

# A unique configuration of genome-wide DNA methylation patterns in the testis

C. C. Oakes<sup>\*†</sup>, S. La Salle<sup>\*†</sup>, D. J. Smiraglia<sup>‡</sup>, B. Robaire<sup>\*§</sup>, and J. M. Trasler<sup>\*†||\*\*</sup>

Departments of <sup>\*</sup>Pharmacology and Therapeutics, <sup>§</sup>Obstetrics and Gynecology, <sup>†</sup>Pediatrics, and <sup>||</sup>Human Genetics, McGill University, Montreal, QC, Canada H3A 1B1; <sup>‡</sup>Montreal Children's Hospital Research Institute, Montreal, QC, Canada H3H 1P3; and <sup>‡</sup>Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY 14250

Edited by Ryuzo Yanagimachi, University of Hawaii, Honolulu, HI, and approved November 6, 2006 (received for review August 29, 2006)

In the mammalian lifecycle, the two periods of genome-wide epigenetic reprogramming are in the early embryo, when somatic patterns are set, and during germ cell development. Although some differences between the reprogrammed states of somatic and germ cells have been reported, overall patterns of genomic methylation are considered to be similar. Using restriction landmark genomic scanning to examine  $\approx 2,600$  loci distributed randomly throughout the genome, we find that the methylation status of testicular DNA is highly distinct, displaying eightfold the number of hypomethylated loci relative to somatic tissues. Identification and analysis of  $>300$  loci show that these regions are generally located within nonrepetitive sequences that are away from CpG islands and 5' regions of genes. We show that a contributing factor for these differences is that the methylation state of non-CpG-island DNA is correlated with the regional level of GC content within chromosomes and that this relationship is inverted between the testis and somatic tissues. We also show that in *Dnmt3L*-deficient mice, which exhibit infertility associated with abnormal chromosomal structures in germ cells, this unique testicular DNA methylation pattern is not established. These special properties of testicular DNA point to a broad, distinct epigenetic state that may be involved in maintaining a unique chromosomal structure in male germ cells.

germ cells | restriction landmark genomic scanning | epigenetic

Although germ cells faithfully carry parental genetic information to the next generation, they edit the epigenetic information that they inherit. Epigenetic marks in the form of DNA methylation are erased in primordial germ cells in favor of the establishment of a sex-specific, *de novo* epigenetic program (1). This initial programming is not only important for gametogenesis (2–4) but also provides the basis for genomic imprinting (5), because some sex-specific marks persist after a second wave of demethylation that occurs during early embryogenesis. Furthermore, it has been shown recently that some epigenetic marks can pass from one generation to the next by avoiding both periods of reprogramming in what is termed “epigenetic inheritance” (6, 7). A clear appreciation of the genome-wide nature of the germ cell epigenetic program is paramount to the understanding of these important processes.

DNA methylation in mammals occurs at cytosine residues in CpG dinucleotides. CpG islands are generally thought to be devoid of methylation, whereas the rest of the genome is highly methylated in nondeveloping tissues (1). A recent study in humans confirmed that CpG islands represent the only major unmethylated genomic compartment in somatic (brain) DNA and implied that germ cells were similar, but the latter was not tested directly (8). Although DNA methyltransferases are differentially regulated and essential during germ-cell development (9, 10), few studies to date have directly examined germ cell genome-wide DNA methylation patterns. Because of the larger number of germ cells produced in the male than in the female, their greater accessibility and their continual renewal, most germ-cell studies focus on spermatozoa and testicular tissues. In the adult, the testis is composed of  $>80\%$

germ cells, mainly in the meiotic and postmeiotic phases of spermatogenesis (11) and, thus, will be referred to in this study as nonsomatic. Although the data are limited, imprinted loci (5), some repetitive sequences (12), and tissue-specific genes (13), point to a unique state of DNA methylation in the testis. Studies using restriction landmark genomic scanning (RLGS) have found methylation differences between the testis and other tissues (14, 15), but these studies either did not identify the sequences or the analysis was restricted to tissue-specific genes. In this study, we have used RLGS as well as other techniques to explore, in a nonbiased fashion, the DNA methylation pattern in the testis.

