D380–D386 *Nucleic Acids Research, 2022, Vol. 50, Database issue* *Published online 27 September 2021*
https://doi.org/10.1093/nar/gkab833

scMethBank: a database for single-cell whole genome DNA methylation maps

Wenting Zong[1,2,3,†], Hongen Kang[1,2,3,†], Zhuang Xiong[1,2,3], Yingke Ma[1,2], Tong Jin[1,2,3], Zheng Gong[1,2,3], Lizhi Yi[1,2], Mochen Zhang[1,2,3], Song Wu[1,2,3], Guoliang Wang[1,2,3], Yiming Bao[1,2,3,*] and Rujiao Li[1,2,*]

[1]National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [2]China National Center for Bioinformation, Beijing 100101, China and [3]University of Chinese Academy of Sciences, Beijing 100049, China

Received August 14, 2021; Revised September 06, 2021; Editorial Decision September 07, 2021; Accepted September 23, 2021

ABSTRACT

Single-cell bisulfite sequencing methods are widely used to assess epigenomic heterogeneity in cell states. Over the past few years, large amounts of data have been generated and facilitated deeper understanding of the epigenetic regulation of many key biological processes including early embryonic development, cell differentiation and tumor progression. It is an urgent need to build a functional resource platform with the massive amount of data. Here, we present scMethBank, the first open access and comprehensive database dedicated to the collection, integration, analysis and visualization of single-cell DNA methylation data and metadata. Current release of scMethBank includes processed single-cell bisulfite sequencing data and curated metadata of 8328 samples derived from 15 public single-cell datasets, involving two species (human and mouse), 29 cell types and two diseases. In summary, scMethBank aims to assist researchers who are interested in cell heterogeneity to explore and utilize whole genome methylation data at single-cell level by providing browse, search, visualization, download functions and user-friendly online tools. The database is accessible at: https://ngdc.cncb.ac.cn/methbank/scm/.

INTRODUCTION

As an important layer of epigenomics, DNA methylation (DNAm) provides important insights into transcriptional regulation and biological processes including genomic imprinting, early embryonic development and cancer progression (1–3). Although bulk whole genome bisulfite sequencing (WGBS) has made great efforts in mapping the DNA methylome landscape across types of tissues, it still has certain deficiencies in explaining the cell heterogeneity and understanding the development dynamics in specific biological status (4). Besides, it is difficult to obtain large numbers of cells for many important physiological issues, such as mammalian early embryogenesis (5,6). At present, advances in sequencing methods have enabled the development of strategies to analyze DNA methylation at single-cell resolution including scRRBS (7) and scBS-seq (8), and multi-omics approaches such as scTrioSeq2 (9) and scM&T-seq (10), which have greatly facilitated the exploration of cellular epigenetic heterogeneity.

However, the continuous accumulation of the vast amount of experiments and datasets poses a great challenge for the integration and reuse of single-cell DNA methylation data. In addition, how to retrieve such huge amount of whole genome methylation data is one of the bottlenecks due to the internet bandwidth limitation (11,12). In this case, most databases only provide the storage and downloading for the raw data, and therefore researchers cannot obtain intuitive and effective information from these data which lack unified metadata. Despite some efforts, there is still a drastically shortage of systematically designed single-cell DNA methylation databases up to now. For example, the only single-cell methylation database HeteroMeth (13), stores only 150 DNA methylation heterogeneity data rather than the whole genome methylation profiles. Additionally, the traditional bulk methylation analysis and visualization tools are intrinsically limited in interpreting the single-cell methylation data because of the special data structure of sparsity and duality. Therefore, it is highly necessary to establish a platform to comprehensively integrate published data from different studies to make them reusable and comparable.

