encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81,** 611–620 (1995).
11. Pelletier, L. *et al.* The Caenorhabditis elegans centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Curr. Biol.* **14,** 863–873 (2004).
12. Hannak, E. *et al.* The kinetically dominant assembly pathway for centrosomal asters in *Caenorhabditis elegans* is gamma-tubulin dependent. *J. Cell Biol.* **157,** 591–602 (2002).
13. Albertson, D. G. Formation of the first cleavage spindle in nematode embryos. *Dev. Biol.* **101,** 61–72 (1984).
14. Strome, S. *et al.* Spindle dynamics and the role of gamma-tubulin in early *Caenorhabditis elegans* embryos. *Mol. Biol. Cell* **12,** 1751–1764 (2001).
15. Kemp, C. A., Kopish, K. R., Zipperlen, P., Ahringer, J. & O'Connell, K. F. Centrosome maturation and duplication in *C. elegans* require the coiled-coil protein SPD-2. *Dev. Cell* **6,** 511–523 (2004).
16. Stevenson, V. A., Kramer, J., Kuhn, J. & Theurkauf, W. E. Centrosomes and the Scrambled protein coordinate microtubule-independent actin reorganization. *Nature Cell Biol.* **3,** 68–75 (2001).
17. Berdnik, D. & Knoblich, J. A. *Drosophila* Aurora-A is required for centrosome maturation and actin-dependent asymmetric protein localization during mitosis. *Curr. Biol.* **12,** 640–647 (2002).
18. Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291,** 1550–1553 (2001).
19. Gonczy, P. *et al.* Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144,** 927–946 (1999).
20. Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. A. Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* **153,** 1209–1226 (2001).
21. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157,** 1217–1226 (2001).
22. Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S. & Hyman, A. A. SAS-4 is a *C. elegans* centriolar protein that controls centrosome size. *Cell* **112,** 575–587 (2003).
23. Grill, S. W., Howard, J., Schaffer, E., Stelzer, E. H. & Hyman, A. A. The distribution of active force generators controls mitotic spindle position. *Science* **301,** 518–521 (2003).
24. Pichler, S. *et al.* OOC-3, a novel putative transmembrane protein required for establishment of cortical domains and spindle orientation in the P(1) blastomere of *C. elegans* embryos. *Development* **127,** 2063–2073 (2000).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank: A. Desai, S. Eaton, C. Hoege, K. Oegema, N. Özlü, L. Pelletier, S. Quintin, A. Schlaitz, S. Schonegg, M. Srayko, J. Stear, A. Tudor-Constantinescu and M. van Breugel for comments on the manuscript; S. Grill for advice on laser ablation; A. Pozniakovsky for modified GFP vectors; and G. Seydoux for the gift of JH1380 worms. Some of the worm strains used in this study were obtained from the Caenorhabditis Genetics Center, funded by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.C. (cowan@mpi-cbg.de).

Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L

Déborah Bourc'his & Timothy H. Bestor

Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 West 168th Street, New York, New York 10032, USA

Mammalian genomes employ heritable cytosine methylation in the long-term silencing of retrotransposons and genes subject to genomic imprinting and X chromosome inactivation. Little is known of the mechanisms that direct cytosine methylation to specific sequences. Here we show that DNA methyltransferase 3-like (Dnmt3L (ref. 1)) is expressed in testes during a brief perinatal period in the non-dividing precursors of spermatogonial stem cells at a stage where retrotransposons undergo *de novo* methylation. Deletion of the *Dnmt3L* gene prevented the *de novo* methylation of both long-terminal-repeat (LTR) and non-LTR retrotransposons, which were transcribed at high levels in spermatogonia and spermatocytes. Loss of Dnmt3L from early germ cells also caused meiotic failure in spermatocytes, which do not express Dnmt3L. Whereas dispersed repeated sequences were demethylated in mutant germ cells, tandem repeats in pericentric regions were methylated normally. This result indicates that the Dnmt3L protein might have a function in the *de novo* methylation of dispersed repeated sequences in a premeiotic genome scanning process that occurs in male germ cells at about the time of birth.