## Results and Discussion

RLGS was used to survey  $\approx 2,600$  genomic NotI restriction sites that are found mostly in CpG islands but also in noncoding unique and repetitive sequences not within CpG islands. Analysis of testis, liver, intestine, and brain from adult mice revealed 241 spots that display tissue-specific intensity (Fig. 1). Although each tissue demonstrates a unique spot profile, more than one-half of all differentially methylated (DM) spots are specific to the testis. The testis displays eightfold the number of exclusively hypomethylated spots compared with the average of the somatic tissues. A comparison of adult testis and mature sperm isolated from the cauda of the epididymis display identical RLGS patterns, consistent with the fact that the adult testis is composed of a majority of mature germ cells [see supporting information (SI) Fig. 7].

Using a combination of the mouse RLGS spot cloning library (16) and the virtual RLGS method (D.J.S., R. Wenger, and C. Plass, unpublished work), we identified the genomic location of approximately half (125) of the DM spots (Fig. 2 and SI Table 2). We further identified the genomic locations of 48 “absent” spots that correspond to unchanged methylated NotI sites and, combining 89 spots identified in ref. 16, 186 unchanged unmethylated spots. Over 90% of loci that are unmethylated and unchanged in all tissues are found in CpG islands and within the 5' region of genes; unchanged methylated loci are found in regions of low CpG content and away from 5' regions (Fig. 3). Interestingly, DM loci are found in sequences of diverse CpG content and in various locations within and outside of known genes. Focusing on the loci that are either DM between testis and somatic tissue (testis-specific DM) or DM between different somatic tissues (somatic-specific DM) reveals

Author contributions: C.C.O., S.L., D.J.S., B.R., and J.M.T. designed research; C.C.O. and S.L. performed research; D.J.S. contributed new reagents/analytic tools; C.C.O. analyzed data; and C.C.O., D.J.S., B.R., and J.M.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: DM, differentially methylated; LTR, long terminal repeat; RLGS, restriction landmark genomic scanning; qAMP, quantitative analysis of DNA methylation by real-time PCR.

\*\*To whom correspondence should be addressed at: McGill University–Montreal Children's Hospital Research Institute, 2300 Tupper Street, Montreal, QC, Canada H3H 1P3. E-mail: jacquetta.trasler@mcgill.ca.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0607521104/DC1](http://www.pnas.org/cgi/content/full/0607521104/DC1).

