


DATABASE

Open Access



# GLEANER: a web server for GermLine cycle Expression ANalysis and Epigenetic Roadmap visualization

Shiyang Zeng, Yuwei Hua, Yong Zhang, Guifen Liu\* and Chengchen Zhao\* 

\*Correspondence:  
gfliu0806@tongji.edu.cn;  
cczhao@tongji.edu.cn  
Institute for Regenerative  
Medicine, Shanghai East  
Hospital, Shanghai Key  
Laboratory of Signaling  
and Disease Research,  
Frontier Science Center  
for Stem Cell Research,  
School of Life Science  
and Technology, Tongji  
University, Shanghai 200092,  
China

## Abstract

**Background:** Germline cells are important carriers of genetic and epigenetic information transmitted across generations in mammals. During the mammalian germline cell development cycle (i.e., the germline cycle), cell potency changes cyclically, accompanied by dynamic transcriptional changes and epigenetic reprogramming. Recently, to understand these dynamic and regulatory mechanisms, multiomic analyses, including transcriptomic and epigenomic analyses of DNA methylation, chromatin accessibility and histone modifications of germline cells, have been performed for different stages in human and mouse germline cycles. However, the long time span of the germline cycle and material scarcity of germline cells have largely limited the understanding of these dynamic characteristic changes. A tool that integrates the existing multiomics data and visualizes the overall continuous dynamic trends in the germline cycle can partially overcome such limitations.

**Results:** Here, we present GLEANER, a web server for GermLine cycle Expression ANalysis and Epigenetics Roadmap visualization. GLEANER provides a comprehensive collection of the transcriptome, DNA methylome, chromatin accessibility, and H3K4me3, H3K27me3, and H3K9me3 histone modification characteristics in human and mouse germline cycles. For each input gene, GLEANER shows the integrative analysis results of its transcriptional and epigenetic features, the genes with correlated transcriptional changes, and the overall continuous dynamic trends in the germline cycle. We further used two case studies to demonstrate the detailed functionality of GLEANER and highlighted that it can provide valuable clues to the epigenetic regulation mechanisms in the genetic and epigenetic information transmitted during the germline cycle.

**Conclusions:** To the best of our knowledge, GLEANER is the first web server dedicated to the analysis and visualization of multiomics data related to the mammalian germline cycle. GLEANER is freely available at <http://compbio-zhanglab.org/GLEANER>.

**Keywords:** Bioinformatics, Epigenetics, Germline cycle



## Background

In mammals, genetic and epigenetic information is transmitted across generations through the germline cell development cycle (i.e., germline cycle) [1]. The germline cycle runs through the whole process of ontogenesis and contains three phases: preimplantation embryogenesis, primordial germ cell (PGC) development in postimplantation embryos and gametogenesis in individuals after birth [2]. During these three phases of the germline cycle, the cell potency of germline cells changes cyclically, accompanied by dynamic transcriptional changes and epigenetic reprogramming. A systematic study of these cyclical changes in the mammalian germline cycle will help us to understand the mechanism of genetic and epigenetic information transmission across generations.

In the past decade, studies that characterized the dynamics of transcriptional and epigenetic features reported the local relationships of these dynamics at different phases of the mammalian germline cycle by using high-throughput sequencing technologies. For example, in the preimplantation embryogenesis phase of the mouse germline cycle, Liu et al. reported that transcriptional activation is related to increasing H3K4me3 distribution in gene promoter regions [3]. In the PGC development phase of the mouse germline cycle, Lesch et al. reported that the transcription activated epigenetic marker H3K4me3 was established and that the repressive epigenetic marker H3K27me3 was removed in the active gene promoters [4]. Recent studies further proved the continuity of the germline cycle by linking adjacent phases; for example, it has been reported that global epigenetic reprogramming of the genome occurs during the preimplantation embryogenesis phase and early PGC development phase in the mammalian germline cycle [5, 6]. Both of these waves of epigenetic reprogramming result in a decrease in global DNA methylation along with epigenetic modification re-establishment to regulate transcription [7–9]. However, these findings are still difficult to extend to the whole germline cycle to explain the inheritance and transmission of genetic information across generations due to the following two limitations. First, owing to the large time span of the germline cycle, previous studies profiled only the transcriptome and epigenome in one or two phases of the germline cycle and lacked an understanding of gene expression dynamics and epigenetic regulation during the whole germline cycle. Second, several important and rare cell types in the mammalian germline cycle (especially migrating PGCs in the early gestational stages in humans and mice and prospermatogonia and mitotic oocytes in the late gestational stage in humans) are difficult to collect and characterize, so transcriptome and epigenome profiles are usually incomplete, leading to a deficiency in continuous dynamic trends of the transcriptome and epigenome during the whole germline cycle. Taking advantage of the continuity and integrality of the germline cycle, it is feasible to overcome these two limitations by integrating the existing transcriptional and epigenetic high-throughput sequencing data in the germline cycle and building an online web server that provides multiomics features, the genes with correlated changes in transcription and the overall continuous dynamic trends of the transcriptome and epigenome for an inputted gene. However, to the best of our knowledge, no such web server is currently available.

