

Original article

DevMouse, the mouse developmental methylome database and analysis tools

Hongbo Liu^{1,†}, Rangfei Zhu^{2,†}, Jie Lv^{1,†}, Hongjuan He¹, Lin Yang³, Zhijun Huang¹, Jianzhong Su², Yan Zhang², Shihuan Yu^{4,*} and Qiong Wu^{1,*}

¹Department of Developmental Biology, School of Life Science and Technology, State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150001, China, ²Department of Computational Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China, ³Department of Food Science, School of Food Science and Engineering, Harbin Institute of Technology, Harbin 150001, China and ⁴Department of Respiratory Medicine, the First Affiliated Hospital of Harbin Medical University, Harbin 150001, China

*Corresponding author: Tel: +86 451 86403181; Fax: +86 451 86403181; Email: kigo@hit.edu.cn

*Correspondence may also be addressed to Shihuan Yu. Tel: +86 451 86416944; Fax: +86 451 86416944; Email: yushihuan2000@126.com

†These authors contributed equally to this work.

Submitted 10 September 2013; Revised 24 October 2013; Accepted 11 December 2013

Citation details: Liu, H., Zhu R., Lv J., et al. DevMouse, the mouse developmental methylome database and analysis tools. *Database*, Vol. 2013, Article ID bat084, doi:10.1093/database/bat084

DNA methylation undergoes dynamic changes during mouse development and plays crucial roles in embryogenesis, cell-lineage determination and genomic imprinting. Bisulfite sequencing enables profiling of mouse developmental methylomes on an unprecedented scale; however, integrating and mining these data are challenges for experimental biologists. Therefore, we developed DevMouse, which focuses on the efficient storage of DNA methylomes in temporal order and quantitative analysis of methylation dynamics during mouse development. The latest release of DevMouse incorporates 32 normalized and temporally ordered methylomes across 15 developmental stages and related genome information. A flexible query engine is developed for acquisition of methylation profiles for genes, microRNAs, long non-coding RNAs and genomic intervals of interest across selected developmental stages. To facilitate in-depth mining of these profiles, DevMouse offers online analysis tools for the quantification of methylation variation, identification of differentially methylated genes, hierarchical clustering, gene function annotation and enrichment. Moreover, a configurable MethyBrowser is provided to view the base-resolution methylomes under a genomic context. In brief, DevMouse hosts comprehensive mouse developmental methylome data and provides online tools to explore the relationships of DNA methylation and development.

Database URL: <http://www.devmouse.org/>

Introduction

DNA methylation is an epigenetic modification involving the addition of a methyl group at the 5'-position cytosine in DNA sequence (1). In mammalian genomes, most cytosines in CpG dinucleotides are methylated by DNA methyltransferases, whereas those in CpG islands are protected from methylation (2). DNA methylation plays crucial roles in transcriptional regulation, genomic imprinting, X chromosome inactivation and long-term repression (3,4). It has been reported that DNA methylation is highly

dynamic during mammalian development and participates in regulation of embryogenesis, cell-lineage determination and the genesis of germ cells (5,6). The aberrant DNA methylation programming and reprogramming during development would cause the inheritance of epigenetic mutations that have been found to be associated with human diseases as suggested by studies on mouse models (7,8). Owing to the important regulatory roles and the improvement of high-throughput technologies, DNA methylation is extensively studied in the development progress of the mouse, which is the model organism most

closely related to humans. Studying the influence of DNA methylation on developmental genes in mouse development should offer insight into the mechanisms affecting mammalian development and human developmental disorders, given the ethical issues of human studies.