© 2006 by The National Academy of Sciences of the USA





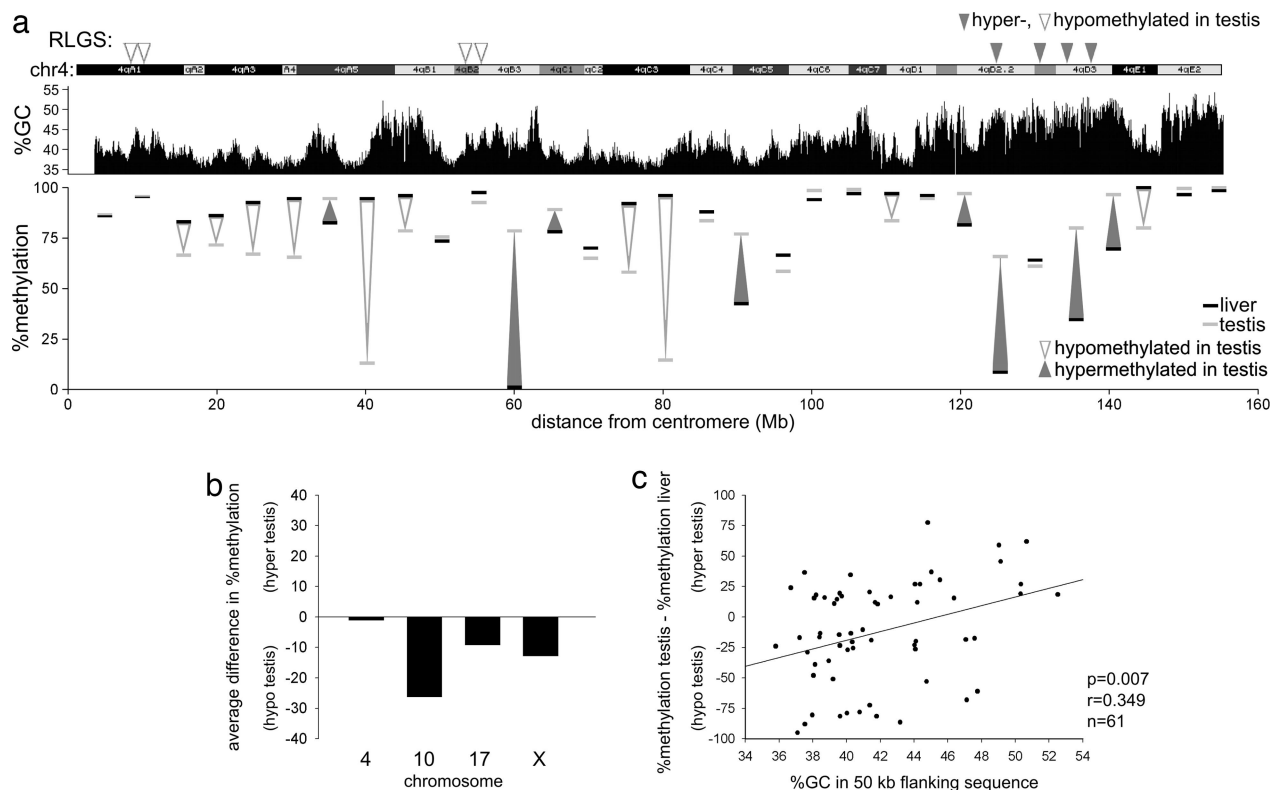
**Table 1. The number and characterization of repetitive elements associated with identified RLGs loci**

DNA methylation pattern	Total loci, <i>n</i>	Repetitive, <i>n</i> (%)	Hypermethylated	Hypomethylated	Classification of repetitive elements					
					Non-LTR-containing		LTR-containing (solitary LTR)			
					SINE	LINE	IAP	ETn	RLTR	MaLR
Unchanged unmethylated	161	0 (0)	0	0	0	0	0	0	0	0
Unchanged methylated	47	23 (49)	23	0	1	3	11	3 (1)	4 (2)	1 (1)
DM testis vs. somatic	60	22 (37)	3	19	0	1	0	1	18 (17)	2 (2)
DM intersomatic	26	3 (12)	0	3	0	0	0	3	0	0

