D54–D58 *Nucleic Acids Research, 2015, Vol. 43, Database issue* *Published online 07 October 2014*
doi: 10.1093/nar/gku920

MethBank: a database integrating next-generation sequencing single-base-resolution DNA methylation programming data

Dong Zou[1,†], Shixiang Sun[1,2,†], Rujiao Li[3], Jiang Liu[1], Jing Zhang[1,*] and Zhang Zhang[1,*]

[1]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [2]University of Chinese Academy of Sciences, Beijing 100049, China and [3]Core Genomic Facility, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

Received August 12, 2014; Revised September 09, 2014; Accepted September 22, 2014

ABSTRACT

DNA methylation plays crucial roles during embryonic development. Here we present MethBank (http://dnamethylome.org), a DNA methylome programming database that integrates the genome-wide single-base nucleotide methylomes of gametes and early embryos in different model organisms. Unlike extant relevant databases, MethBank incorporates the whole-genome single-base-resolution methylomes of gametes and early embryos at multiple different developmental stages in zebrafish and mouse. MethBank allows users to retrieve methylation levels, differentially methylated regions, CpG islands, gene expression profiles and genetic polymorphisms for a specific gene or genomic region. Moreover, it offers a methylome browser that is capable of visualizing high-resolution DNA methylation profiles as well as other related data in an interactive manner and thus is of great helpfulness for users to investigate methylation patterns and changes of gametes and early embryos at different developmental stages. Ongoing efforts are focused on incorporation of methylomes and related data from other organisms. Together, MethBank features integration and visualization of high-resolution DNA methylation data as well as other related data, enabling identification of potential DNA methylation signatures in different developmental stages and accordingly providing an important resource for the epigenetic and developmental studies.

INTRODUCTION

DNA methylation is a major epigenetic mark that is crucial for embryogenesis and highly dynamic during embryonic development (1,2). According to our previous studies (3,4), it is found that, in *Danio rerio* (zebrafish), paternal methylome is discovered to be stably inherited, while maternal methylome is reprogrammed to the sperm pattern (3,5). In *Mus musculus* (mouse), the paternal methylome and at least a significant proportion of maternal methylome go through active demethylation during embryonic development (4). The strategies for reprogramming parental methylomes are fundamentally different between vertebrates (6), suggesting that the underlying developmental programs may be distinct in different species (3,4,7–10). Therefore, studying DNA methylation at multiple developmental stages in different species may extend the knowledge of the inheritance and reprogramming of methylomes.

In contrast to region-wide methods for DNA methylation profiling and reduced representation bisulfite sequencing (RRBS), whole-genome bisulfite sequencing (WGBS) powered with high-throughout sequencing technologies enables the single-base nucleotide measurement of DNA methylation and accordingly the generation of genome-wide high-resolution DNA methylome. The comprehensive integration of DNA methylomes in different species, therefore, bears promises to help us systemically study DNA methylation reprogramming in early embryos at full aspects (10–14). Over the past years, several methylation-related databases have been developed for managing methylation data (15–22). However, none of them is designed to support developmental studies by integrating high-resolution whole-genome DNA methylomes from multiple developmental stages in different organisms. Specifically, MethylomeDB (15) is focused only on a specific tissue. DiseaseMeth (16), MethyCancer (17) and PubMeth (18) link methylome dedicatedly with human cancer/disease.

*To whom correspondence should be addressed. Tel: +86 10 8409 7261; Fax: +86 10 8409 7845; Email: zhangzhang@big.ac.cn
Correspondence may also be addressed to Jing Zhang. Tel: +86 10 8409 7488; Fax: +86 10 8409 7845; Email: zhangj@big.ac.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