Using bisulfite conversion coupled with second-generation sequencing, researchers from around the world have profiled DNA methylomes for different developmental stages of the mouse and analyzed the DNA methylation during mouse development (9–16). By reduced representation bisulfite sequencing, Meissner *et al.* mapped the genome-scale DNA methylation maps of pluripotent and differentiated cells, and found DNA methylation undergoes demethylation and then remethylation during the development of the mouse nervous system (9). By profiling and studying a genome-scale base-resolution timeline of DNA methylation in the pre-specified embryo, Smith and colleagues found DNA methylation dynamics in the early mammalian embryo (13). A base-resolution allele-specific DNA methylation map in the mouse genome revealed the roles of differential methylation in the regulation of imprinting and allele-specific gene expression in mammalian cells (14). Most recently, Kobayashi *et al.* profiled the high-resolution DNA methylomes of primordial germ cells via whole-genome shotgun bisulfite sequencing and found gender-specific reprogramming from E10.5 to E16.5 (15). The integration and depth mining of these methylomes in temporal order should help us to gain further knowledge about dynamic DNA methylation during development from a global perspective. It should be possible to discover potentially novel developmental genes/regions regulated by DNA methylation via integrating methylomes across multiple developmental stages and identifying differentially methylated genes during development. Integrative analysis of methylomes scattered in different data resources is a great idea, but costly and difficult to implement for experimental biologists with limited bioinformatics experience.

Currently, there are a few databases involved in DNA methylation. The data resource databases National Center for Biotechnology Information (NCBI) Epigenomics (17), NGSmethDB (18) and MethDB (19) were designed as a great data pool for epigenetic modification data or DNA methylation data stored according to the experiments and samples. NCBI Epigenomics provides genome-wide maps of DNA and histone modifications from a diverse collection of epigenomic data sets. NGSmethDB is a database for storage and retrieval of methylation data by next-generation sequencing. MethDB focuses on environmental effects on DNA methylation. The disease methylation databases MethyCancer (20), DiseaseMeth (21) and MethylomeDB (22) were developed to study the aberrant DNA methylation alterations in human disorders. MethyCancer focuses on the integrated cancer-related DNA methylation data.

DiseaseMeth is a web-based resource focused on the aberrant methylomes of human diseases, and MethylomeDB is a database containing genome-wide DNA methylation profiles for neurodevelopmental and neuropsychiatric disorders. However, there is no specialized and comprehensive database that focuses on storage of mouse developmental methylomes in temporal order or provides convenient analysis tools for in-depth mining of methylation dynamics from these data.

Thus, DevMouse was developed to store the mouse developmental methylomes in temporal order and provide online analysis tools for mining of the developmental genes/regions with dynamic DNA methylation during mouse development. The current version of DevMouse stores temporally ordered DNA methylomes covering multiple developmental stages, which should be useful for a wide range of developmental biologists. DevMouse supports users to search for the methylation patterns of various genome items such as genes, microRNAs (miRNAs), long non-coding RNAs (lncRNAs), CpG islands and other genome regions, which should benefit broad researchers focusing on molecular biology from genes to specific genome regions. Furthermore, the convenient analysis tools are provided for depth mining of novel knowledge from integrated methylomes in a global perspective. DevMouse also includes a configurable methylation browser, MethyBrowser, by which base-resolution developmental methylomes can be shown under a mouse genome context. All the search and analysis results can be viewed as graphs for view and downloaded as figures or tables for further analysis.

Database construction and content

DevMouse was designed to store high-throughput DNA methylation data during mouse development in temporal order. The current version of DevMouse consists of 32 DNA methylomes in single-base resolution across 15 mouse developmental stages that were collected from public DNA methylation resources (23–28) and genome information (genes, mRNAs, lncRNA, CpG islands and others) obtained from public genome databases (29–33) (Figure 1 and Table 1). These methylomes are profiled by the next-generation sequencing technologies coupled with bisulfite conversion. In these methylomes, methylated cytosine can be distinguished from unmethylated cytosine by the presence of a cytosine versus thymine residue during sequencing. The proportion of methylated cytosine is treated as the methylation level, which ranges from 0% representing unmethylated cytosine to 100% fully methylated cytosine. All methylation data were subsequently processed according to the same procedure. Before being finally stored, the methylomes from assemblies other than the University of California, Santa Cruz (UCSC) July 2007 mouse reference sequence (mm9, NCBI build 37) were