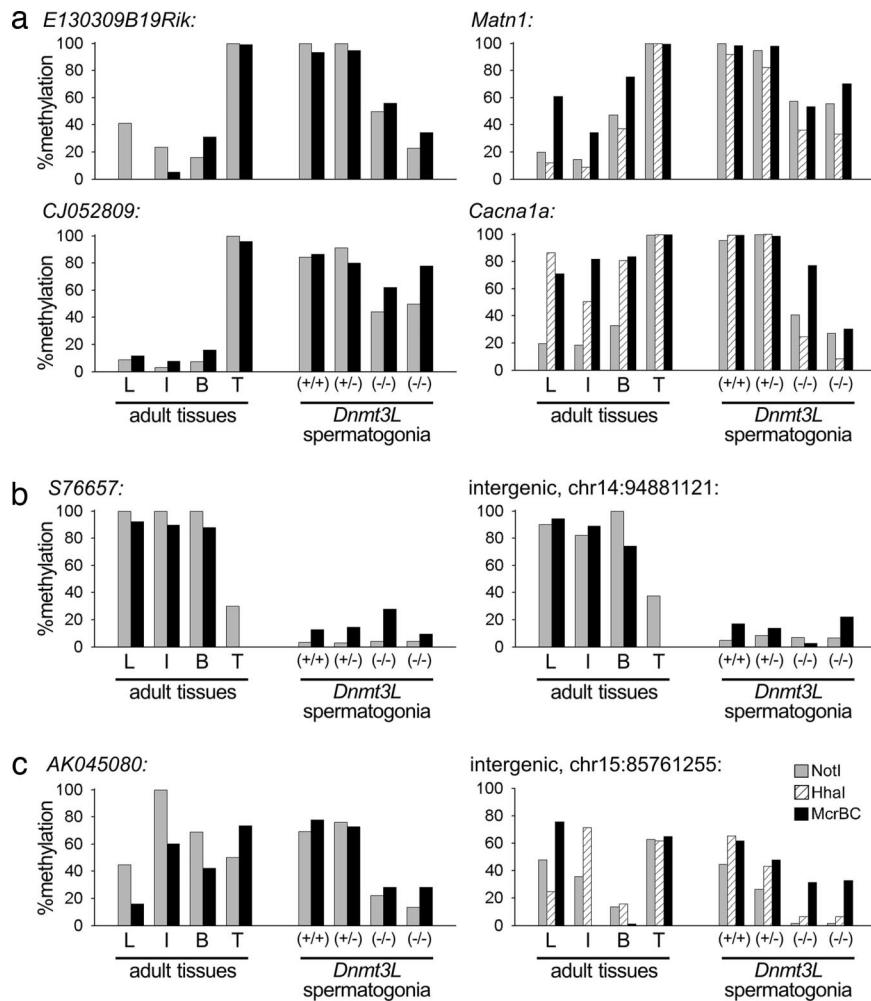
RLTR, retrotransposon LTR; ERVK, endogenous retrovirus K; IAP, intracisternal-A particle; ETn, early transposon; MT, mouse transcript; MaLR, mammalian retrotransposon-like; SINE, short interspersed transposable; LINE, long interspersed transposable.

The majority of loci that are found to be DM in this study occur within non-CpG-island sequences in the testis. As a potential explanation, we observed that those loci that are hypermethylated in the testis are significantly more commonly located within R (reverse)-type bands on chromosomal ideograms, and that loci hypomethylated in the testis are commonly located in G (Giemsa) bands (Fig. 4*a* and SI Fig. 9). Heterogeneous banding patterns reveal regional differences in chromosomal structure and are related to the disparity in the long-range G+C nucleotide composition that occurs along chromosomes (24). R-bands generally contain above-average and G-bands contain below-average GC content. We tested whether the sequence that surrounds each locus

in these two groups differs in overall GC content. We found that loci that are hypomethylated specifically in the testis are found in regions of lower GC content (SI Fig. 9). Consistent with the banding-pattern results, loci that are hypomethylated in somatic tissue are located in regions of higher GC content, opposite to what is found in the testis. To further compare the differential nature of the relationship between methylation and GC content between testis and somatic tissues, the level of methylation of all identified non-CpG-island hypomethylated loci in each tissue was compared with its regional GC content. In all somatic tissues, there is a significant negative relationship between the level of methylation and GC content, demonstrating that lower GC-content regions are



**Fig. 5.** DNA methylation analysis of non-CpG-island DNA on chromosomes 4, 10, 17, and X using the qAMP assay. Primers were designed to flank McrBC and HhaI or HpaII restriction sites placed at roughly 5-Mb intervals along each chromosome. Amplified regions were chosen only on the basis of the sequence not being in the proximity of a CpG island or a known 5' region, or within a repetitive sequence. (a) Analysis of chromosome 4. The percentage of DNA methylation of each amplified region in liver (black dash) and testis DNA (gray dash) is shown. Differences in methylation are shown by arrows. The ideogram and GC percentage are shown, as are the positions of hyper- and hypomethylated loci identified by using RLGs. (b) The average difference in percentage of methylation of all DM regions analyzed on each chromosome shows an overall decreased level of methylation in testis vs. liver in three of four chromosomes. Unchanged regions (<10% difference) were excluded. (c) The correlation between the difference in methylation in each region on all three chromosomes and the GC percent of flanking sequence shows a lower relative level of methylation in the testis in sequences of lower GC content. Chromosome 4 ideogram and percentage of GC were obtained from the University of California, Santa Cruz, CA, genome browser.