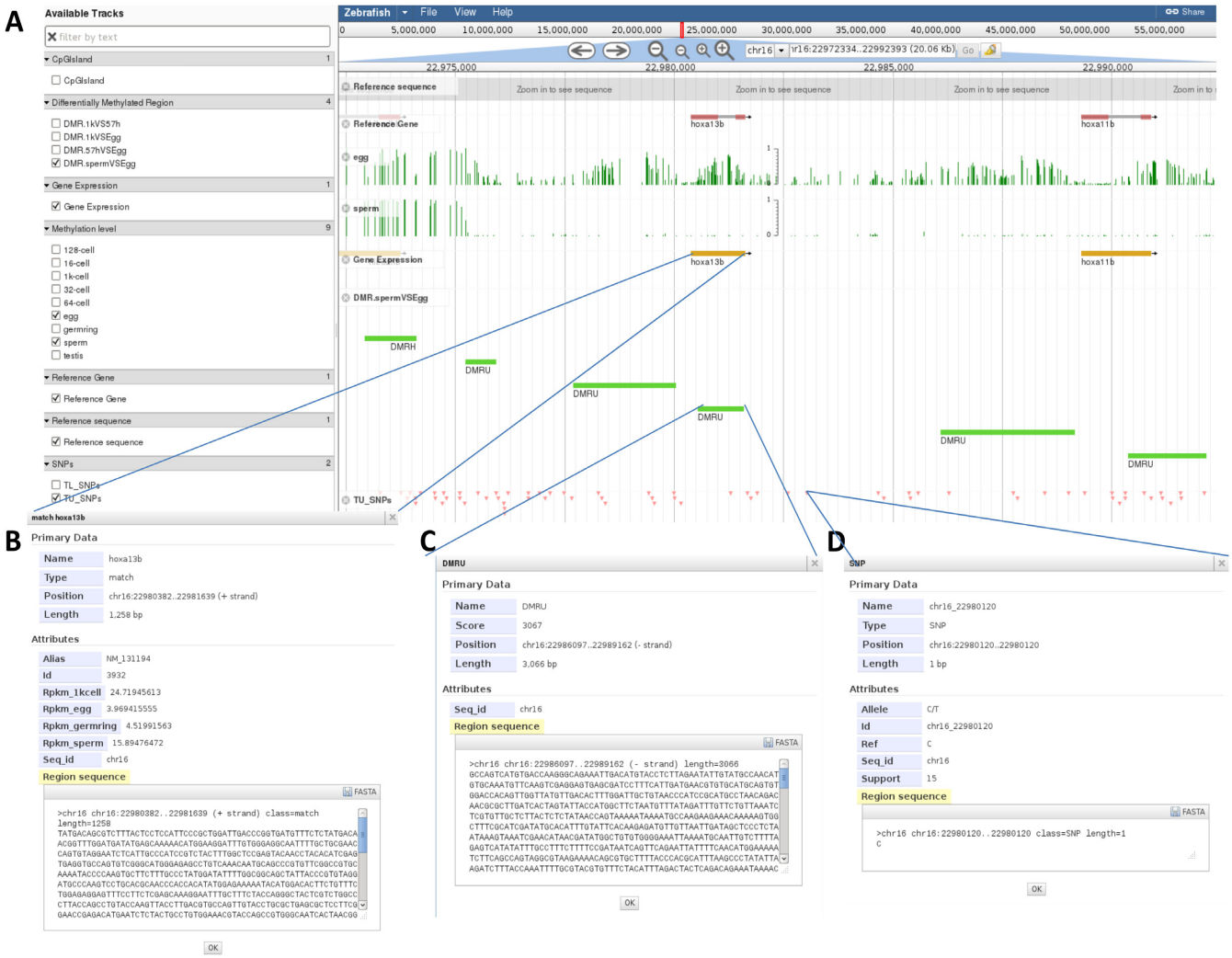

Figure 1. Screenshots of the methylome browser in MethBank. (**A**) Overview of the methylome browser that corresponds to two zebrafish developmental-associated *hoxa* genes, showing methylation levels, gene expression profiles, DMRs and SNPs. (**B**) Detailed gene expression information for *hoxa13b*. (**C**) Detailed DMR information for *hoxa13b*. (**D**) Detailed SNP information for *hoxa13b*.

MethDB (19) and PCMdb (20) are not committed to storing whole-genome single-base-resolution methylome data. NGSmethDB does not contain methylome data or other related omics data in gametes and early embryos (21). Cyclonet is under construction as of 28 June 2014 (http://cyclonet.biouml.org) and what we can learn is that it is mainly centered on cell cycles and does not include methylome data of gametes and early embryos. MethBase does not focus on methylomes of gametes and early embryos for developmental studies (22). NCBI Epigenomics Resources is designed for general research purposes, mainly providing pre-computed methylation results at individual cytosimes. Clearly, it can be seen that there is a lack of a specialized database that contains high-resolution genome-wide developmental methylomes with the aim to exploit the full potential of methylomes for systematic analysis of methylation dynamics during development.