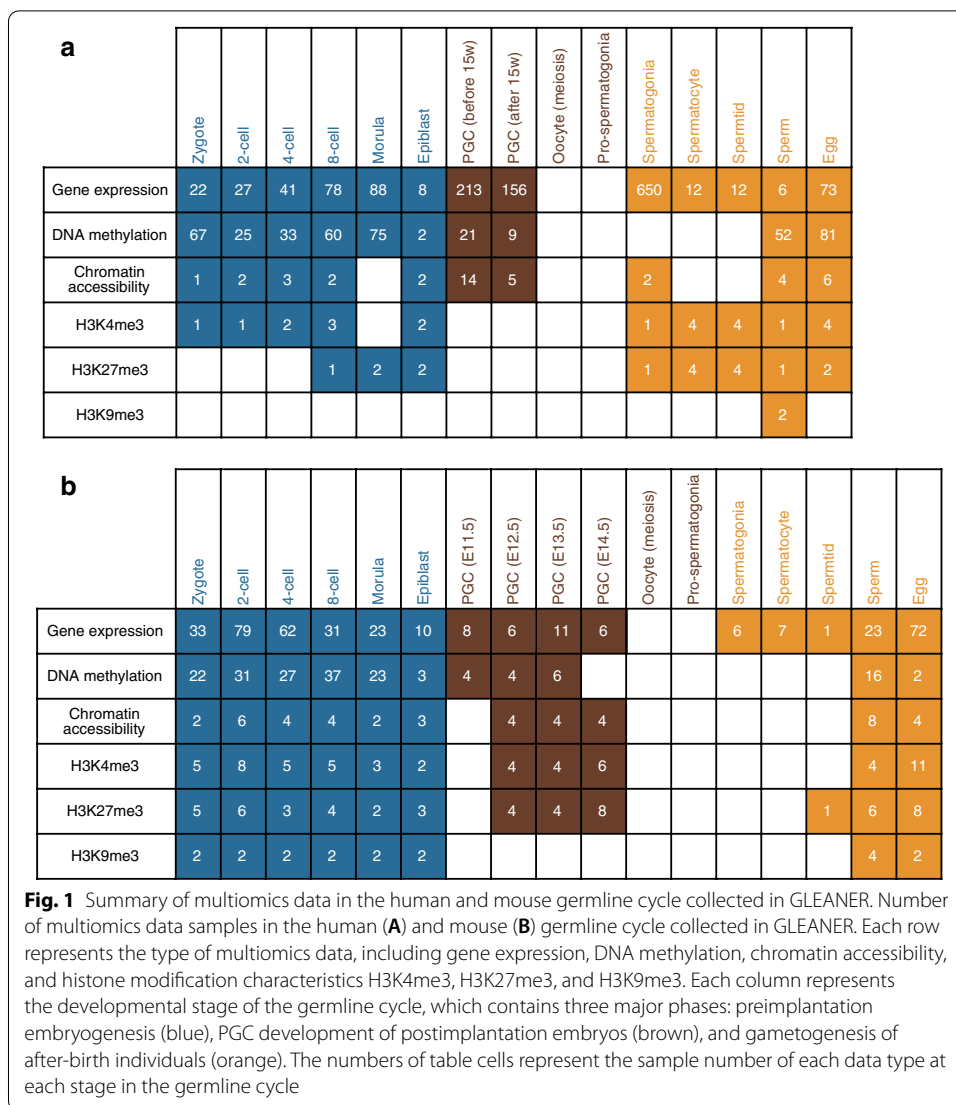
Here, we present GLEANER, a web server for human and mouse germline cycle expression analysis and epigenetic roadmap visualization, which integrates and provides visualization of transcriptional and epigenetic features in the germline cycle. GLEANER

collected 1,764 RNA sequencing (RNA-seq) samples, 600 bisulfite sequencing (BS-seq) samples characterized for DNA methylation, 25 DNase I hypersensitive sites sequencing (DNase-seq) samples and 61 ATAC-seq samples characterized for chromatin accessibility, and 151 ChIP sequencing (ChIP-seq) samples and 20 Cleavage Under Targets & Release Using Nuclease (CUT&RUN) samples characterized for histone modification (80 H3K4me3 samples, 71 H3K27me3 samples, and 20 H3K9me3 samples) in the three phases of the human and mouse germline cycle. GLEANER takes a gene or a genomic region as input and reports the following outputs: (1) integrated analysis of the transcriptional and epigenetic dynamics of the input element in the germline cycle; (2) the genes whose transcriptional changes are most positively or negatively correlated with those of the input element, as well as the integrated data matrix download; (3) overall continuous change trends of transcriptional and epigenetic features at the different stages of germline cycle, including the uncharacterized intermediate cell stages. In summary, GLEANER unifies the transcriptional and epigenetic feature visualization of three phases of the germline cycle and provides similarity analysis and regression analysis, which can yield valuable clues for understanding the transmission mechanisms and regulatory functions of epigenetic modification in the mammalian germline cycle.