Here, we present scMethBank (The single-cell Methylation Bank), a comprehensive and curated database that

*To whom correspondence should be addressed. Tel: +86 10 84097638; Email: lirj@big.ac.cn
Correspondence may also be addressed to Yiming Bao. Tel: +86 10 84097858; Email: baoym@big.ac.cn
†The authors wish it to be known that, in their opinion, first two authors should be regarded as Joint First Authors.

integrates single-cell methylation data and metadata from publicly available datasets (Figure 1). By collecting data produced by several kinds of main bisulfite-based strategies and characterizing methylation status at single-base resolution with a standardized workflow, scMethBank builds the whole genome methylation maps at single-cell level for human and mouse. It provides with the genome-wide single-cell DNA methylation profiles and value-added curation metadata across multiple biological conditions including cell types, developmental stages, disease states, and treatment methods. Moreover, random access to methylation data of any genome interval has been achieved by optimizing the underlying database storage and inquiry. A user-friendly interface is equipped to support interactive exploration and visualization of single-cell methylation profiles with single-base accuracy. Other functions in scMethBank include data searching and downloading, interactive genome browser with custom tracks, and online analysis and graphical tools specifically designed for the single-cell methylation data. We expect this database can provide valuable single-cell methylation data resources to accelerate scientific discoveries and lead to a new paradigm into epigenetic heterogeneity.

MATERIALS AND METHODS

Data collection and curation

scMethBank collected and organized single-cell methylation sequencing data from NCBI sequence read archive (SRA) (14) through keyword and literature search. For meta data, sample descriptions were sourced from NCBI GEO database (15) and manually reviewed supplementary information submitted by the authors in the original publication if there is one.

Data processing and analysis

A standardized analysis workflow was applied to the following processes including sequencing reads format conversion, quality control, alignment, methylation level estimation, differential methylation analysis, cell cluster analysis, and gene set enrichment analysis.

All bisulfite sequences were subjected to quality control by FastQC v0.11.7 (http://www.bioinformatics.babraham.a.uk/projects/fastqc/). Adaptors and low-quality bases were trimmed using trim galore v0.6.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Next, reads those passed quality control were mapped to the reference genome (hg38 for human and mm10 for mouse). To recover more informative cytosines, scBS-map (16), a tool applying local alignment to enhance single-cell bisulfite sequencing data efficiency, was used to conduct reads alignment on account of the low ratio of unique mapped reads. Methylation calls were extracted from BAM files after the duplicated sequences had been removed using CGmapTools (11). The read coverage threshold used to call the DNA methylation level for single cytosine was $1\times$ for single-cell samples (17). Specifically, a cytosine site with greater than 90% DNA methylation was considered methylated, whereas one with $<10\%$ DNA methylation was considered unmethylated (8,17). All processed data were normalized to 0 (unmethylated) and 1 (methylated). After the methylation BED file was obtained, the subsequent steps including visualization and analysis were performed with customized R and Perl scripts. To verify the data process of scMethBank, we used a real biological dataset (GSE81233) of human preimplantation embryo to reproduce the dynamic development progress of methylation level (17).

Database implementation

scMethBank were organized with Spring Boot (http://spring.io/), deployed in Centos Linux environment. The processed data and annotated files were stored in MySQL (https://dev.mysql.com). The front-end display was achieved using HTML5 rendered by Thymeleaf (https://www.thymeleaf.org), Bootstrap4 (https://getbootstrap.com) and Semantic UI (https://semantic-ui.com). Several JavaScript libraries such as zTReeJS, HighchartJS EchartsJS, PlotlyJS and AJAX strategies were used to construct interactive and dynamic web pages. JBrowse (https://jbrowse.org) (18) was integrated to visualize single-cell methylation data on genes of interest. The interactive application Lollipop Plotter for single-cell methylation data was implemented based on customized Perl and R scripts. The background analysis of the enrichment tool was also underpinned by the in-house scripts which used R packages including ChIPseeker (19) and clusterProfiler (20).