Lifelong gene silencing is imposed on target sequences by *de novo* methylation in germ cells and early embryos. The only factors known to be involved in the establishment of methylation patterns in germ cells are Dnmt3A (ref. 2) and Dnmt3L, a protein that lacks the catalytic motifs that characterize the DNA cytosine-5-methyltransferases but is related to the active DNA methyltransferases Dnmt3A and Dnmt3B in framework regions[1]. Deletion of *Dnmt3L* does not prevent oogenesis, but the heterozygous offspring of homozygous mutant females die before mid-gestation as a result of biallelic expression of imprinted genes normally methylated and silenced on the allele of maternal origin[3]. Male mice that lack Dnmt3L are viable but sterile, with a complete absence of germ cells in adult males[3].

As shown in Fig. 1, expression of Dnmt3L is first seen in non-dividing prospermatogonia after 12.5 days post coitus (d.p.c.) and is highest at about the time of birth; expression declines rapidly after birth and is extinguished by 6 days post partum (d.p.p.), when most prospermatogonia have differentiated into dividing spermatogonial stem cells. Other data show that the decline in Dnmt3L expression over this period is more than 200-fold[4]. Dnmt3L shows striking sexual dimorphism in expression patterns; it is present in females only in growing oocytes, which are arrested in the diplotene stage of meiosis I, but in males it is restricted to diploid prospermatogonia, which differentiate into spermatogonia and undergo many mitotic divisions before entry into meiosis.

The germ cell population of testes lacking Dnmt3L was normal at birth, but only spermatogonia and leptotene and zygotene spermatocytes were seen in testes of young adult males; progressive loss of spermatogonia caused complete azoospermia in older mutant animals. Examination of meiotic chromosome spreads stained with antibodies against synaptonemal complex proteins 1 and 3 (Scp1 and Scp3 (ref. 5)) showed widespread nonhomologous synapsis with branching and anastomosing arrays of synaptonemal complex proteins in most leptotene and zygotene spermatocytes (compare control in Fig. 2a to mutant in Fig. 2b). Some leptotene spermatocytes had complexes of synaptonemal proteins in the form of interlocked rings and other highly abnormal structures (Fig. 2c, d). All spermatocytes showed asynapsis or abnormal synapsis, and none progressed to the full pachytene stage, as assessed by the

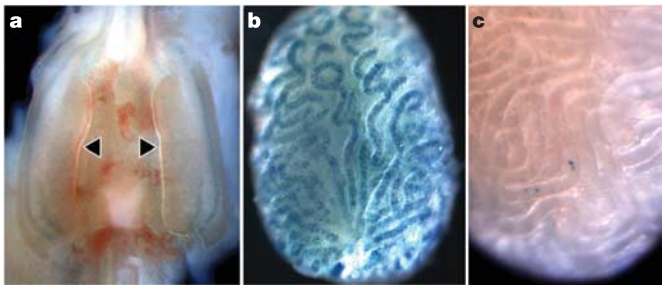

Figure 1 Expression of Dnmt3L in perinatal male germ cells. Dnmt3L transcription was revealed by staining whole testes for β-galactosidase in animals heterozygous for the deletion–*β-geo* insertion allele of *Dnmt3L* described in ref. 3. **a**, Lack of *Dnmt3L* expression in testes (black arrows) at 12.5 d.p.c. **b**, High-level expression of *Dnmt3L* in germ cells at 2 d.p.p.; the β-galactosidase staining visible in the seminiferous tubules was confirmed to be in prospermatogonia by examination of sectioned specimens. **c**, Loss of *Dnmt3L* expression in germ cells in seminiferous tubules of 6 d.p.p. testis.

lack of pachytene-specific histone H1t and of sex bodies after staining with antibodies against histone γ-H2AX (data not shown). The meiotic abnormalities do not indicate a role for Dnmt3L during normal chromosome synapsis, because the protein is not expressed in meiotic cells, and at the end of the reproductive life span normal meiotic male germ cells can be more than 2 years distant from exposure to Dnmt3L.