## Construction and content

### Data collection and preprocessing

The high-throughput sequencing datasets characterized typical omics (i.e., transcriptomics, epigenomics of DNA methylation, chromatin accessibility, and three types of histone modifications) in mouse and human germline cycle are collected from a public database (Gene Expression Omnibus, GEO) [10, 11] (Fig. 1). To represent and investigate the transcriptional regulatory effects of epigenetic modifications, we focused on the gene promoter regions, defined as those regions  $-2000$  base pairs (bp) and  $+500$  bp around the transcription start sites. For these datasets with preprocessed data available, the preprocessed expression levels or the signals of epigenetic features were adopted. For those datasets that had only raw data available, raw reads were first trimmed using TrimGalore and then mapped to the reference genome (mm9 for mouse and hg19 for human). For RNA-seq data, the sequenced reads were mapped to the reference genome using TopHat (v2.1.1) [12] with default parameters. To make the expressions comparable between different samples, expression levels were measured as TPM values. For BS-seq data, BSMAP [13] was used to map the sequenced reads to the reference genome with parameters “-n 1 -r 0 -s 16” for WGBS data and parameters “-n 1 -r 0 -s 12 -D C-CGG” for RRBS data. The DNA methylation levels were quantified by using the mcall function in MOABS [14]. For ChIP-seq and CUT&RUN data, sequenced reads were mapped to reference genome using bowtie2 [15] with default parameter. The signals of epigenetic features were calculated from the piled-up reads using MACS [16] with parameters “-SPMR”, which generate profiling signals normalized to 1 million reads. For DNase-seq data and ATAC-seq data, bowtie2 was used to map the sequenced reads to the reference genome with default parameters. MACS pileup function was used to calculate the chromatin openness signal by using filtered fragments (less than 120 bp) with “-ext-size=50 -SPMR” to normalize signals to 1 million reads background. Transcriptional and epigenetic features from biological replicates were averaged for subsequent analysis.



GLEANER also provides a visualized and downloadable page of these collected data including a data summary, the public accession numbers, and the related publication information.

**Data normalization**

Samples from different researches may have batch effects. For different types of data, we used different strategies to eliminate the effects of different data sources or experimental batches. For the chromatin accessibility data and histone modification data, signal profiles were normalized per one million reads and calculated on genomic regions by using parameters “-SPMR” in MACS to control the sequencing depth to obtain relatively comparable values among different cell types. For DNA methylation data from BS-seq, due to the quantitative methylation levels calculation above, no more extra normalization processes were performed.

### Correlated changing genes selection

To select the genes with correlated transcriptional changes, we calculated the Pearson correlation coefficient (PCC) and the correlation test  $p$ -value between genes based on the expression levels in the germline cycle. The correlation calculations were performed using the R package Hmisc. Genes with a  $p$ -value  $< 0.01$  and a  $PCC > 0.8$  were selected as the positively correlated candidates, and genes with a  $p$ -value  $< 0.01$  and a  $PCC < -0.8$  were selected as the negatively correlated candidates.

### Regression model construction and overall continuous dynamic trend visualization

For each gene, different regression models, which included linear model, logarithmic model, and polynomial (quadratic, cubic, and quartic) model, were constructed and implemented by the ECharts Statistics module using the gene expression level or the average epigenetic signals within the gene promoter. With these regression models and their specific parameters, the overall continuous dynamic trends of the input transcriptional level or epigenetic features among the developmental stages in the mammalian germline cycle were demonstrated. The mean square errors of the predictions were calculated and displayed as a validity estimation measurement.

### Webserver implementation

In this study, the LAMP (i.e., Linux operating system, Apache web server, MySQL database, and PHP programming language) architecture was used to build an online platform (Additional file 1: Fig. 1). Through the data preprocessing above, we obtained normalized transcriptional and epigenetic features in the mouse and human germline cycles. A MySQL database was built as the back end of GLEANER. Programmed PHP scripts were used to construct the front end of the web page and connect with the back-end database. Subsequently, we used the open-source web framework Bootstrap for interactive interface establishment. The ECharts [17] library in JavaScript was used for statistical calculation and feature visualization in GLEANER.