DATABASE CONTENTS AND FEATURES

Overview of scMethBank

Presently, scMethBank provides genome-wide single-cell DNA methylation profiles and curated metadata of 8328 samples from both human and mouse, covering 15 projects, 29 cell types and two disease conditions. Embryonic cells (11.0%), cancer cells (14.4%), germ cells (10.7%), nerve cells (54.5%), stem cells (7.9%) and other cell types (2.3%) are recorded in the database. The biological contexts involved including early embryonic development, cancer progression, cell differentiation and aging. To our knowledge, the data we collected cover almost all of the known single-cell bisulfite-based sequencing methods. All processed data in scMethBank are freely accessible.

The homepage displays cell types information stored in scMethBank through a tree structure, which are linked to corresponding samples of interest. Meanwhile, the homepage supports searching function of multiple items including datasets accessions, tissues, cell types, treatment methods and diseases. In addition, users can navigate the whole database through four featured function modules: Browse, Visualize, Tools and Download. Apart from this, instructions for the usage of the database are accessible in the Documentation page.

Browsing interface

scMethbank provides a browsing interface and users can browse and perform some extended operations, such as searching, filtering and downloading. Besides, the browse

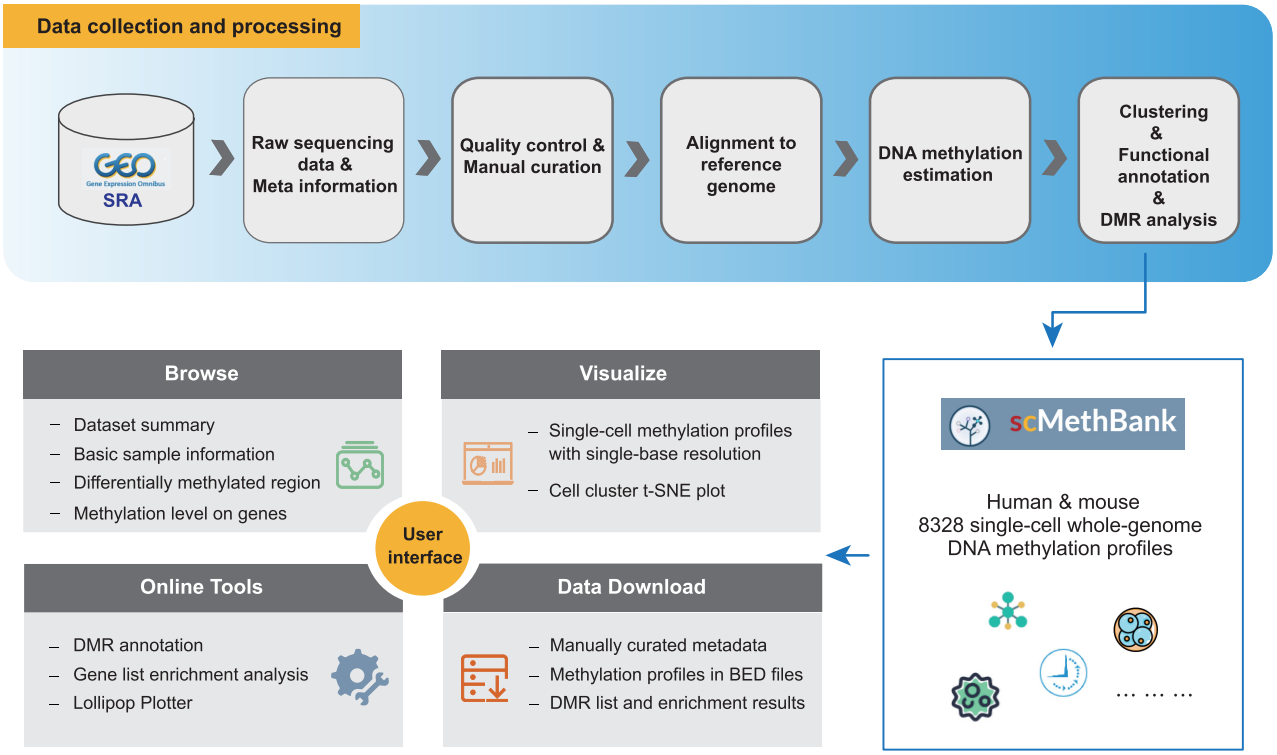

Figure 1. Workflow for data processing and overview of scMethBank.

module also stores information of samples and corresponding datasets as well as methylation levels of genes and differentially methylated regions (DMRs) between different cell types.