The loss of post-zygotene spermatocytes from Dnmt3L-deficient males is likely to arise through the activation of an apoptotic checkpoint triggered before pachynema by non-synapsed chromosomal regions[6], but the death of spermatogonia in adult males implies that a heritable epigenetic defect is incurred in the perinatal prospermatogonia while under Dnmt3L deprivation and is transmitted by mitotic inheritance during the division of spermatogonia. To test for DNA methylation abnormalities that might underlie the heritable defect, germ cells were isolated from dissociated testes by fluorescence-activated cell sorting after labelling with antibodies against germ cell nuclear antigen-1 (GCNA1 (ref. 7)), and DNA was purified and tested for resistance to the methylation-sensitive restriction endonuclease *Hpa*II. As shown in Fig. 3a, DNA of mutant germ cells was less methylated than DNA from wild-type germ cells or mutant somatic cells from testis. The DNA blot hybridization of Fig. 3b shows that the promoter regions of LINE-1 retrotransposons were largely demethylated, as were LTR sequences of retrotransposons of the intracisternal A particle (IAP) class (Fig. 3c). These findings are consistent with the overall demethylation shown in Fig. 3a, because retrotransposons and their remnants contain the large majority of the 5-methylcytosine present in the mammalian genome[8]. These data show that dispersed repeats within the euchromatic compartment of the genome require Dnmt3L for methylation, but the pericentric tandem repeat DNAs known as major and minor satellites are not dependent on Dnmt3L

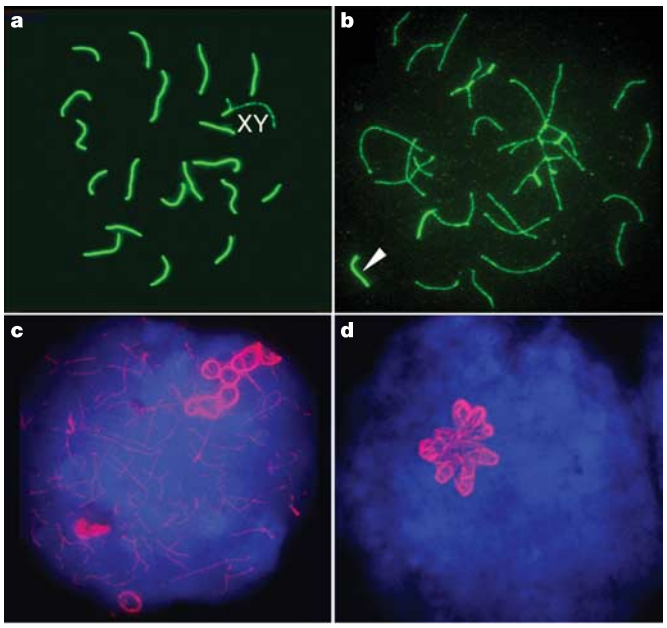

Figure 2 Meiotic catastrophe in spermatocytes derived from Dnmt3L-deficient prospermatogonia. **a**, Normal meiotic synapsis in wild-type spermatocytes. Note the formation of complete linear synaptonemal complexes except for the XY chromosome pair, in which synapsis is restricted to the pseudoautosomal regions. **b**, Branching and anastomosing synaptonemal complexes in Dnmt3L-deficient spermatocytes. Nearly all the chromosomal regions are unpaired or engaged in non-homologous synapsis; a single apparently normal synaptonemal complex is indicated by the white arrowhead. **c**, **d**, Formation of highly aberrant complexes of synaptonemal proteins in the form of interlocked rings (**c**) and complex three-dimensional structures (**d**) in *Dnmt3L* mutant spermatocytes. Similar staining patterns were seen after labelling with antibodies against Scp3 or a combination of Scp1 and Scp3 antibodies. Meiotic spreads were prepared as described[22]. Synaptonemal complex proteins are stained in green in **a** and **b**, and in red in **c** and **d**.