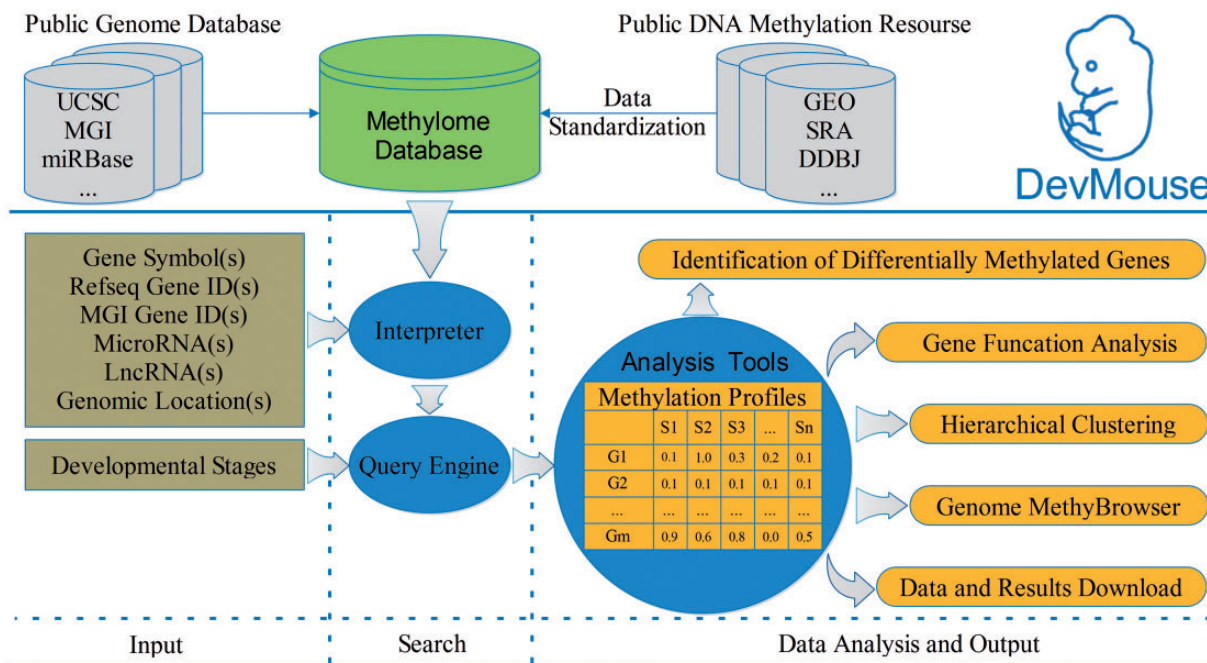


Figure 1. Overview of building and workflow of DevMouse. DevMouse integrates DNA methylomes of mouse development from public DNA methylation resources and the mouse genome information. Users can input multiple genome items to the query engine to gain the methylation profiles of genes/regions in different developmental stages. The analysis tools enable users to carry out online analysis on the methylation profiles, including identification of differentially methylated genes/regions, gene function analysis, hierarchical clustering and visualization in MethyBrowser. All search and analysis results can be downloaded as flat format for further analysis.

converted into mm9 by the LiftOver tool from UCSC (30). All data available can be downloaded from the download page, which lists detailed information about the data including experiment name, experimental technology, cell/tissue type, developmental stage, sex, author information, download links and external database links. Based on these high-throughput cytosine methylomes, DevMouse provides the basic operations, search, analysis, view and download (Figure 1). A flexible query engine is provided for acquisition and investigation of the methylation profiles of genes/regions of interest. Powerful analysis tools written in Java facilitate in-depth mining of novel knowledge about DNA methylation and development. Moreover, the methylation information and novel findings can be viewed by the visualization modules based on Apache Batik Scalable Vector Graphics (SVG) toolkit, and can be downloaded before exiting the browser for further reuse.