Datasets and samples. In the datasets browsing module (Figure 2A), kinds of relevant datasets information are included such as accession numbers, library protocols, cell types and their counts, technology platforms and others. To establish connections between single-cell multi-omics datasets and better characterize cell function and regulatory processes, links to single-cell expression databases including scRNASeqDB (21), Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) and Gene Expression Nebulas (GEN) (https://ngdc.cncb.ac.cn/gen/) are provided. Detailed information of every sample recorded in the database is displayed in the sample browse module (Figure 2A). Each sample is assigned 13 manually-curated attributes so users can narrow down the query results by corresponding keywords including the cell type, development stage, treatment, disease, sex and age. It's also allowed to select high-quality data from the detailed sequencing information in the drop-down secondary menu including bisulfite conversion efficiencies, sequencing platforms and sequencing depths. For search results, one or more samples can be added into the carts and downloaded with the corresponding processed BED format file or detailed meta table.

Genes. scMethBank provides easy access to the gene of interest by gene symbol, gene ID or genomic location, that will be of benefit to users to explore the methylation status and variation of genes in specific biological contexts in a greater depth. scMethBank returns the genes satisfying the criteria with gene alias, detailed description, genomic location and a link to the page where methylation details will be displayed. Here, we provide a box plot (Figure 2B) to show the average methylation levels of the gene across all cell types stored in the database, which makes it easy to compare the methylation level of the genes among different cell types. Beyond that, the methylation browser allows users to interactively visualize the single-cell methylation patterns of samples by loading or unloading the tracks for processed data and other related data at the single-base resolution, thereby helpful to make simple side-by-side comparison across different samples.

DMRs. Detection of differentially methylated regions which are altered during development or perturbed by disease (22) has important benefits to understand the underlying causes of differential gene expression (23). scMethBank provides easy access to DMRs between different cell types that are pre-identified by using published method (17). Based on the single-cell DNA methylation data, scMethBank has identified 165 918 DMRs across 6 datasets and 29 pairs of different cell types. DMRs were characterized with features including genomic locations, associated genes, annotation of genomic elements, and GO and KEGG enrichment (24,25) analysis results (Figure 2C). For the specific DMR region of a group of cells, clicking '+' under 'Methylation Distribution' displays a scatter plot of the methylation level distribution of the two cell types. All details displayed in the forms of tables and images are downloadable. These are significant to highlight the regulation and func-

Figure 2. Screenshot of the browse module of scMethBank. (**A**) Browse page for datasets (right) and samples (left) in the database. (**B**) Left: Detailed information can be viewed by selecting the gene or position in the gene browse module. Right: The top table shows the basic information of the selected gene (PGAM5), and the box plot in the middle shows the average methylation level of the gene of all samples included in the database. Genome browser provides users with the position of the gene (PGAM5) and CpG islands on it. (**C**) Browse DMRs by selecting the datasets and cell types pairs. Results shows the distribution statistical information and enrichment results about the DMR lists.

tional roles of DMRs by associating single-cell methylation profiles of different cell types, conditions or diseases with specific gene functions.