### Utility and discussion

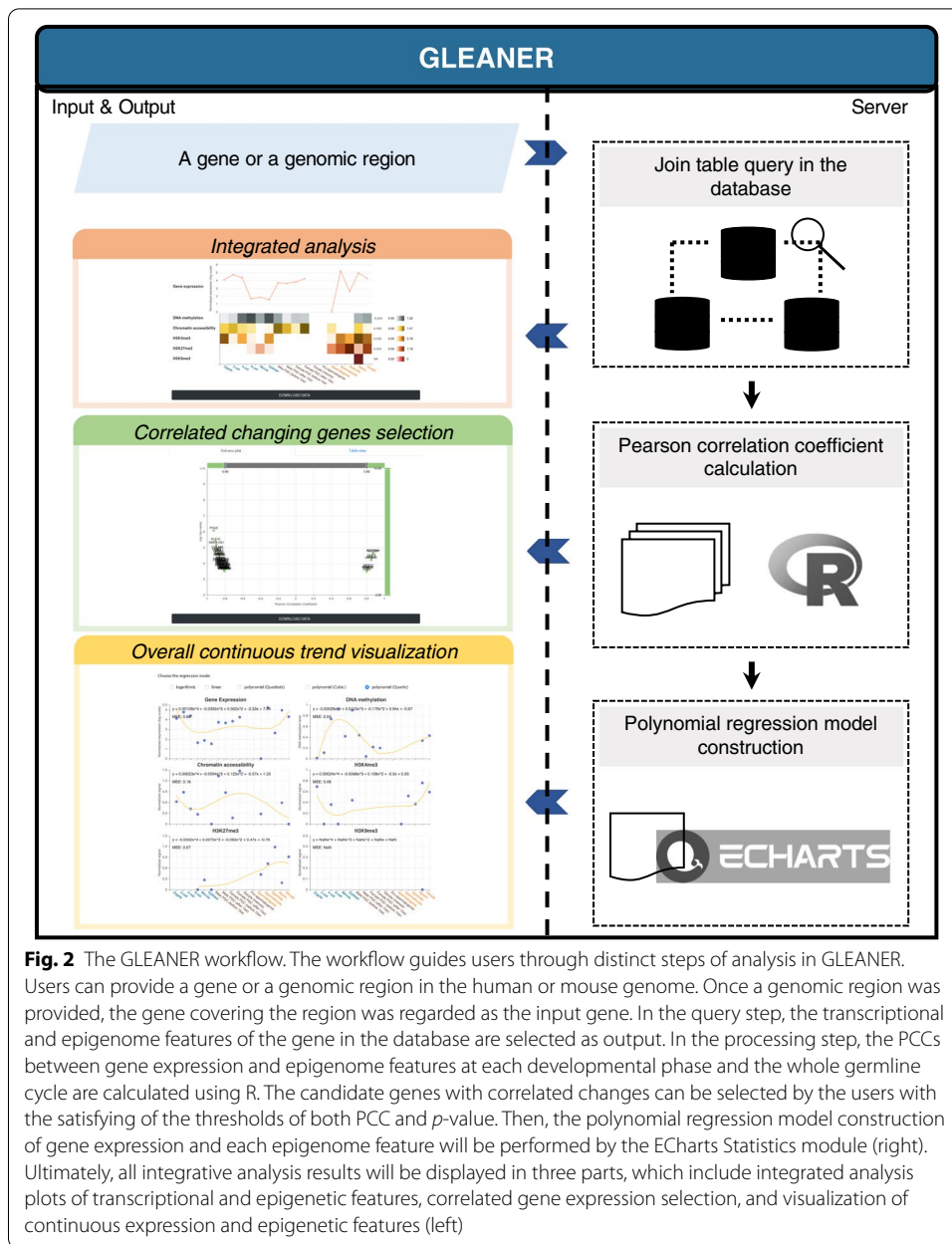
The protocol of GLEANER taking a gene or a genomic region as input for integrated analysis of the transcriptional and epigenetic dynamics, selection of correlated genes, and overall continuous trend visualization during the mammalian germline cycle is presented as a workflow in Fig. 2.

#### *Input*

Users can begin a query by inputting a gene name in different formats (i.e., official gene symbol and RefSeq ID) or a genomic region less than 10,000 bp in the format of “chromosome:start–end”. For gene name input, GLEANER recognizes the format type for subsequent analysis. For genomic region input, the gene within the region will be used for subsequent analysis.

#### *Integrated analysis*

To reflect the dynamics, integrality, and continuity in the mammalian germline cycle, an integrated analysis of the input gene’s expression levels, average DNA methylation levels,



average chromatin accessibility signals, and histone modification features on promoters is conducted and displayed as a line plot combined with heatmaps. Separate PCCs in three phases and the overall PCCs between these features are calculated to provide clues about epigenetic regulation and interaction. This information can be downloaded by clicking the “DOWNLOAD DATA” button.