**Fig. 6.** Levels of DNA methylation in *Dnmt3L*-deficient spermatogonia and adult tissues. The percent methylation of DM RLGs loci that are hypermethylated in testis (a), hypomethylated in testis (b), and somatic-specific DM (c) was determined by using qAMP. Loci that are hypermethylated in the testis fail to gain normal levels of methylation in *Dnmt3L*-deficient spermatogonia. Analyzed loci were all in non-5' regions. L, liver; I, intestine; B, brain; T, testis.

more likely to be hypermethylated (Fig. 4b). In the testis, this relationship is inverted, illustrating that lower GC content regions are specifically hypomethylated.

Because of the GC-rich nature of the NotI recognition site (GCGGCCGC), RLGs sites are relatively rare in non-CpG-island DNA and are biased toward sequences of higher GC content. To avoid this bias and the potential influence of nearby genes and repeat elements on DNA methylation, we focused on HhaI, HpaII, and McrBC sites in non-5', non-CpG-island, and nonrepetitive sequences using the qAMP assay (25). A survey of 104 small amplified regions evenly spaced along chromosomes 4, 10, 17, and X in liver and testis shows striking differences in the pattern of DNA methylation between the two tissues (Fig. 5a and SI Fig. 10). Consistent with the RLGs findings, lower levels of methylation are generally found in the testis (Fig. 5b). In liver, non-CpG-island sequences are generally highly methylated in GC-poor regions, whereas lower levels are observed in some GC-rich regions. In the testis, the methylation status of these sites is again found to be inverted by using this alternate approach. A comparison of the regional GC content with the difference in methylation between the two tissues reveals a significant correlation between lower GC content and hypomethylation on these four chromosomes in the testis (Fig. 5c).

These data show that, by taking into account broad regional characteristics of chromosomes, we are able to explain a significant

proportion of the differences of DNA methylation that occur at single sites or small regions in non-CpG-island sequences at both a whole-chromosome and a genome-wide scale. Because these differences occur broadly, are more likely to occur away from regulatory regions of genes, and correlate with broad attributes of chromosomes, the possibility is raised that the testicular DNA methylation pattern may have a role in maintaining a unique chromosomal state that is capable of undergoing germ cell-specific processes, such as meiosis. The *Dnmt3L* gene is highly expressed during the perinatal period of reprogramming in male germ cells; yet, the primary defect that results from the disruption of the *Dnmt3L* gene is observed after birth in meiotic germ cells in which highly abnormal chromosomal structures are seen (2). If the testis-specific methylation patterns we describe for non-CpG-island sequences have a role in chromosome structure or meiosis in germ cells, we would predict that these patterns might be altered in the germ cells of *Dnmt3L*<sup>-/-</sup> mice. To address this issue, spermatogonia were isolated from *Dnmt3L*<sup>-/-</sup> mice, and the methylation levels at various DM loci determined by RLGs were analyzed by using qAMP (Fig. 6). In the spermatogonia from *Dnmt3L*<sup>-/-</sup> mice, all loci examined that were hypermethylated in the testis failed to gain the normal levels of methylation found in wild-type spermatogonia and testis. Other DM loci are also hypomethylated. Testis-specific hypomethylated loci remained normal, indicating that the primary influence of *Dnmt3L* is to promote DNA methylation. Further-

more, levels of methylation in wild-type spermatogonia resemble the levels found in the adult testis, demonstrating that the testis-specific pattern was acquired in early germ cell development at these non-CpG-island sequences.