Here we develop MethBank (http://dnamethylome.org), a database of DNA methylome programming that integrates the genome-wide single-base nucleotide methylomes of gametes and early embryos covering multiple diverse developmental stages and spanning different model organisms. Unlike extant databases, MethBank includes multiple whole-genome methylomes of gametes and early embryos at different developmental stages, incorporates related data (e.g. expression, genetic polymorphism) to facilitate systematic integrative investigation of DNA methylome reprogramming in embryonic development, and provides a configurable and interactive methylome browser to visualize high-resolution methylation data as well as other related data.

IMPLEMENTATION

MethBank has been implemented using MySQL (http://www.mysql.org; a free and popular relational database management system), JSP (JavaServer Pages; a technology facilitating rapid development of dynamic web pages based on the Java programming language) and Apache Tomcat (http://tomcat.apache.org; an open source web server for

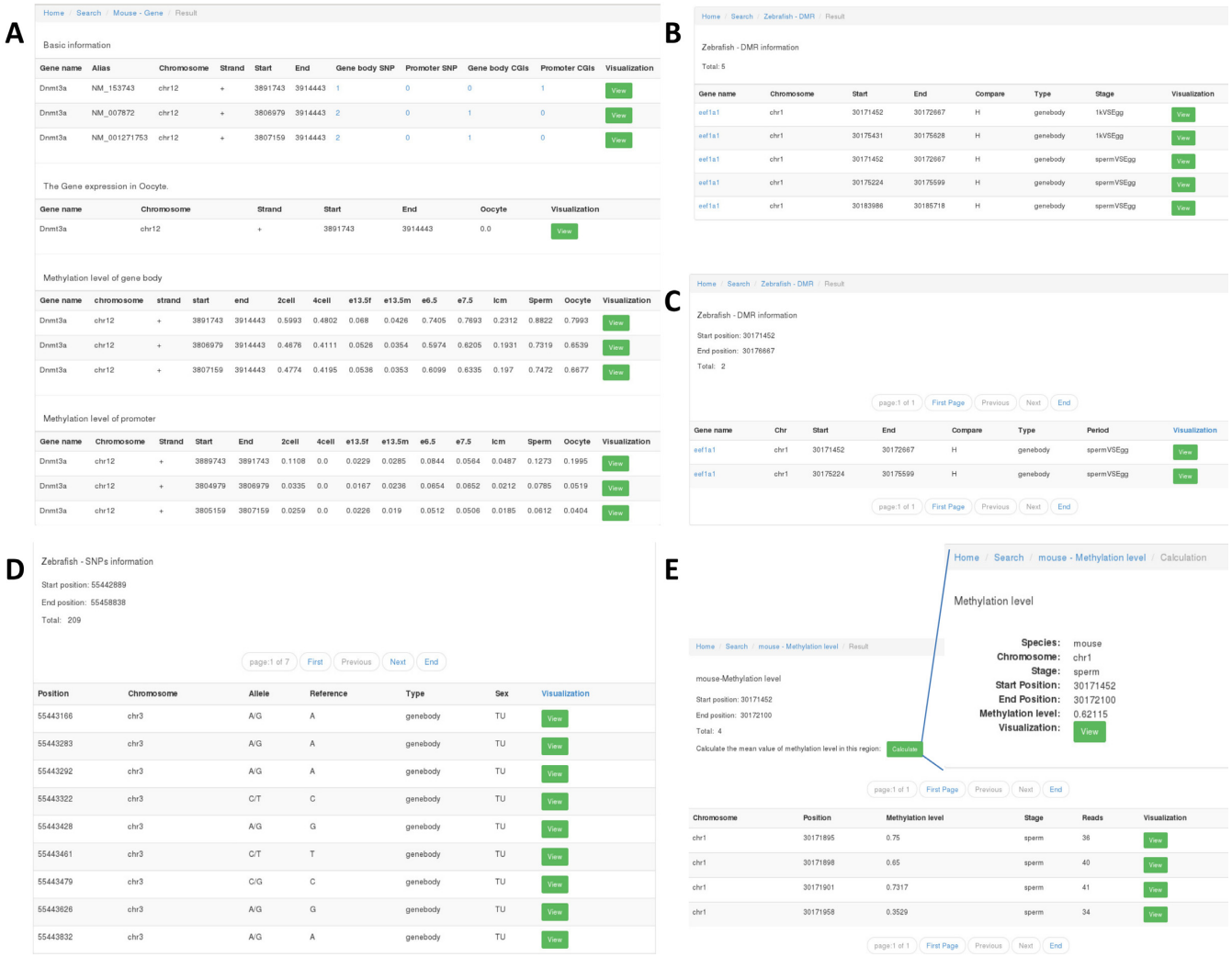

Figure 2. Screenshots of information query in MethBank. (**A**) Methylation states at promoter and gene body as well as basic information and gene expression, by searching the mouse gene *DNMT3a*. (**B**) DMR information for the zebrafish gene *EEF1A1*. (**C**) DMR information, by searching a genomic region from 30171452 to 30176667 in zebrafish chromosome 1. (**D**) SNP information, by searching a genomic region from 55442889 to 55458838 in zebrafish chromosome 3. (**E**) Methylation levels for all CG sites in a genomic region from 30171452 to 30172100 in mouse chromosome 1 and its average methylation level (0.62115).