Database use and access

Search methylation profiles of genes across developmental stages

The search toolkits in the homepage of DevMouse can be used to acquire the methylation states for any given genes/regions of interest across multiple developmental

stages. In the current version, DevMouse supports six kinds of gene IDs including the official gene symbol, Refseq gene, MGI gene, mRNA and lncRNA. For a gene >500 bp, the methylation state is calculated as the mean methylation level of the cytosines located in the promoter region from upstream 2 kb to downstream 500 bp of the transcriptional start site, and the methylation state of a gene <500 bp is calculated as the mean methylation level of the cytosines located in the whole gene region. As an example, we imported 19 gene symbols (listed in gene example 1 in the homepage) that were used for induction of pluripotent stem cells (34) and selected 13 developmental stages during mouse sperm development (stage example 1 in the homepage). The search result is displayed by default as an overview table that summarizes the methylation profiles of genes across multiple developmental stages as well as gene information and chromosomal location (Figure 2). The table contains links to a methylation pattern panel in which the methylation profile of a selected gene can be viewed, and links to the MethyBrowser in which the user can view the methylation profile as well as genomic information, according to the specified view parameters. The whole query result data can be downloaded to local computers from the download links in the overview panel.

Table 1. DevMouse data content and statistics as of 1 September 2013

Data content	Data statistics
DNA methylation	
Single base DNA methylome	32
Profiled by RRBS	26
Profiled by BS-Seq/WGSBS	6
Developmental stage	15
Genome information	
Cytosine	1068446432
With methylation data	42685552
Gene symbol	23417
With methylation data	23321
RefSeq gene	31263
With methylation data	31094
MGI gene	62083
With methylation data	52448
MicroRNA	3212
With methylation data	1478
LncRNA	1335
With methylation data	1335
CpG island	16026
With methylation data	15991
Other integrated data	
Gene expression	31840
Transcription factor binding site	13483
Repeat	4881442
Single-nucleotide polymorphism	14893502
Class of function annotation	11

Analysis tools to explore methylation dynamics and gene function

DNA methylation is highly dynamic during mouse development, and hypermethylation of gene promoter inhibits a few genes including developmental genes at a specific developmental stage (35,36). However, few genes were documented to exhibit differential promoter methylation during a whole development process due to less availability of DNA methylation data. As bisulfite sequencing enables profiling of mouse developmental methylomes on an unprecedented scale, experimental biologists with limited bioinformatics experience challenges in integrating and mining these data. To facilitate integrative analysis of methylomes, DevMouse offers online tools for quantitative analysis: (i) entropy-based quantification of methylation variation, (ii) identification of differentially methylated genes/regions, (iii) hierarchical clustering analysis of methylation profiles and (iv) gene function annotation and enrichment (Figure 2). An entropy-based approach is

integrated to quantify methylation variation of a gene across multiple developmental stages, with lower entropy indicating greater methylation variation as described in our previous study (37). The genes with entropy lower than a threshold are identified as differentially methylated genes that are marked in red in the result table. Furthermore, the stage specificity of a differentially methylated gene in each stage is measured by an algorithm embedded in DevMouse. The hierarchical clustering analysis tool can be used to study the methylation similarity among genes or stages, and identify the genes with similar methylation patterns across developmental stages. The functional similarity in 11 kinds of functions for a differentially methylated gene set can be retrieved directly by the tools for gene annotation and enrichment, which is based on DAVID API (38). Using these analysis tools, we identified differentially methylated genes across multiple stages in each of the development lines stored in the current database, and performed clustering of methylation profiles for differentially methylated genes. These differentially methylated genes and clustering maps can be obtained from the DMGenes page. One merit of these tools is that they are highly automatic for analyzing given genes, facilitating specific analysis focusing on identification of a functional gene set such as potentially novel developmental genes. All analysis results by these tools can be downloaded as in the figures or tables for further analysis.