Interactive visualization

Although the average methylation level within a region reflects the basic characteristics of a heterogeneous population to some extent, it will also mask the complexity of methylation patterns (26). Therefore, the way to unleash the real value of single-cell methylation data is to make the genome-wide data searchable and visualized at single-base accuracy in the database. To correctly characterize the heterogeneity of methylation patterns, scMethBank builds up a vast amount of single-cell whole genome methylation data pool of Terabase level storage that allows users to retrieve methylation profiles with single-base precision from over 8000 different samples. In addition, scMethBank provides two ways to interactively visualize CpG methylation patterns for samples of interest (Figure 3A). First, a heatmap mode plot enables users to browse a relatively broader methylation pattern of the local region by inputting the gene name or specifying the genomic location. The heatmap provides an intuitive comparison of the methylation status of different samples. Secondly, for the patterns of certain sites of particular concern, a more refined lollipop-like plot is also allowed to help users scan and interpret single-cell methylation profiles with a greater precision. Besides, t-SNE analysis results of all single-cell samples from different datasets are displayed in the cell cluster module (Figure 3B) with point colors representing different cell groups.

Analysis and graphical tools

scMethBank provides researchers with user-friendly online tools for simple and practical downstream analysis functions including Lollipop plotter, DMR annotation and Enrichment analysis tools. Unlike the traditional bulk methylation methods, the series of online tools are designed for downstream analysis and visualization of methylation data especially single-cell methylation data on account of the duality for methylated and unmethylated.

Lollipop plotter, an easy-to-use graphical tool, is developed to help users interpret and visualize the patterns of single-cell DNA methylation data. Users can upload their own tab-delimited text file which describes the methylation status and define some simple parameters to generate a customized lollipop-like graphics (Figure 3C), where filled circles represent fully methylated sites and open circles represent unmethylated ones. This graphical tool offers an intuitive manner to visualize the methylation pattern of a local region at the single-cell level and is often used to indicate dynamically determined regions and DMRs.

DMR annotation tool (Figure 3D) offers capabilities for genomic element annotation and gene function enrichment to DMRs. The DMR list that users upload will be assigned to related genomic elements including promoter, 5′UTR, 3′UTR, exon, intron, downstream and intergenic. Based on the annotated gene list, GO and KEGG analysis will be performed and the results are also accessible. scMethBank also provides a direct enrichment analysis tool for a given gene set. DMR annotation and gene enrichment tools allow users to combine epigenetic information with gene functions and underlying regulatory relationships, which is of great significance for many biological problems. Users can easily use tools in scMethBank through a web server without consuming any local computational resources and all data will be removed from the server once the browser session terminates. In addition, we generate a corresponding job id for each task based on the IP address and task file and users can view their analysis results by job id later.

Data download

All processed single-cell DNA methylation data and metadata can be downloaded through the browse page. Further table files including DMR lists, annotation results, and methylation levels on genes are also accessible. Additionally, scMethBank provides a packaged download for files on the download page, where the methylation files of samples are stored in a compressed BED file format. Of course, due to the huge size of methylation data and the network bandwidth problems, we also provide users with a URL-based FTP download method, and users can perform download tasks on the command line.

DISCUSSION AND FUTURE DEVELOPMENTS

The emergence of single-cell DNA methylation sequencing has provided a new perspective and thinking for biological research. To the best of our knowledge, scMethBank is the first single-cell DNA methylation database which provides comprehensively curated metadata and analysis results processed by unified pipeline for all publicly available datasets. Powered by the optimized database structure, we set up a vast single-cell genome-wide methylation data pool to help users retrieve interested single-cell methylation profiles with single-base resolution. And for the first time, we provide a series of friendly online downstream tools to help interpret and apply single-cell methylation data through the web server. In the future, our work will be carried out in the following directions: (i) Continue to update and explore database and add more species, cell types and cell states. (ii) Update our pipeline actively to better assess the single-cell methylation data and downstream analysis results. (iii) Combine multi-omics data such as scRNA-seq and scATAC-seq (27) to explore associations between different epigenetic layers and potential associations with the transcriptome. (iv) Considering the high cost and technical difficulty of single-cell DNA methylation sequencing technology, some computational algorithms (28–30) that perform deconvolution of bulk DNA methylation data could be a good way to expand the richness of scMethBank. And a detailed resolution of cell types based on multi-omics integration will help to identify markers at epigenetics level. As an important supplement in the single-cell dimension to MethBank (31) and in the genome-wide dimension to EWAS (32,33), as well as one of the important resources in the National Genomics Data Center (NGDC) (34), scMethBank will be devoted to giving insights to more important biological issues such as early embryonic development, aging and diseases development.