Java code to run in) on a Red Hat Enterprise Linux Server. The web pages were developed using Eclipse (http://www.eclipse.org), an integrated development environment (IDE) that features rapid development of Java-based web applications and simplified connection with database management systems. To provide friendly and interactive web pages, browser-based interfaces were coded in JSP and AJAX (Asynchronous JavaScript and XML, a collection of web development technologies for creating highly interactive web applications), enabling data transfer between server and browser asynchronously without interfering with the display of the current web page. MethBank is freely available at http://dnamethylome.org.

DATABASE CONTENT AND USAGE

MethBank is a database of DNA methylome programming, dedicating to storing, browsing and visualizing single-base nucleotide whole-genome DNA methylation data of gametes and early embryos in different animals. Based on our previous studies (3,4), MethBank incorporates large cohorts of gametes and early embryo methylomes at single-base nucleotide resolution on two well-studied species, *D. rerio* and *M. musculus*. For each species, there are nine different developmental stages involving two gametes and seven embryos (for details see http://dnamethylome.org/about). For each developmental stage, >17 and ∼19 million methylated CG sites are included for zebrafish and mouse, respectively, covering about 90% of all CpGs in the whole genome. To enable in-depth investigation of DNA methylation data, MethBank profiles methylation level for each methylated CG site and identifies differentially methylated regions (DMRs) between oocyte and sperm, facilitating users to explore the dynamic methylation changes between different developmental stages.

As genetic polymorphisms may disrupt the methylation status and gene expression correlates closely with DNA

Table 1. MethBank data content and statistics as of 1 July 2014

Data content	Data statistics
Methylation data (WGBS[a])	
Sperm (zebrafish)	20 606 757 CpG sites
Oocyte (zebrafish)	20 987 274 CpG sites
16-Cell (zebrafish)	18 948 935 CpG sites
32-Cell (zebrafish)	20 825 693 CpG sites
64-Cell (zebrafish)	17 836 722 CpG sites
128-Cell (zebrafish)	19 889 625 CpG sites
1k-Cell (zebrafish)	20 815 960 CpG sites
Germ-ring (zebrafish)	19 258 632 CpG sites
Testis (zebrafish)	20 095 377 CpG sites
Sperm (mouse)	19 455 209 CpG sites
Oocyte (mouse)	19 100 312 CpG sites
2-Cell (mouse)	19 844 915 CpG sites
4-Cell (mouse)	19 637 338 CpG sites
E7.5 (mouse)	19 644 831 CpG sites
ICM (mouse)	19 523 556 CpG sites
E13.5 female (mouse)	19 253 566 CpG sites
E13.5 male (mouse)	19 352 586 CpG sites
SNP[b]data	
TU female (zebrafish)	3 653 795 SNPs
TL male (zebrafish)	3 423 892 SNPs
Mouse	2 561 574 SNPs
Geneexpression data	
Sperm (zebrafish)	13 529 genes
Oocyte (zebrafish)	11 191 genes
1k-Cell (zebrafish)	12 168 genes
Germ-ring (zebrafish)	12 129 genes
Oocyte (mouse)	18 259 genes
CpG islanddata	
Zebrafish	8768 CpG islands
Mouse	37 730 CpG islands
DMR[c]data	
Sperm versus oocyte (zebrafish)	53 680 DMRs
1k-cell versus oocyte (zebrafish)	51 886 DMRs
1k-cell versus germ-ring (zebrafish)	95 DMRs
Germ-ring versus oocyte (zebrafish)	36 004 DMRs
Sperm versus oocyte (mouse)	2000 DMRs

[a]WGBS: whole-genome bisulfate sequencing.
[b]SNP: single nucleotide polymorphism.
[c]DMR: differentially methylated region.

methylation, a large number of single nucleotide polymorphisms (SNPs) and expression profiles are also included in MethBank. Consequently, interconnections among different omics data are established and presented in MethBank. The detailed data statistics in MethBank are summarized in Table 1 and maintained online at http://dnamethylome.org/about. For a given region or gene, MethBank is able to profile methylation levels, locate DMRs inside, and unveil the corresponding expression and SNP information. Considering that advanced users may download raw data for their own analysis, the 'Download' page provides links to whole-genome single-base-resolution methylomes at different developmental stages.