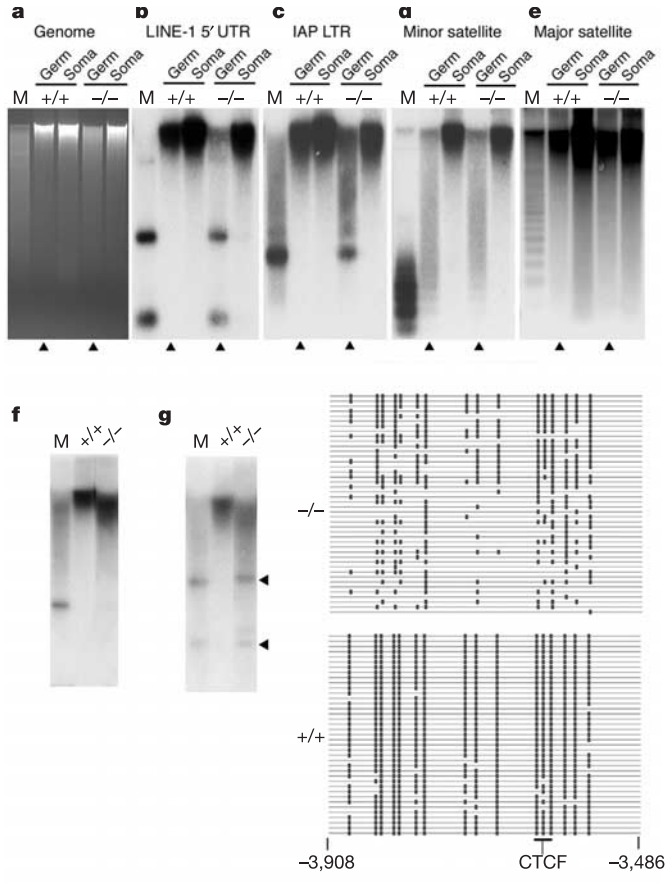

Figure 3 Abnormal genome methylation in Dnmt3L-deficient male germ cells. Germ cells from 17 d.p.p. testes were purified by fluorescence-activated cell sorting after being stained for GCNA1. DNA was digested with the methylation-sensitive restriction endonuclease *Hpa*II before DNA blot analysis with the indicated probes. Lanes headed M contained DNA that had been cleaved with *Msp*I, a methylation-insensitive isoschizomer of *Hpa*II. Soma, somatic cells from testis. **a**, Demethylation of mutant germ-cell genomes revealed by staining DNA with ethidium bromide; arrowheads at the bottom indicate wild-type and Dnmt3L mutant germ cells. **b**, **c**, Demethylation of the promoter region of A-type LINE-1 elements (**b**) and of LTR sequences of IAP elements (**c**) in the absence of Dnmt3L. **d**, **e**, Equal methylation of minor (**d**) and major (**e**) satellites in Dnmt3L-deficient germ cells; satellite DNA in normal male germ cells is not fully methylated[23,24]. The gel shown in **a** was blotted to nylon and the blot was stripped and reprobed to provide the data for **b**–**e**. **f**, Dnmt3L-independent methylation of the intergenic differentially methylated region of the *Dlk1–Gtl2*-imprinted cluster. **g**, Partial demethylation of the *H19* differentially methylated region in Dnmt3L-deficient male germ cells. The DNA blot hybridization at the left shows partial demethylation at *Hpa*II sites between −766 and −3,994 bases 5′ of the *H19* start site. The bisulphite genomic sequencing data at the right show partial demethylation of CpG sites within the region −3,486 to −3,908 base pairs 5′ of the H19 start site that contains a CTCF-binding site. A total of 39 bisulphite-converted DNA molecules were sequenced for the control and 45 for the Dnmt3L mutant. The 5′ LTR IAP probe was as described[13], and the LINE-1 5′ untranslated-region probe was a PCR product that spanned the promoter region of a type A LINE-1 element from nucleotides 515–1,628 in GenBank accession no. M13002. The *Dlk1–Gtl2* DMR probe covered positions 82,879–83,729 in GenBank accession no. AG320506. The *H19* DMR probe spanned the 5′ region from position −766 to −3,994 upstream of the *H19* transcription start site in GenBank accession no. M19619. The major satellite probe was from plasmid pMR196 (ref. 25), and minor satellite probe was an end-labelled oligonucleotide of sequence 5′-AACAGTGTATATCAATGAGTTACAATGAG-3′.