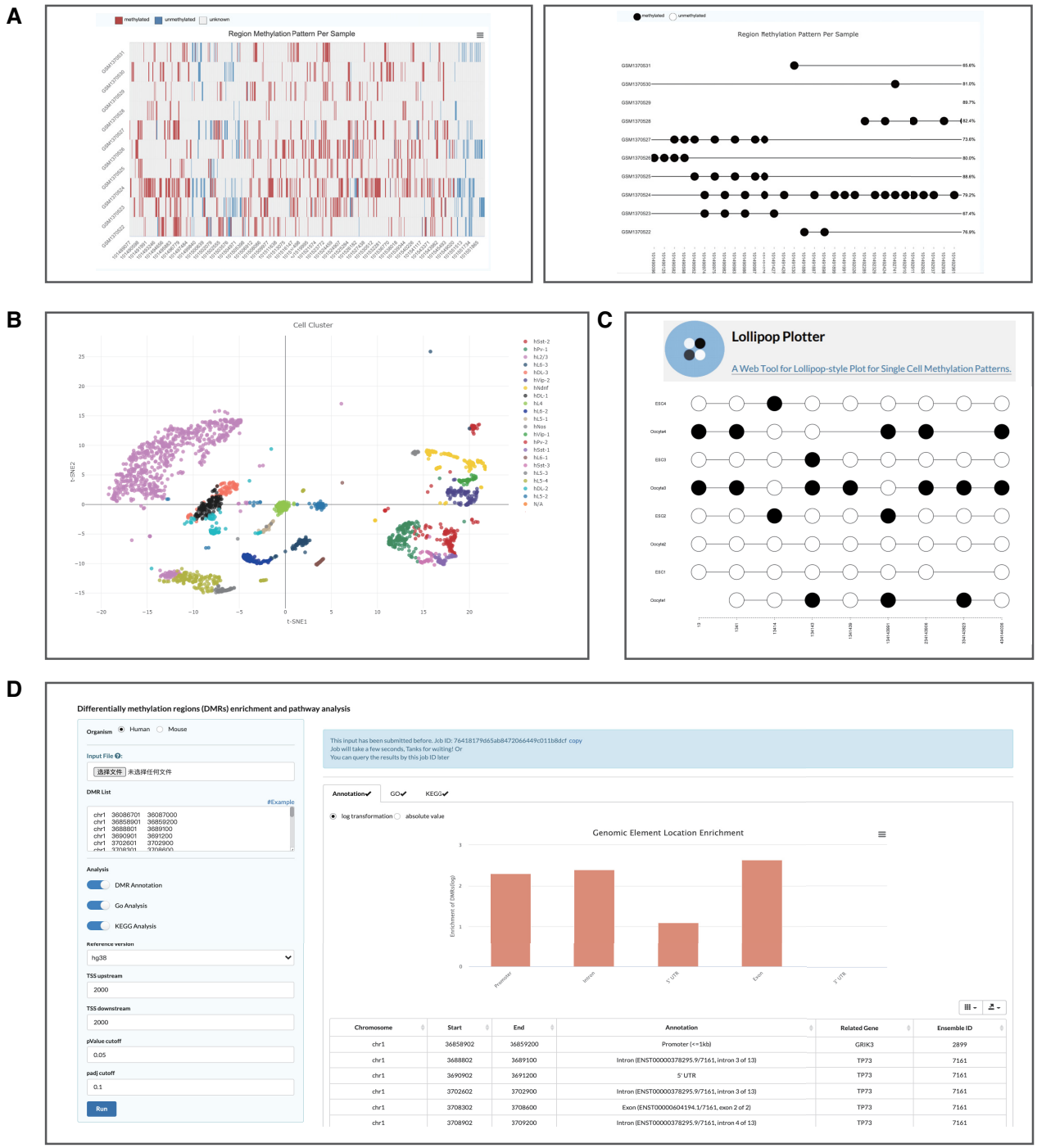

Figure 3. Visualize module and online tools of scMethBank. (**A**) Visualization of methylation patterns at single-base accuracy in a region. Left: Heatmap mode graph describing methylation patterns of Brca1 for 10 samples (MII oocyte) of GSE56879. Right: lollipop-style graph. (**B**) t-SNE analysis for human neuron cells in GSE97179. (**C**) Lollipop Plotter: The methylation patterns of a local region uploaded by users will be displayed in the form of lollipop-style graphs at single-base accuracy. (**D**) Annotation and enrichment analysis tools. The results including annotation results associated with gene and genome elements, GO and KEGG enrichment results are displayed on the web in real time.

DATA AVAILABILITY

scMethBank is a single-cell DNA methylation database which is freely available online and all data can be accessed at https://ngdc.cncb.ac.cn/methbank/scm/.

ACKNOWLEDGEMENTS

We thank a number of users for reporting bugs and providing constructive suggestions, and two anonymous reviewers for their valuable comments.

FUNDINGS

Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38030200]; Key Technology Talent Program of the Chinese Academy of Sciences; Genomics Data Center Construction of Chinese Academy of Sciences [WX145XQ07-04]; the Professional Association of the Alliance of International Science Organizations [ANSO-PA-2020-07]; the Open Biodiversity and Health Big Data Program of IUBS. Funding for open access charge: Key Technology Talent Program of the Chinese Academy of Sciences.
Conflict of interest statement. None declared.