To visualize single-base-resolution DNA methylation data, an interactive and user-friendly methylome browser built on JBrowse (23) is deployed in MethBank (Figure 1). For each species, the methylome browser includes a variety of data tracks (namely, CpG island, DMR, gene expression, methylation level, reference gene, reference sequence and SNP) and allows users to choose tracks of interest and to zoom and scroll any region along the genome. Therefore, the methylome browser is of usefulness to investigate methylation status of specific genes/regions across different developmental stages by taking account of multiple relevant data tracks (Figure 1A). Moreover, when clicking a gene/region on a specific track, its corresponding details are displayed and accessible for download (Figure 1B–D). Equipping with the methylome browser, MethBank is able to visualize high-resolution DNA methylation profiles as well as DMRs, gene expression levels, SNPs, CpG islands, etc., in an interactive manner and thus is of great utility to investigate methylation patterns of gametes and early embryos at different developmental stages within specific genes/regions.

To support information search and exploration, MethBank provides friendly web interfaces to retrieve a diversity of information for a specific gene or region (Figure 2). By specifying a gene symbol, users can obtain its methylation states at promoter and gene body across multiple developmental stages, as well as its basic information, gene expression, etc. (Figure 2A). For a given gene symbol (Figure 2B) or a specified genomic region (Figure 2C), MethBank can also provide all relevant DMRs between two developmental stages. Detailed SNP information is available for any genomic region, including allele information and genomic locus (Figure 2D). Moreover, MethBank provides not only the detailed methylation levels for all CG sites in a specific genomic region, but also the averaged methylation level for this region (Figure 2E). Additionally, all these information can be interactively and integratedly displayed in the methylome browser just by clicking the 'View' link.

DISCUSSION AND FUTURE DEVELOPMENTS

Different from extant databases, MethBank features (1) integrating the whole-genome single-base-resolution methylomes of gametes and early embryos at multiple developmental stages; (2) storing vast amounts of methylated CG sites in zebrafish and mouse, with >17 million and ∼19 million in count for each developmental stage, respectively; (3) incorporating other related omics data, interconnecting them with methylomes and building a methylome browser for visualization of all types of data in a genomic context; and (4) allowing the online query of methylation levels, DMRs, CpG islands, expression profiles and SNP information for a specific region or gene. Taken together, MethBank integrates and visualizes high-resolution genome-wide DNA methylomes as well as gene expression profiles and genetic polymorphisms, enabling identification of potential DNA methylation signatures in different developmental stages and accordingly providing an important resource for the epigenetic and developmental studies.

MethBank is committed to integrating the genome-wide single-base nucleotide methylomes of gametes and early embryos in different animals. With the rapid advancements of high-throughput sequencing technologies, more and more single-base-resolution developmental-related methylation data will become available in the following years, which bear great promises in unveiling fundamentals of DNA methylation in the development and differentiation of various cell types in different organisms. Therefore, future developments for MethBank include incorporation of whole-

genome methylomes at multiple developmental stages from other species. Accordingly, MethBank will continue to integrate related types of data (e.g. expression, SNP) from different resources and add more methylation analysis tools. Considering the increasing volume of methylation data, it is also important to develop web pages and tools to allow the easy incorporation of new data. Furthermore, MethBank will also provide orthologous genes in different species and develop web interfaces to facilitate cross-species comparison of DNA methylation at different developmental stages. The methylome browser will be further improved to support interactive visualization of big methylation data as well as other related data. In addition to our DNA methylation data generated in-house, we also invite the scientific community to submit their methylation data to MethBank and to build collaborations in improving the functionalities of MethBank.

ACKNOWLEDGEMENT

We thank Dr Jun Yu for valuable discussions on this work and members of the Zhang Lab for reporting bugs and sending comments.