MethyBrowser

Moreover, a configurable methylation browser, MethyBrowser, is developed using the Apache Batik SVG toolkit for users to view the methylomes and genome information simultaneously (Figure 2). The genome MethyBrowser connecting to a MySQL backend is used to show the methylation profiles of imported gene/region across multiple developmental stages at single-base precision. A color gradient from green (methylation value = 0%) to red (methylation value = 100%) is used to display the numeric methylation states of the cytosines. MethyBrowser also visualizes other available genomic annotations including chromosomal location, base sequence, gene structure, lncRNA, miRNA, CpG island, transcription factor binding site, gene expression, single-nucleotide polymorphism, repeat elements, sequence tagged site along the mouse draft assembly and the alignment in human genome (hg19) of this region. It should be noted that base sequence is shown only when the region is <150 bp. The alignment in the human genome (hg19) may be helpful for exploring the potential roles of orthologous genes/regions in the human genome. Features of MethyBrowser include the ability to view the region by specifying the genomic coordinates, to zoom and move the given region, to show and hide certain feature tracks by mouse-click configuring, as well as to access the function of the genes by links to Entrez

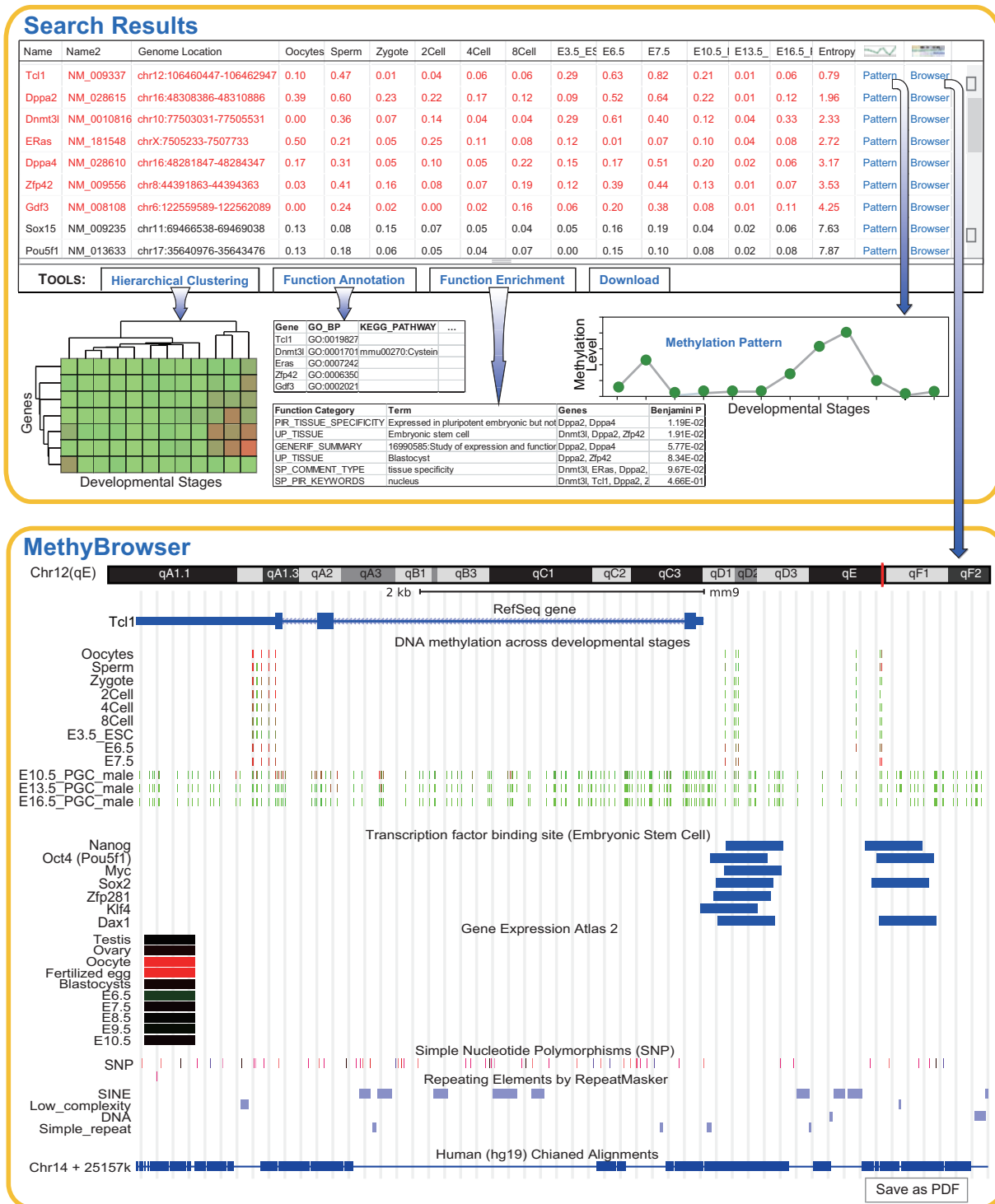


Figure 2. Search and analysis results on 19 genes during sperm development. Shown at the top is the default table and view generated by the query and analysis tools. The table shows the methylation profiles of 19 genes across 13 developmental stages and the sorted entropy values representing the methylation variation. The genes in red are differentially methylated across multiple stages. Using the buttons in the toolbar, the user can easily carry out online analysis including hierarchical clustering, gene function annotation and enrichment. The analysis results are shown as images or tables. The methylation pattern can be shown in a pop-up panel after clicking 'Pattern' link. When the 'Browser' link is clicked, the gene-centric methylation profile of gene *Tcl1* and related genomic information are viewed in MethyBrowser as shown at the bottom.

Gene in NCBI (39). As shown in Figure 2, links to MethyBrowser are also available in the search result page. By the 'Save as PDF' button, the browser graphic can be saved as PDF images that can be printed with Acrobat Reader and edited by many drawing programs such as Adobe Illustrator.

System design and implementation

