surfaces	153 (2010)
P. A. Janmey and R. T. Miller	mechanisms of **mechanical** signaling in development and disease	133 (2011)

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507−4540