We have demonstrated the finding that male germ cells have highly unique patterns of DNA methylation, with most of the hypomethylation found in nonrepetitive, non-CpG-island sequences. These hypomethylated sequences do not correlate with expression patterns, which is consistent with the fact that they are generally not found in the regulatory regions of genes. Instead, they are correlated with regional chromosome features, such as banding patterns and GC content. Although another study inferred a paucity of non-CpG-island hypomethylation in germ cells (8), experimental evidence in the present study indicates a prominent role for regulation of DNA methylation in this compartment of the testis genome. Because greater than half of the total genomic sequence and CpG dinucleotide content belongs to nonrepetitive, non-CpG-island DNA (26), these modifications might be expected to have prevalent effects on the overall structure of chromosomes in the testis. We also find that in the spermatogonia of *Dnmt3L*<sup>-/-</sup> mice, a model that displays abnormal chromosomal structures, at the sites examined, the unique testicular DNA methylation pattern is not established. We propose that the maintenance of a unique, germ cell-specific chromosomal structure may require a distinct configuration of DNA methylation patterns in non-CpG-island DNA.

## Materials and Methods

**RLGS and Spot Identification.** Adult male C57BL/6 mice were from Charles River Laboratories (St-Constant, QC, Canada). Genomic DNA was isolated from tissues by using proteinase K, followed by phenol extraction (27). RLGS was done as described (27). Three replicate RLGS profiles were generated for each tissue. Each spot was assessed a score between 0 and 4 based on its intensity relative to surrounding, fully unmethylated spots. Intensity scores of 0, 1, 2, 3, and 4 are representative of 100, 75, 50, 25, and 0% methylation, respectively. Visual assessment of spot intensity was confirmed by densitometry. All DM spots were observed to be changed in all three replicates. Spot identities were determined by using either the RLGS cloning library (16) or the virtual RLGS method (D.J.S., R. Wenger, and C. Plass, unpublished work). Loci identified by using virtual RLGS were confirmed by obtaining the corresponding BAC clone (Roswell Park Microarray Core Facility, Buffalo, NY) and

running RLGS mixing gels. The methylation status of 36 identified loci was confirmed by using the qAMP methylation assay (25). Briefly, DNA is digested with various methylation-sensitive and methylation-dependent restriction enzymes and amplified by using real-time PCR. Changes in cycle threshold values were used to calculate the percentage of methylation at each restriction enzyme site. For the analysis of chromosomes 4, 10, 17, and X, the percentage of methylation of each locus was determined by averaging the values obtained for all of the restriction enzyme sites analyzed within the amplified region. To obtain purified *Dnmt3L*-deficient spermatogonia, heterozygous *Dnmt3L*-deficient mice (28) were crossed with GOF18deltaPE-Oct4/GFP mice (29), and spermatogonia were isolated from the testes of 6-dpp offspring by using fluorescence-activated cell sorting; 3–5 mice were pooled in each group. All primers are listed in SI Tables 3 and 4.

**Data Analysis.** All positional and sequence information was obtained from the University of California, Santa Cruz, CA, genome browser (30) Mouse August 2005 assembly, mm7 (<http://genome.ucsc.edu/>). Chromosomal bands were defined as a G-band if they appeared to be black or gray and as an R-band if they appeared to be white on ideograms. CpG islands were defined by the criteria of Takai and Jones (31) ( $G+C \geq 55\%$ , CpGobs/exp  $\geq 0.6$ ,  $\geq 500$  bp), 5' regions were defined as being within -3 to +1 kb of the transcriptional start site. DM loci were divided into the subcategories of "somatic-specific DM" and "testis-specific DM" if the unique state of hyper- or hypomethylation occurred solely in a single somatic tissue or in the testis, respectively. SigmaStat v3.0 software (SPSS, Chicago, IL) was used for statistical analysis, with  $P < 0.05$  considered to be significant. Gene expression data were obtained from the GNF SymAtlas v1.2.4 expression database (<http://symatlas.gnf.org/SymAtlas/>) by using the GNF1M gcRNA data set.