and their methylation status was equivalent in wild-type and mutant germ cells (Fig. 3d, e). This indicates the independent regulation of *de novo* methylation of dispersed and tandem repeated sequences. Figure 3f shows that the intergenic imprinting control region (the differentially methylated region (DMR)) of the *Dlk1–Gtl2*-imprinted cluster[9] was not detectably demethylated in Dnmt3L-deficient germ cells, and there was only partial (about 50%) demethylation of the DMR of the *H19* gene[10,11] (Fig. 3g). This is in contrast to the phenotype of Dnmt3L-deficient females, in which meiosis and the methylation of repeated sequences are normal and obvious methylation defects are limited to single-copy sequences associated with maternally imprinted genes[3].

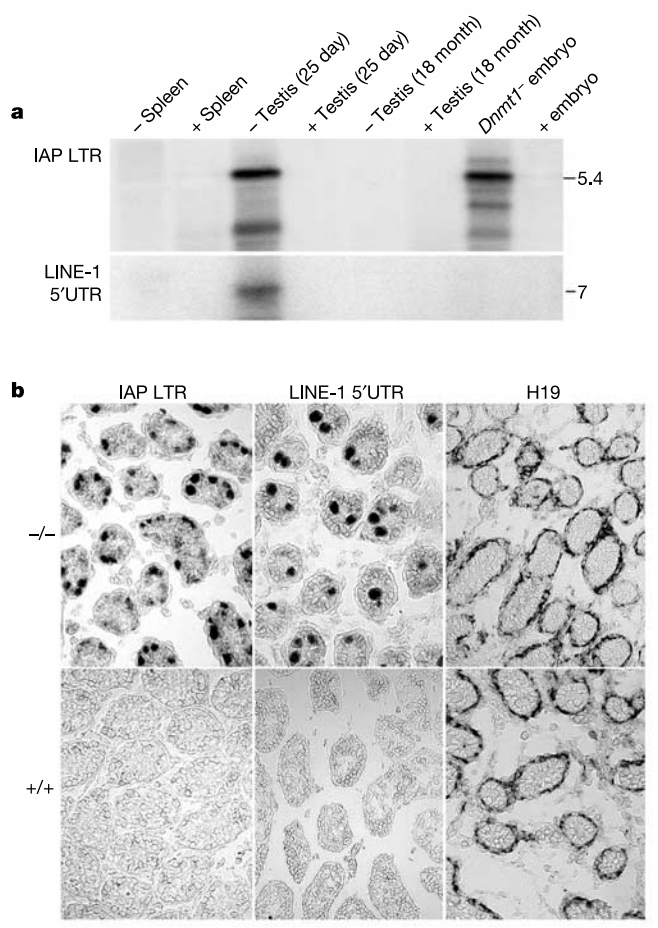

Figure 4 Transcriptional reactivation of retrotransposons in Dnmt3L-deficient germ cells. **a**, High-level expression of retrotransposons in Dnmt3L-deficient testis by RNA blot hybridization assay. Both LINE-1 and IAP transcripts are present at high concentrations in *Dnmt3L* mutant testis (third lane), whereas somatic cells of *Dnmt1*$^{n/n}$ embryos at 9.5 d.p.c. (second lane from right) reactivate only IAP elements[9]. The predominant IAP subtype is the active IΔ1 type at 5.4 kilobases (ref. 13). At 18 months, testes contain no detectable germ cells and there are no detectable retrotransposon transcripts (lane 5). **b**, Specific expression of LINE-1 and IAP transcripts in Dnmt3L-deficient germ cells. *In situ* hybridization against mutant (top row) and wild-type (bottom row) sections of testes from mice at 2 d.p.p. showed expression of high concentrations of LINE-1 and IAP transcripts in prospermatogonia. Further analysis showed expression of LINE-1 and IAP transposons in dividing spermatogonia and spermatocytes as well. The imprinted *H19* gene was expressed at similar levels in mutant and control testes only in peritubular myoid cells, which confirms that the partial demethylation shown in Fig. 3g was not sufficient to reactivate *H19* expression in *Dnmt3L* mutant germ cells. Probes in both **a** and **b** were as in the DNA blot hybridization of Fig. 3.