FUNDING

Strategic Priority Research Program of the Chinese Academy of Sciences [XDB13040000 to Z.Z. and J.L.]; National Natural Science Foundation of China [31200958 to J.Z., 91219104 to J.L. and 31000584 to R.L.]; Youth Innovation Promotion Association of the Chinese Academy of Sciences [to J.Z.]; the '100-Talent Program' of the Chinese Academy of Sciences [Y1SLXb1365 to Z.Z.]; National Programs for High Technology Research and Development [863 Program; 2012AA020409 to Z.Z.]; the Ministry of Science and Technology of the People's Republic of China. Funding for open access charge: National Programs for High Technology Research and Development [2012AA020409].
Conflict of interest statement. None declared.

REFERENCES

1. Li,E., Bestor,T.H. and Jaenisch,R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
2. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
3. Jiang,L., Zhang,J., Wang,J., Wang,L., Zhang,L., Li,G., Yang,X., Ma,X., Sun,X., Cai,J. *et al.* (2013) Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, **153** , 773–784.
4. Wang,L., Zhang,J., Duan,J., Gao,X., Zhu,W., Lu,X., Yang,L., Zhang,J., Li,G., Ci,W. *et al.* (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell*, **157**, 979–991.
5. Potok,M.E., Nix,D.A., Parnell,T.J. and Cairns,B.R. (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell*, **153**, 759–772.
6. Hackett,J.A. and Surani,M.A. (2013) Beyond DNA: programming and inheritance of parental methylomes. *Cell*, **153**, 737–739.
7. Wossidlo,M., Nakamura,T., Lepikhov,K., Marques,C.J., Zakhartchenko,V., Boiani,M., Arand,J., Nakano,T., Reik,W. and Walter,J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.*, **2**, 241.
8. Gu,T.P., Guo,F., Yang,H., Wu,H.P., Xu,G.F., Liu,W., Xie,Z.G., Shi,L., He,X., Jin,S.G. *et al.* (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, **477**, 606–610.
9. Inoue,A. and Zhang,Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, **334**, 194.
10. Smith,Z.D., Chan,M.M., Mikkelsen,T.S., Gu,H., Gnirke,A., Regev,A. and Meissner,A. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**, 339–344.
11. Beck,S. and Rakyan,V.K. (2008) The methylome: approaches for global DNA methylation profiling. *Trends Genet.*, **24**, 231–237.
12. Laird,P.W. (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.*, **11**, 191–203.
13. Ball,M.P., Li,J.B., Gao,Y., Lee,J.H., LeProust,E.M., Park,I.H., Xie,B., Daley,G.Q. and Church,G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.*, **27**, 361–368.
14. Harris,R.A., Wang,T., Coarfa,C., Nagarajan,R.P., Hong,C., Downey,S.L., Johnson,B.E., Fouse,S.D., Delaney,A., Zhao,Y. *et al.* (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.*, **28**, 1097–1105.
15. Xin,Y., Chanrion,B., O'Donnell,A.H., Milekic,M., Costa,R., Ge,Y. and Haghighi,F.G. (2012) MethylomeDB: a database of DNA methylation profiles of the brain. *Nucleic Acids Res.*, **40**, D1245–D1249.
16. Lv,J., Liu,H., Su,J., Wu,X., Liu,H., Li,B., Xiao,X., Wang,F., Wu,Q. and Zhang,Y. (2012) DiseaseMeth: a human disease methylation database. *Nucleic Acids Res.*, **40**, D1030–D1035.
17. He,X., Chang,S., Zhang,J., Zhao,Q., Xiang,H., Kusonmano,K., Yang,L., Sun,Z.S., Yang,H. and Wang,J. (2008) MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res.*, **36**, D836–D841.
18. Ongenaert,M., Van Neste,L., De Meyer,T., Menschaert,G., Bekaert,S. and Van Criekinge,W. (2008) PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res.*, **36**, D842–D846.
19. Grunau,C., Renault,E., Rosenthal,A. and Roizes,G. (2001) MethDB—a public database for DNA methylation data. *Nucleic Acids Res.*, **29**, 270–274.
20. Nagpal,G., Sharma,M., Kumar,S., Chaudhary,K., Gupta,S., Gautam,A. and Raghava,G.P. (2014) PCMdb: pancreatic cancer methylation database. *Sci. Rep.*, **4**, 4197.
21. Hackenberg,M., Barturen,G. and Oliver,J.L. (2011) NGSmethDB: a database for next-generation sequencing single-cytosine-resolution DNA methylation data. *Nucleic Acids Res.*, **39**, D75–D79.
22. Song,Q., Decato,B., Hong,E.E., Zhou,M., Fang,F., Qu,J., Garvin,T., Kessler,M., Zhou,J. and Smith,A.D. (2013) A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One*, **8**, e81148.
23. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.