### **Correlated gene selection**

Genes with similar transcriptional dynamics trends in the germline cycle may be coexpressed or share similar regulatory mechanisms. To select these correlated genes, GLEANER used the preprocessed annotation matrixes and calculated the PCCs of gene expression levels between genes. For each type of data, the PCCs and  $p$ -values of correlation tests are displayed as a volcano plot and a table including the significant positively and negatively correlated candidates that satisfy the adjustable thresholds of both PCC and  $p$ -value. To benefit the users, a button, which allows the download of the underlying data being visualized can be found at the bottom of this panel.

### **Overall continuous trend visualization**

According to the continuity and cyclicity of the germline cycle, the examination of the overall continuous trends of transcriptional and epigenetic features is an effective and feasible way to investigate the transmission of genetic information between generations. Different types of regression models, which included linear model, logarithmic model, and polynomial (quadratic, cubic, and quartic) model, were built to fit the separate points across the time-series data in different stages and demonstrate the overall continuous trend of transcriptional and epigenetic features (see Methods and Materials for details). This trend was visualized as a line plot with points at distinct stages, including those cell types with unavailable data in the germline cycle. It is worth noting that signal values of ChIP-seq samples are not suitable for direct quantitative comparison due to the semi-quantitative characteristic of ChIP-seq if important parameters are unknown.

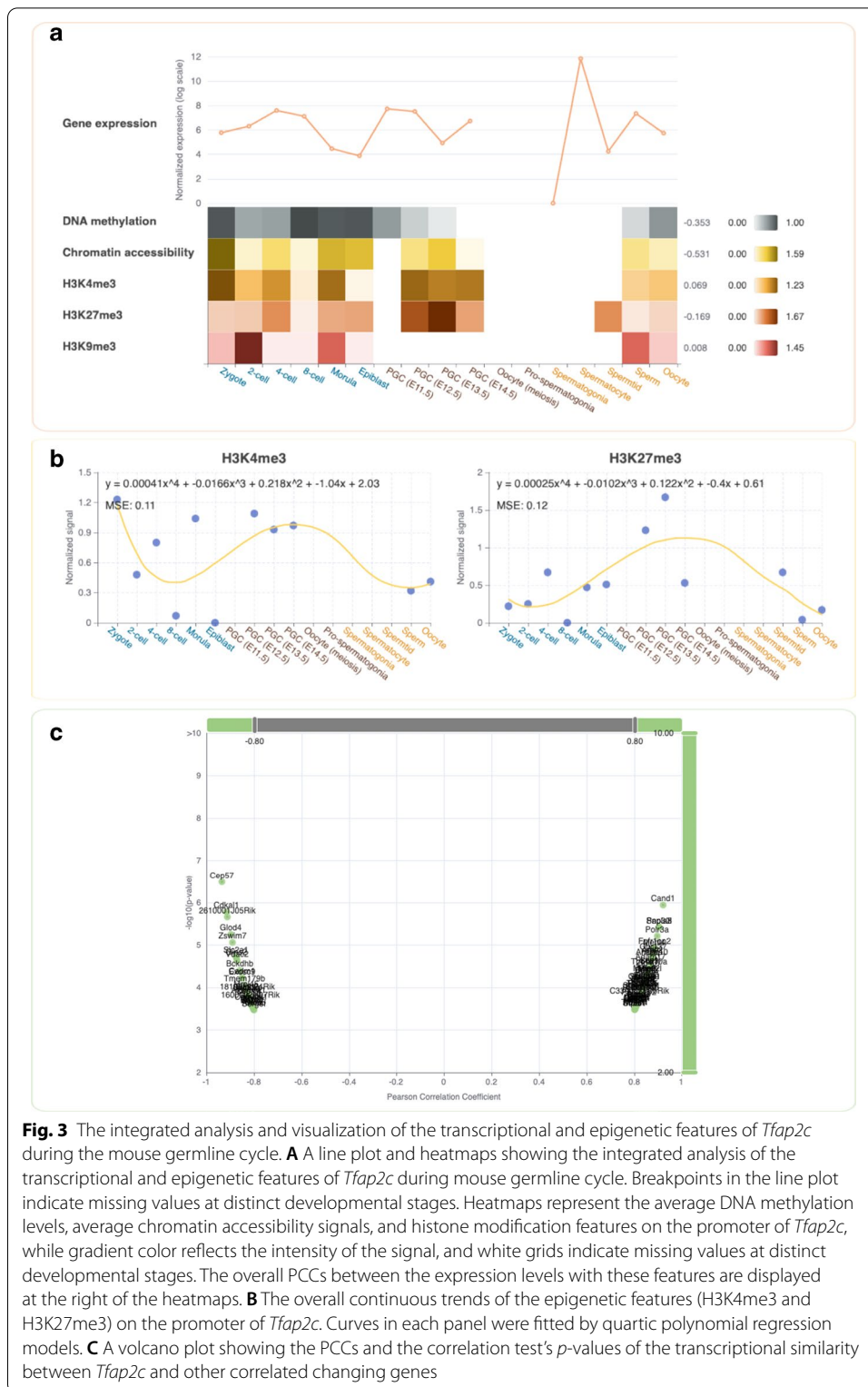
In summary, in a general workflow, GLEANER takes a gene or a genomic region as input to start a query of the transcriptional and epigenetic features in the human or mouse germline cycle. GLEANER conducts the analysis, and visualizes the results in three modules, namely, the integrated analysis of gene expression levels and epigenetic characteristics on the corresponding regulatory region, the identification of genes with correlated changes and the visualization of the overall continuous dynamic transcriptional and epigenetic trends in the germline cycle. In the cluster page, GLEANER also navigates the similarity of input genes globally by representing them in the clusters of transactional and each epigenetic feature.