The host defence hypothesis[8,12] predicts that demethylation will reactivate retrotransposons in germ cells, as occurs in somatic cells whose genomes have become partly demethylated as a result of mutations in the *Dnmt1* gene[13]. RNA blot hybridization of RNA from whole testis showed that transcripts of IAP (reviewed in ref. 14) and LINE-1 (reviewed in ref. 15) retrotransposons were present in *Dnmt3L* mutant testes but undetectable in wild-type testes (Fig. 4a). Full-length transcripts of IΔ1 IAP elements (5.4 kilobases) and A-type LINE-1 elements (about 7 kilobases), which are active retrotransposons, were present at high concentrations. *In situ* RNA hybridization revealed that IAP and LINE-1 retrotransposon transcripts were present at high concentrations only in mutant germ cells as early as 2 d.p.p., when the first spermatogonia begin to divide, and were present in all mutant spermatogonia and spermatocytes (Fig. 4b). Reactivation of retrotransposons in demethylated germ-cell genomes confirmed that cytosine methylation has an essential role in the suppression of retrotransposons in the mammalian germ line[8,12]. Figure 4b also shows that the partial demethylation observed at the DMR of *H19* (Fig. 3) was not sufficient to reactivate this imprinted gene in germ cells, and expression of *H19* was confined to peritubular myoid cells in both mutant and control specimens. A more pronounced demethylation of *H19* in Dnmt3L-deficient male germ cells at earlier stages than those tested here was reported recently[16].

The methylation defect in Dnmt3L-deficient germ cells is not due to a general defect in the *de novo* methylation of repeated sequences, as shown by the normal methylation of the tandem repeats of satellite DNA (which are largely free of retrotransposons); the defect is instead seen specifically at retrotransposons, which are interspersed repeated sequences. A major difference between populations of tandem and interspersed repeats is the abundance of homology–heterology boundaries in the latter; these boundaries have been proposed to be marks by which retrotransposons could be identified by the host[17]. Under this model, Dnmt3L is involved in a premeiotic genome scan that is proposed to occur in prospermatogonia within a brief perinatal period that establishes the lifelong silencing of retrotransposons. Silencing is proposed to involve heritable cytosine methylation mediated by the Dnmt3L-dependent recruitment and activation of Dnmt3A (refs 2, 18) and other silencing factors at homology–heterology boundaries. As the genome is scanned against itself, the frequency of contacts between dispersed repeats will follow first-order kinetics and will therefore increase as the second power of copy number. The premeiotic genome scan proposed here differs from a meiotic homology search in that the former identifies non-allelic repeats whereas the latter aligns homologous chromosomes. The postulated premeiotic genome scan is likely to occur without double-strand breaks or the invasion of duplex DNA by free DNA ends and is not therefore directly comparable to current models for homology search in meiosis. The most similar phenomena might be repeat-induced point mutation (RIP) in *Neurospora crassa*[19], in which interactions between repeated sequences cause their mutual methylation and silencing. As in retrotransposon and imprinted gene methylation in mouse prospermatogonia, RIP occurs in the sexual cycle before the start of meiosis and is not known to share mechanisms with meiotic recombination.