DevMouse was constructed based on three major software components: an Apache Tomcat web server, a MySQL relational database and Java-based computational services. The backstage processing programs were written in Java, which are available on request. The web services were developed using Apache Struts2, a Java web application framework, and iBATIS, a persistence framework that automates the mapping between MySQL databases and objects in Java, both of which help guarantee the high performance and stability of the web services. Browser-based interfaces were built in JSP and AJAX. The Apache Batik SVG toolkit was used to render, generate and manipulate the SVG dynamically. DevMouse allows users to access all of the key features of the web application through their mobile device. DevMouse is available at <http://www.devmouse.org>.

Future perspective

The current version of DevMouse is the first release of our database. Although it contains a wealth of development-specific DNA methylomes in the mouse, which will be of great use both for experimental and bioinformatics researchers, the available data and functionality are still limited. Aiming to build a DNA methylome database focusing on the mouse development, continued efforts will be made to update the DevMouse data, add more methylation analysis tools and improve the functionality of the database and MethyBrowser. As the rapid profiling of DNA methylomes in more and more samples based on high-throughput bisulfite sequencing, we will continuously collect the latest data sets in different developmental stages of the mouse to keep DevMouse up-to-date. We would like to invite and encourage the scientific community to submit their methylation data about mouse development to keep DevMouse updated and comprehensive. As a resource to study the potential roles of DNA methylation in mouse development, DevMouse could be extended with utilities for the identification and confirmation of developmental markers related to DNA methylation from large-scale methylome data (40). The MethyBrowser will be improved to display strand-specific methylation in higher resolution and be extended by more configurable functionalities. Because chromatin modifications including histone modification have also been reported as dynamic marks of mouse

developmental genes, we would extend the research scope and integrate chromatin modification data into DevMouse. We hope our continuous efforts working on the database will contribute to the understanding of epigenetic regulation in mouse development and modeling human development and disease.

Acknowledgement

The authors thank Dr Yaoping Lei for revising the manuscript.

Funding

National Natural Science Foundation of China (31171383, 31371478, 61075023, 31371334, and 61203262); the Fundamental Research Funds for the Central Universities (HIT.NSRIF.2010027); the Natural Science Foundation of Heilongjiang Province (C201217); and the Scientific Research Fund of Heilongjiang Provincial Education Department (12511272 and 12521270). Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest. None declared.