### **Application case 1**

In this section, a tripartite transcription factor network (PRDM1, PRDM14, and AP2 $\gamma$ ) [18] for the specification of mouse early germline cells is chosen as an example to show the functionality of GLEANER and to discuss the results in detail. It has been reported that PRDM1 and PRDM14 are mutually interdependent and induce the expression of the transcription factor AP2 $\gamma$  in mouse PGC development [19, 20]. This network upregulates the expression of pluripotency genes, such as *Nanog*, *Pou5f1*, and *Sox2*, leading to the initiation of epigenetic reprogramming during the specification and migration process in the mouse PGC development phase [21]. However, the epigenetic regulatory effect of this network in the PGC phase remains unclear.

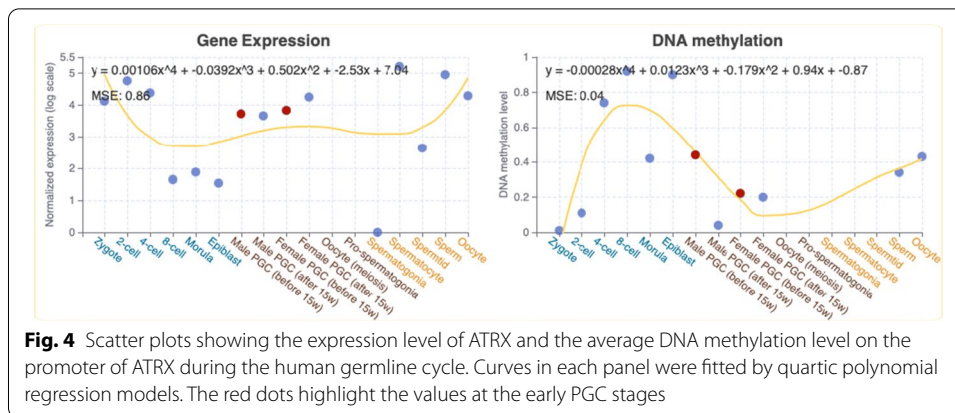
To investigate the potential function of epigenetic information transmission across the germline cycle phases, the coding genes of this network (*Prdm1*, *Prdm14*, and *Tfap2c*) were queried as inputs of GLEANER to explore the dynamic changes and information





about potential epigenetic regulation. GLEANER first conducted an integrated analysis of the transcriptional levels and epigenetic features. In the PGC development phase, *Prdm1*, *Prdm14*, and *Tfap2c* were expressed from Embryonic Day 11.5 (E11.5) to E14.5, which was consistent with previous studies [22]. The expression trend of *Tfap2c* was





positively correlated to the dynamic trend of the H3K4me3 signal of its promoter region (PCC = 0.894) (Fig. 3A) and negatively correlated to the dynamic trend of the H3K27me3 signal of its promoter region (PCC = -0.581). It is worth noting that the DNA methylation level on the promoter region of *Tfap2c* decreased rapidly as well as the H3K27me3 signals increased, which indicated that the dominant regulating epigenetic factors of *Tfap2c*'s expression might have changed during this period. Besides, the increasing H3K4me3 and H3K27me3 signals on the *Tfap2c* promoter from the late preimplantation embryogenesis phase to the PGC phase indicate a potential bivalent regulatory mechanism of AP2γ during these two phases of the mouse germline cycle (Fig. 3B). To further explore the other potential regulatory factors involved in this network, GLEANER selected the genes whose changes most correlated with those of *Tfap2c* at the expression level. As shown in Fig. 3C, *Cand1*, an assembly factor of SKP1-CUL1-F-box protein complexes that related to the regulation of transcription and cell differentiation processes, shows the most positively correlated expression pattern.