Dnmt3L is required for normal male meiosis but is not expressed in spermatocytes, and the heritable defect in Dnmt3L-deficient male germ cells is likely to arise from a failure to establish methylation patterns on dispersed repeated sequences in perinatal prospermatogonia. The non-homologous synapsis that characterizes 1meiotic prophase in Dnmt3L-deficient spermatocytes could have several sources: it might result from a perturbation of gene expression, from illegitimate interactions between dispersed repeated sequences that were unmasked by demethylation[20], from single-strand or double-strand breaks produced during replicative retrotransposition or from an increased number of double-strand

breaks induced at ectopic transcription units[21] activated by demethylation.

The data presented here show that Dnmt3L is required for heritable silencing of retrotransposons in male germ cells. Temporary deprivation of Dnmt3L through drug-induced conditional alleles of the *Dnmt3L* gene might allow the controlled mobilization of endogenous retrotransposons without meiotic catastrophe and could form the basis of a system of retrotransposon insertional mutagenesis that would be useful in forward genetic screens in the mouse. □

Received 22 May; accepted 26 July 2004; doi:10.1038/nature02886.
Published online 18 August 2004.

1. Aapola, U. *et al.* Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65,** 293–298 (2000).
2. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. *Development* **129,** 1983–1993 (2002).
3. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536–2539 (2001).
4. La Salle, S. *et al.* Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev. Biol.* **268,** 403–415 (2004).
5. Dobson, M. J., Pearlman, R. E., Karaiskakis, A., Spyropoulos, B. & Moens, P. B. Synaptonemal complex proteins: occurrence, epitope mapping and chromosome disjunction. *J. Cell Sci.* **107,** 2749–2760 (1994).
6. Odorisio, T., Rodriguez, T. A., Evans, E. P., Clarke, A. R. & Burgoyne, P. S. Meiotic checkpoint monitoring synapsis eliminates spermatocytes via p53-independent apoptosis. *Nature Genet.* **18,** 257–261 (1998).
7. Enders, G. C. & May, J. J. Developmentally regulated expression of a mouse germ cell nuclear antigen examined from embryonic day 11 to adult in male and female mice. *Dev. Biol.* **163,** 331–340 (1994).
8. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13,** 335–340 (1997).
9. Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1–Gtl2* imprinted cluster on mouse chromosome 12. *Nature Genet.* **35,** 97–102 (2003).
10. Davis, T. L., Trasler, J. M., Moss, S. B., Yang, G. J. & Bartolomei, M. S. Acquisition of the *H19* methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* **58,** 18–28 (1999).
11. Ueda, T. *et al.* The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* **5,** 649–659 (2000).
12. Bestor, T. H. Cytosine methylation mediates sexual conflict. *Trends Genet.* **19,** 185–190 (2003).
13. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20,** 116–117 (1998).
14. Kuff, E. L. & Lueders, K. K. The intracisternal A-particle gene family: structure and functional aspects. *Adv. Cancer Res.* **51,** 183–276 (1988).
15. Ostertag, E. M. & Kazazian, H. H. Jr Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35,** 501–538 (2001).
16. Kaneda, M. *et al.* Essential role for *de novo* methyltransferase 3a in paternal and maternal imprinting. *Nature* **429,** 900–903 (2004).
17. Bestor, T. H. & Tycko, B. Creation of genomic methylation patterns. *Nature Genet.* **12,** 363–367 (1996).
18. Chedin, F., Lieber, M. R. & Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl Acad. Sci. USA* **99,** 16916–16921 (2002).
19. Cambareri, E. B., Jensen, B. C., Schabtach, E. & Selker, E. U. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244,** 1571–1575 (1989).
20. Maloisel, L. & Rossignol, J.-L. Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev.* **12,** 1381–1389 (1998).
21. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nature Rev. Genet.* **2,** 360–369 (2001).
22. Peters, A. H., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res.* **5,** 66–68 (1997).
23. Ponzetto-Zimmerman, C. & Wolgemuth, D. J. Methylation of satellite sequences in mouse spermatogenic and somatic DNAs. *Nucleic Acids Res.* **12,** 2807–2822 (1984).
24. Sanford, J., Forrester, L. & Chapman, V. Methylation patterns of repetitive DNA sequences in germ cells of *Mus musculus*. *Nucleic Acids Res.* **12,** 2822–2835 (1984).
25. Pietras, D. F. *et al.* Construction of a small *Mus musculus* repetitive DNA library: identification of a new satellite sequence in *Mus musculus*. *Nucleic Acids Res.* **11,** 6965–6983 (1983).