References

1. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
2. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
3. Boyes, J. and Bird, A. (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell*, **64**, 1123–1134.
4. Miranda, T.B. and Jones, P.A. (2007) DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.*, **213**, 384–390.
5. Reik, W., Dean, W. and Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–1093.
6. Saitou, M., Kagiwada, S. and Kurimoto, K. (2012) Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development*, **139**, 15–31.
7. Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
8. Zhao, X., Pak, C., Smrt, R.D. et al. (2007) Epigenetics and Neural developmental disorders: Washington DC, September 18 and 19, 2006. *Epigenetics*, **2**, 126–134.
9. Meissner, A., Mikkelsen, T.S., Gu, H. et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
10. Popp, C., Dean, W., Feng, S. et al. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, **463**, 1101–1105.
11. Guibert, S., Forne, T. and Weber, M. (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.*, **22**, 633–641.

12. Seisenberger,S., Andrews,S., Krueger,F. *et al.* (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell*, **48**, 849–862.
13. Smith,Z.D., Chan,M.M., Mikkelsen,T.S. *et al.* (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**, 339–344.
14. Xie,W., Barr,C.L., Kim,A. *et al.* (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*, **148**, 816–831.
15. Kobayashi,H., Sakurai,T., Miura,F. *et al.* (2013) High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.*, **23**, 616–627.
16. Bock,C., Tomazou,E.M., Brinkman,A.B. *et al.* (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.*, **28**, 1106–1114.
17. Fingerman,I.M., Zhang,X., Ratzat,W. *et al.* (2013) NCBI Epigenomics: what's new for 2013. *Nucleic Acids Res.*, **41**, D221–D225.
18. 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.
19. Amoreira,C., Hindermann,W. and Grunau,C. (2003) An improved version of the DNA methylation database (MethDB). *Nucleic Acids Res.*, **31**, 75–77.
20. He,X., Chang,S., Zhang,J. *et al.* (2008) MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res.*, **36**, D836–D841.
21. Lv,J., Liu,H., Su,J. *et al.* (2012) DiseaseMeth: a human disease methylation database. *Nucleic Acids Res.*, **40**, D1030–D1035.
22. Xin,Y., Chanrion,B., O'Donnell,A.H. *et al.* (2012) MethylomeDB: a database of DNA methylation profiles of the brain. *Nucleic Acids Res.*, **40**, D1245–D1249.
23. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
24. Barrett,T., Troup,D.B., Wilhite,S.E. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
25. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
26. Kodama,Y., Shumway,M. and Leinonen,R. (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
27. Tateno,Y., Miyazaki,S., Ota,M. *et al.* (2000) DNA data bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucleic Acids Res.*, **28**, 24–26.
28. Kaminuma,E., Mashima,J., Kodama,Y. *et al.* (2010) DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.*, **38**, D33–D38.
29. Karolchik,D., Hinrichs,A.S., Furey,T.S. *et al.* (2004) The UCSC table browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.
30. Meyer,L.R., Zweig,A.S., Hinrichs,A.S. *et al.* (2013) The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64–D69.
31. Blake,J.A., Richardson,J.E., Davison,M.T. *et al.* (1997) The mouse genome database (MGD). a comprehensive public resource of genetic, phenotypic and genomic data. The mouse genome informatics group. *Nucleic Acids Res.*, **25**, 85–91.
32. Griffiths-Jones,S., Grocock,R.J., van Dongen,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
33. Su,J., Zhang,Y., Lv,J. *et al.* (2010) CpG_MI: a novel approach for identifying functional CpG islands in mammalian genomes. *Nucleic Acids Res.*, **38**, e6.
34. Takahashi,K. and Yamanaka,S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
35. Holliday,R. and Pugh,J.E. (1975) DNA modification mechanisms and gene activity during development. *Science*, **187**, 226–232.
36. Borgel,J., Guibert,S., Li,Y. *et al.* (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.*, **42**, 1093–1100.
37. Zhang,Y., Liu,H., Lv,J. *et al.* (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res.*, **39**, e58.
38. Huang da,W., Sherman,B.T., Tan,Q. *et al.* (2007) DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, **35**, W169–W175.
39. Maglott,D., Ostell,J., Pruitt,K.D. *et al.* (2011) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
40. Bock,C. (2012) Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.*, **13**, 705–719.