### Application case 2

In this section, an X-chromosome reactivation-associated gene (*ATRX*) at the human PGC development stage is chosen as an example to show how GLEANER provides clues about epigenetic regulatory mechanisms during the human germline cycle. It has been reported that X-chromosome reactivation is one of the representative events of the epigenetic reprogramming process during mammalian female PGC development [23–27]. However, studies also observed that the total expression level of the genes on the X chromosome in female PGCs did not reach expected twofold higher than that in male PGCs at a similar stage [28, 29], indicating that reactivation of genes in the X-chromosome of female PGCs is incomplete. A recent study also reported that the expression of X-linked genes on both alleles is transcriptionally reduced but not silenced, representing a compensated state referred to as X-chromosome dampening [30], and *ATRX* was one of these biallelically expressed X-linked genes [31]. By querying *ATRX* in the human germline cycle using GLEANER, we found that the expression level of *ATRX* was 1.07-fold higher in the early female PGCs than in the early male PGCs, which indicated that the reactivation of *ATRX* is also incomplete (Fig. 4). Besides, the *ATRX* gene promoter showed a relatively higher methylation level in the early male PGCs (average methylation

level = 0.44) than in the early female PGCs (average methylation level = 0.22), suggesting that DNA methylation might be related to the incomplete or ongoing reactivation of the *ATRX* gene, which indicated that epigenetic factors were involved in X-chromosome reactivation in the human PGC development process. However, due to the lack of other epigenetic features (i.e., histone modifications), we cannot rule out the possible regulatory effects of these features on the unexpectedly low accessibility of the *ATRX* gene promoter at the early female PGC stage.

## Discussion

Although an increasing amount of multiomics data has been generated in human and mouse germline cycles, to the best of our knowledge, there is no web server to facilitate the integrated analysis and visualization of transcriptional and epigenetic characteristics in the germline cycle. In this study, we developed GLEANER, a web server for germline cycle expression analysis and epigenetics roadmap visualization. Using normalized transcriptional and epigenetic high-throughput sequencing data collected in human and mouse germline cycles, we investigated the genetic and epigenetic information that is transmitted between generations. For each input, GLEANER shows integrated information about the gene transcription, DNA methylation, chromatin accessibility, and histone modification characteristics on its corresponding regulatory region, both positively and negatively correlated changing genes during the germline cycle and the overall continuous dynamic transcriptional and epigenetic trends in the germline cycle. Case studies of a tripartite transcription factor network (PRDM1, PRDM14, and AP2 $\gamma$ ) and an X-chromosome reactivation associated gene (*ATRX*) demonstrated the functional utility of GLEANER and highlighted that the integrative analysis of transcriptional and epigenetic features could provide a new angle to investigate the mechanisms underlying the epigenetic regulatory effects in genetic and epigenetic information transmission during the germline cycle. In addition to visualizing the characteristics of individual genes, users may also be interested in navigating relationships of multiple genes globally. In the cluster page, GLEANER performed clustering analysis for the transcriptional and each epigenetic feature, and also provided the user with the ability to query a list of genes for the clusters they belong to. By highlighting these input genes in the heatmap of the clusters, the user was given a clue to evaluate whether these genes are more inclined to be clustered, or are random.

The long time span of the germline cycle and material scarcity have largely limited the understanding of the genetic and epigenetic information transmitted during this process. GLEANER takes advantage of the continuity and integrality of the germline cycle to partially overcome this limitation through a computational approach. However, the inferred data cannot perfectly represent the actual values in the unavailable cell types in the germline cycle. With further breakthroughs in technology making more data available in the future, GLEANER will be updated to represent the cyclical change of genetic and epigenetic characteristics more completely and present the overall continuous dynamic trends more accurately. On the other hand, future data will provide more various aspects of the integrated analysis. For example, there is still little high-throughput sequencing data available from the PGC development phase (i.e., PGC migration in human and mouse early gestational stages, pro-spermatogonia, and mitotic oocyte

migration in human late gestational stages). With the complement of these unavailable data, a systematic comparison of the transcriptional and epigenetic features between humans and mice can be included in future updates of GLEANER. GLEANER will be maintained and updated periodically to collect and integrate new transcriptional and epigenetic datasets in future updates.

## Conclusions

In summary, by using GLEANER, users can easily obtain integrated information on the transcriptional and epigenetic features in the germline cycle by querying a gene or a genomic region. In addition, through correlation analysis and regression analysis in GLEANER, users could not only obtain candidates with similar transcriptional and epigenetic feature dynamic changes that indicate the potential underlying mechanisms of epigenetic regulation in the mammalian germline cycle but could also predict missing features by extrapolating from the overall continuous dynamic trends of the transcriptional and epigenetic features. Thus we believe that GLEANER could be a valuable tool to help understand the transcriptional and epigenetic information transmitted in the mammalian germline cycle.

## Abbreviations

PGC: Primordial germ cell; BS-seq: Bisulfite sequencing; DNase-seq: DNase I hypersensitive sites sequencing; CHIP-seq: Chromatin immunoprecipitation sequencing; CUT&RUN: Cleavage Under Targets & Release Using Nuclease; GEO: Gene Expression Omnibus; PCC: Pearson correlation coefficient; LAMP: Linux operating system, Apache web server, MySQL database, and PHP programming language.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-021-04217-1>.