Acknowledgements We thank B. Tycko for suggestions, P. Moens for antibodies against synaptonemal complex proteins and H1t and for discussions, G. Enders for antibodies against GCNA1, J. Walonoski for advice on *in situ* hybridization, and K. Anderson, M. Damelin and M. Goll for criticism of the manuscript. This work was supported by grants from the National Institutes of Health to T.H.B. and by a fellowship from the Rett Syndrome Research Foundation to D.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H.B (thb12@columbia.edu).
Dnmt3L^{tm1Bes} mutant mice are available from the Jackson Labs (http://jaxmice.jax.org/).

Transcriptional regulatory code of a eukaryotic genome

Christopher T. Harbison[1,2]*, D. Benjamin Gordon[1]*, Tong Ihn Lee[1], Nicola J. Rinaldi[1,2], Kenzie D. Macisaac[3], Timothy W. Danford[3], Nancy M. Hannett[1], Jean-Bosco Tagne[1], David B. Reynolds[1], Jane Yoo[1], Ezra G. Jennings[1], Julia Zeitlinger[1], Dmitry K. Pokholok[1], Manolis Kellis[1,3,4], P. Alex Rolfe[3], Ken T. Takusagawa[3], Eric S. Lander[1,2,4], David K. Gifford[3,4], Ernest Fraenkel[1,3] & Richard A. Young[1,2,4]

[1]*Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA*
[2]*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA*
[4]*Broad Institute, One Kendall Square, Building 300, Cambridge, Massachusetts 02139, USA*

* These authors contributed equally to this work

DNA-binding transcriptional regulators interpret the genome's regulatory code by binding to specific sequences to induce or repress gene expression[1]. Comparative genomics has recently been used to identify potential *cis*-regulatory sequences within the yeast genome on the basis of phylogenetic conservation[2–6], but this information alone does not reveal if or when transcriptional regulators occupy these binding sites. We have constructed an initial map of yeast's transcriptional regulatory code by identifying the sequence elements that are bound by regulators under various conditions and that are conserved among *Saccharomyces* species. The organization of regulatory elements in promoters and the environment-dependent use of these elements by regulators are discussed. We find that environment-specific use of regulatory elements predicts mechanistic models for the function of a large population of yeast's transcriptional regulators.

We used genome-wide location analysis[7–10] to determine the genomic occupancy of 203 DNA-binding transcriptional regulators in rich media conditions and, for 84 of these regulators, in at least 1 of 12 other environmental conditions (Supplementary Table 1, Supplementary Fig. 1; http://web.wi.mit.edu/young/regulatory_code). These 203 proteins are likely to include nearly all of the DNA-binding transcriptional regulators encoded in the yeast genome. Regulators were selected for profiling in an additional environment if they were essential for growth in that environment or if there was other evidence implicating them in the regulation of gene expression in that environment. The genome-wide location data identified 11,000 unique interactions between regulators and promoter regions at high confidence ($P \leq 0.001$).

To identify the *cis*-regulatory sequences that are likely to serve as recognition sites for transcriptional regulators, we merged information from genome-wide location data, phylogenetically conserved sequences, and prior knowledge (Fig. 1a). We used six motif discovery methods[11–13] to discover 68,279 DNA sequence motifs for the 147 regulators that bound more than ten probes (Supplementary Methods, Supplementary Fig. 2). From these motifs we derived the most likely specificity for each regulator through clustering and stringent statistical tests. This motif discovery process identified highly significant ($P \leq 0.001$) motifs for each of 116 regulators. We determined a single high-confidence motif for 65 of these regulators by using additional criteria including the requirement for conservation across three of four related yeast species. Examples of discovered and rediscovered motifs are depicted in Fig. 1b, and comparisons of the discovered motifs with those described previously are shown in Supplementary Table 2.