We thank T. Bestor (Columbia University, New York, NY) for the gift of the *Dnmt3L*<sup>-/-</sup> mice and H. Scholer (Max Planck Institute for Molecular Biomedicine, Münster, Germany) for the gift of the GOF18deltaPE-Oct4/GFP mice. C.C.O. and S.L. are recipients of Canadian Institutes of Health Research (CIHR) Doctoral Research Awards and Montreal Children's Hospital Studentship Awards. J.M.T. is a William Dawson Scholar of McGill University and a Scholar of the Fonds des Recherches en Santé du Québec. B.R. is a James McGill Professor of McGill University. This work was supported by funding from the CIHR.

- Reik W, Dean W, Walter J (2001) *Science* 293:1089–1093.
- Bourc'his D, Bestor TH (2004) *Nature* 431:96–99.
- Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) *Nature* 429:900–903.
- Hayashi K, Yoshida K, Matsui Y (2005) *Nature* 438:374–378.
- Ferguson-Smith AC, Surani MA (2001) *Science* 293:1086–1089.
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) *Nat Genet* 23:314–318.
- Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E (2003) *Proc Natl Acad Sci USA* 100:2538–2543.
- Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J, Bestor TH (2006) *Genome Res* 16:157–163.
- La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH, Trasler JM (2004) *Dev Biol* 268:403–415.
- La Salle S, Trasler JM (2006) in *The Sperm Cell: Production, Maturation, Fertilization, Regeneration*, eds De Jonge C, Barratt C (Cambridge Univ Press, New York), 279–322.
- Bellvé AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, Dym M (1977) *J Cell Biol* 74:68–85.
- Sanford J, Forrester L, Chapman V, Chandley A, Hastie N (1984) *Nucleic Acids Res* 12:2823–2836.
- Maclean JA, Wilkinson MF (2005) *Curr Top Dev Biol* 71:131–197.
- Shiota K, Kogo Y, Ohgane J, Imamura T, Urano A, Nishino K, Tanaka S, Hattori N (2002) *Genes Cells* 7:961–969.
- Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA (2005) *Proc Natl Acad Sci USA* 102:3336–3341.
- Yu L, Liu C, Bennett K, Wu YZ, Dai Z, Vandeusen J, Opavsky R, Raval A, Trikha P, Rodriguez B, et al. (2004) *Genomics* 84:647–660.
- Fazzari MJ, Grealley JM (2004) *Nat Rev Genet* 5:446–455.
- Aota S, Gojobori T, Shigesada K, Ozeki H, Ikemura T (1987) *Gene* 56:1–12.
- Baust C, Gagnier L, Baillie GJ, Harris MJ, Juriloff DM, Mager DL (2003) *J Virol* 77:11448–11458.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) *Cytogenet Genome Res* 110:462–467.
- Peaston AE, Evsikov AV, Graber JH, De Vries WN, Holbrook AE, Solter D, Knowles BB (2004) *Dev Cell* 7:597–606.
- Shapiro JA (2005) *BioEssays* 27:122–125.
- Shima JE, McLean DJ, McCarrey JR, Griswold MD (2004) *Biol Reprod* 71:319–330.
- Bernardi G, Olofsson B, Filipki J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985) *Science* 228:953–958.
- Oakes CC, La Salle S, Robaire B, Trasler JM (2006) *Epigenetics*, 1:146–152.
- Fazzari MJ, Grealley JM (2004) *Nat Rev Genet* 5:446–455.
- Okazaki Y, Okuzumi H, Sasaki N, Ohsumi T, Kuromitsu J, Hirota N, Muramatsu M, Hayashizaki Y (1995) *Electrophoresis* 16:197–202.
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) *Science* 294:2536–2539.
- Yoshimizu T, Sugiyama N, De FM, Yeom YI, Ohbo K, Masuko K, Obinata M, Abe K, Scholer HR, Matsui Y (1999) *Dev Growth Differ* 41:675–684.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) *Genome Res* 12:996–1006.
- Takai D, Jones PA (2002) *Proc Natl Acad Sci USA* 99:3740–3745.