**Additional file 1.** Framework of the GLEANER construction.

## Acknowledgements

Not applicable.

## Authors' contributions

S.Z., C.Z., G.L. and Y.Z., conceived the study. S.Z., G.L., and C.Z. wrote the manuscript. S.Z. and C.Z. analyzed the data. S.Z., Y.H., and C.Z. developed the web server. All authors approved the final version of the manuscript.

## Funding

This work has been supported by the National Key Research and Development Program of China (Grant Number 2017YFA0102600), National Natural Science Foundation of China (Grant Numbers 32030022, 31970642, 31721003, 31900491), National Program for Support of Top-notch Young Professionals, China Postdoctoral Science Foundation (Grant Number 2018M642073), Major Program of Development Fund for Shanghai Zhangjiang National Innovation Demonstration Zone (Grant Number ZJ2018-ZD-004), Shanghai Municipal Science and Technology Major Project (Grant Number 2018SHZDZX01), LCNBI, ZJLab. Funding for open access charge: National Natural Science Foundation of China.

## Availability of data and materials

All datasets analyzed during this study are included in this published article and the sources are cited accordingly. The analyzed data are available at GLEANER's data page: <http://compbio-zhanglab.org/GLEANER/data.html>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 31 March 2021 Accepted: 25 May 2021

Published online: 31 May 2021

## References

1. Tang WW, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. *Nat Rev Genet.* 2016;17(10):585–600.
2. Wen L, Tang F. Human germline cell development: from the perspective of single-cell sequencing. *Mol Cell.* 2019;76(2):320–8.
3. Liu X, Wang C, Liu W, Li J, Li C, Kou X, Chen J, Zhao Y, Gao H, Wang H, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature.* 2016;537(7621):558–62.
4. Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc Natl Acad Sci.* 2013;110(40):16061–6.
5. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* 2014;28(8):812–28.
6. Saitou M, Kagiwada S, Kurimoto K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development.* 2012;139(1):15–31.
7. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev.* 2002;117(1–2):15–23.
8. Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, Ogura A, Ishino F. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development.* 2002;129(8):1807–17.
9. Guibert S, Forne T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* 2012;22(4):633–41.
10. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30(1):207–10.
11. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* 2013;41(Database issue):D991–995.
12. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–11.
13. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinform.* 2009;10:232.
14. Sun D, Xi Y, Rodriguez B, Park HJ, Tong P, Meong M, Goodell MA, Li W. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* 2014;15(2):R38.
15. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
16. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.
17. Li D, Mei H, Shen Y, Su S, Zhang W, Wang J, Zu M, Chen W. ECharts: a declarative framework for rapid construction of web-based visualization. *Vis Inform.* 2018;2(2):136–46.
18. Magnusdottir E, Dietmann S, Murakami K, Gunesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim Surani M. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol.* 2013;15(8):905–15.
19. Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet.* 2008;40(8):1016–22.
20. Grabole N, Tischler J, Hackett JA, Kim S, Tang F, Leitch HG, Magnusdottir E, Surani MA. Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* 2013;14(7):629–37.
21. Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development.* 2007;134(14):2627–38.
22. Weber S, Eckert D, Nettersheim D, Gillis AJ, Schafer S, Kuckenberger P, Ehlermann J, Werling U, Biermann K, Looijenga LH, et al. Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol Reprod.* 2010;82(1):214–23.
23. Heard E, Clerc P, Avner P. X-chromosome inactivation in mammals. *Annu Rev Genet.* 1997;31(1):571–610.
24. Tam PP, Zhou SX, Tan SS. X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development.* 1994;120(10):2925–32.
25. Monk M, McLaren A. X-chromosome activity in foetal germ cells of the mouse. *J Embryol Exp Morphol.* 1981;63:75–84.
26. Sugimoto M, Abe K. X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet.* 2007;3(7):e116.
27. Chuva de Sousa Lopes SM, Hayashi K, Shovlin TC, Mifsud W, Surani MA, McLaren A. X chromosome activity in mouse XX primordial germ cells. *PLoS Genet.* 2008;4(2):e30.
28. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell.* 2015;161(6):1437–52.
29. Li L, Dong J, Yan L, Yong J, Liu X, Hu Y, Fan X, Wu X, Guo H, Wang X, et al. Single-cell RNA-Seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell.* 2017;20(6):891–2.
30. Vertesy A, Arindartoro W, Roost MS, Reinius B, Torrens-Juaneda V, Bialecka M, Moustakas I, Ariyurek Y, Kuijk E, Mei H, et al. Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nat Commun.* 2018;9(1):1873.

31. Chitiashvili T, Dror I, Kim R, Hsu FM, Chaudhari R, Pandolfi E, Chen D, Liebscher S, Schenke-Layland K, Plath K, et al. Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat Cell Biol.* 2020;22(12):1436–46.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