REFERENCES

1. Baylin,S.B. and Jones,P.A. (2011) A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer*, **11**, 726–734.
2. Reik,W. and Walter,J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet*, **2**, 21–32.
3. Hackett,J.A. and Surani,M.A. (2013) DNA methylation dynamics during the mammalian life cycle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **368**, 20110328.
4. Schwartzman,O. and Tanay,A. (2015) Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.*, **16**, 716–726.
5. Smith,Z.D., Chan,M.M., Humm,K.C., Karnik,R., Mekhoubad,S., Regev,A., Eggan,K. and Meissner,A. (2014) DNA methylation dynamics of the human preimplantation embryo. *Nature*, **511**, 611–615.
6. Guo,H., Zhu,P., Yan,L., Li,R., Hu,B., Lian,Y., Yan,J., Ren,X., Lin,S., Li,J. *et al.* (2014) The DNA methylation landscape of human early embryos. *Nature*, **511**, 606–610.
7. Guo,H., Zhu,P., Wu,X., Li,X., Wen,L. and Tang,F. (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.*, **23**, 2126–2135.
8. Smallwood,S.A., Lee,H.J., Angermueller,C., Krueger,F., Saadeh,H., Peat,J., Andrews,S.R., Stegle,O., Reik,W. and Kelsey,G. (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods*, **11**, 817–820.
9. Bian,S., Hou,Y., Zhou,X., Li,X., Yong,J., Wang,Y., Wang,W., Yan,J., Hu,B., Guo,H. *et al.* (2018) Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, **362**, 1060–1063.
10. Angermueller,C., Clark,S.J., Lee,H.J., Macaulay,I.C., Teng,M.J., Hu,T.X., Krueger,F., Smallwood,S., Ponting,C.P., Voet,T. *et al.* (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods*, **13**, 229–232.
11. Guo,W., Zhu,P., Pellegrini,M., Zhang,M.Q., Wang,X. and Ni,Z. (2018) CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*, **34**, 381–387.
12. Grunau,C., Renault,E., Rosenthal,A. and Roizes,G. (2001) MethDB–a public database for DNA methylation data. *Nucleic. Acids. Res.*, **29**, 270–274.
13. Huan,Q., Zhang,Y., Wu,S. and Qian,W. (2018) HeteroMeth: a database of cell-to-cell heterogeneity in DNA methylation. *Genomics Proteomics Bioinformatics*, **16**, 234–243.
14. Kodama,Y., Shumway,M., Leinonen,R. and International Nucleotide Sequence Database, C. (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
15. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
16. Wu,P., Gao,Y., Guo,W. and Zhu,P. (2019) Using local alignment to enhance single-cell bisulfite sequencing data efficiency. *Bioinformatics*, **35**, 3273–3278.
17. Zhu,P., Guo,H., Ren,Y., Hou,Y., Dong,J., Li,R., Lian,Y., Fan,X., Hu,B., Gao,Y. *et al.* (2018) Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat. Genet.*, **50**, 12–19.
18. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
19. Yu,G., Wang,L.G. and He,Q.Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
20. Wu,T., Hu,E., Xu,S., Chen,M., Guo,P., Dai,Z., Feng,T., Zhou,L., Tang,W., Zhan,L. *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **3**, 100141.
21. Cao,Y., Zhu,J., Jia,P. and Zhao,Z. (2017) scRNASeqDB: a database for RNA-Seq based gene expression profiles in human single cells. *Genes*, **8**, 368.
22. Li,S., Garrett-Bakelman,F.E., Akalin,A., Zumbo,P., Levine,R., To,B.L., Lewis,I.D., Brown,A.L., D'Andrea,R.J., Melnick,A. *et al.* (2013) An optimized algorithm for detecting and annotating regional differential methylation. *BMC Bioinformatics*, **14**, S10.
23. Zhang,Y., Liu,H., Lv,J., Xiao,X., Zhu,J., Liu,X., Su,J., Li,X., Wu,Q., Wang,F. *et al.* (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res.*, **39**, e58.
24. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
25. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
26. Kapourani,C.A. and Sanguinetti,G. (2018) BPRMeth: a flexible Bioconductor package for modelling methylation profiles. *Bioinformatics*, **34**, 2485–2486.
27. Fang,R., Preissl,S., Li,Y., Hou,X., Lucero,J., Wang,X., Motamedi,A., Shiau,A.K., Zhou,X., Xie,F. *et al.* (2021) Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.*, **12**, 1337.
28. Teschendorff,A.E., Zhu,T., Breeze,C.E. and Beck,S. (2020) EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. *Genome Biol.*, **21**, 221.
29. Arneson,D., Yang,X. and Wang,K. (2020) MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. *Commun. Biol.*, **3**, 422.
30. Li,H., Sharma,A., Luo,K., Qin,Z.S., Sun,X. and Liu,H. (2020) DeconPeaker, a deconvolution model to identify cell types based on chromatin accessibility in ATAC-Seq data of mixture samples. *Front. Genet.*, **11**, 392.
31. Li,R., Liang,F., Li,M., Zou,D., Sun,S., Zhao,Y., Zhao,W., Bao,Y., Xiao,J. and Zhang,Z. (2018) MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res.*, **46**, D288–D295.
32. Li,M., Zou,D., Li,Z., Gao,R., Sang,J., Zhang,Y., Li,R., Xia,L., Zhang,T., Niu,G. *et al.* (2019) EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.*, **47**, D983–D988.
33. Xiong,Z., Li,M., Yang,F., Ma,Y., Sang,J., Li,R., Li,Z., Zhang,Z. and Bao,Y. (2020) EWAS Data Hub: a resource of DNA methylation array data and metadata. *Nucleic Acids Res.*, **48**, D890–D895.
34. CNCB-NGDC Members and Partners. (2021) Database resources of the national genomics data center, china national center for bioinformation in 2021. *Nucleic Acids Res.*, **49**, D18–D28.