Authors for correspondence:
Lei Yang
e-mail: yanglei_hmu@163.com
Yongchun Zuo
e-mail: yczuo@imu.edu.cn

EmExplorer: a database for exploring time activation of gene expression in mammalian embryos

Bosu Hu[1,†], Lei Zheng[1,†], Chunshen Long[1], Mingmin Song[1], Tao Li[2], Lei Yang[3] and Yongchun Zuo[1]

[1]State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, College of Life Sciences, Inner Mongolia University, Hohhot 010070, People's Republic of China
[2]College of Life Sciences, Inner Mongolia Agricultural University, Hohhot 010018, People's Republic of China
[3]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, People's Republic of China

LZ, 0000-0002-8531-6949; YZ, 0000-0002-6065-7835

Understanding early development offers a striking opportunity to investigate genetic disease, stem cell and assisted reproductive technology. Recent advances in high-throughput sequencing technology have led to the rising influx of omics data, which have rapidly boosted our understanding of mammalian developmental mechanisms. Here, we review the database EmExplorer (a database for exploring time activation of gene expression in mammalian embryos), which systematically organizes the genes from development-related pathways, and which we have already established and continue to update it. The current version of EmExplorer incorporates over 26 000 genes obtained from 306 functional pathways in five species. The function annotations of development-related genes were also integrated into EmExplorer. To facilitate data extraction, the database also contains the following information. (i) The dynamic expression values for each development stage are matched to the corresponding genes. (ii) A two-layer search tool which supports multi-option searching, such as by official symbol, pathway name and function annotation. The returned entries can directly link to the analysis results for the corresponding gene or pathway in the analysis module. (iii) The analysis module provides different gene comparisons at the multi-species level and functional pathway level, which shows the species specificity and stage specificity at the gene or pathway level. (iv) The analysis based on the hypergeometric distribution test reveals the enrichment of gene functions at a particular stage of one organism's pathway. (v) The browser is designed for users with ambiguous searching goals and greatly helps new users to get a general idea of the contents of the database. (vi) The experimentally validated pathways are manually curated and shown on the home page. EmExplorer will be helpful for elucidating early developmental mechanisms and exploring time activation genes. EmExplorer is freely available at http://bioinfor.imu.edu.cn/emexplorer.

1. Introduction

Preimplantation embryo is the initial stage of mammalian embryonic development, the process of which is rather complex. During this stage, the networks comprising various molecular interactions have a critical role during a particular stage of embryogenesis [1]. Most mammalian preimplantation embryonic developments are extremely similar and highly ordered. During the early embryo process, the single-cell zygote continues to proliferate and finally forms a blastocyst, which is composed of a certain number of cells. Mammalian preimplantation embryonic development consists of a series of important

events, such as oocyte maturation, the first cell mitosis, maternal to zygotic transition, embryonic genome activation and the cell fate decision [2–5]. Fertilization is thought to be controlled by specialized gene regulatory programmes which stringently control different gene expressions during each development stage [6–9]. The significance of the studies in the preimplantation field was previously emphasized at the molecular level, which demonstrated that the transcriptome in the early development stage plays critical roles in controlling and maintaining the cellular identity. The reprogramming implementation also relies on a deep understanding of the mRNA transcriptome [10].

With the wide use of high-throughput sequencing technology, knowledge of the expression data of early embryonic development genes has increased significantly [11]. After the first transcriptome profiling obtained on Genome Array, dynamic transcriptomes from metaphase II oocytes to blastocyst, which is the terminal stage of preimplantation embryos, were successfully tested [12]. The data obtained from this technology are described as the timing sequence expression value. The experimental data greatly help researchers to decipher the regulatory networks that contribute to morphological changes in embryo development. Data analysis has demonstrated that the major maternal to zygotic transition begins from the four-cell to eight-cell stages, during which the largest genes of the zygotic genome were activated [13]. Biologists are finding effective methods to collect and classify the larger number of genes. With a highly ordered frame we can make comparisons between diverse genes and can even make comparisons at a cross-species level [14]. Through these comparisons, researchers may find stage-specific expression genes of a particular species. Furthermore, it seems to be of great value to extract species-specific expression genes.

Hitherto a few online databases have been developed for in-depth mining of development-related genes and their dynamic expression at different stages of development. Previous studies have also suggested that analysis of the gene networks and the molecular pathways at the transcriptome level may offer an entry point to the understanding of the timing mechanism [15]. Further analysis of the transcriptome revealed several patterns of changes in the transcriptional activity during the transition through compaction and blastocyst formation [8]. There are already several existing databases, such as DevMouse [16], GED [17], DBTMEE [18], MetaImprint [19] and EMAGE [1]. The time-ordered data in these online sources vividly shows the changes in each stage. However, all the databases above have the following problems: (i) they are mainly developed for mouse embryo development, so that a database for multiple species urgently needs to be established; (ii) genes are still discrete units in these datasets as they cannot be well ordered only in the timing sequence; (iii) the specificity of gene expression is closely related to its function, however there are few published databases in this field; (iv) the current online databases have not managed to make comparisons between different species. Thus, the new challenge is to implement a highly efficient organization method between genes cross multiple species. Our research team had previously reported the timing of genome-wide activation of the Kyoto Encyclopedia of Genes and Genomes (KEGG) functional pathway in bovine preimplantation embryogenesis [20]. The results have confirmed the key functional pathways of bovine embryos, including the organelle-related functional pathway, junction pathways and receptor pathways,

and that they undergo a successive time-ordered activation [21,22]. Thus, it seemed that gathering diverse genes through different functional pathways is an efficient way to study the network between different genes. Furthermore, the functional pathways have great similarity across different mammals. So collecting, categorizing and analysing each functional pathway not only can sort diverse genes, but also seems to be a way to achieve comparisons among different species.

To our best of our knowledge, EmExplorer is the first Web-based multi-species resource focused on the temporal expression of development-related genes. The latest release of EmExplorer provides a clear molecular pathway, function annotation and dynamic expression information across different stages of preimplantation development. In addition, it offers an online analysis tool to search, browse and visualize the developmental genes with their dynamic expressions during different development stages. Users can also make expression comparisons for one specific gene or functional pathway between different species by using the online tool. EmExplorer can help researchers to understand the timing expression of specific genes in each developmental stage, which may inspire biologists and enable them to carry out further studies in cell reprograming or embryo development. This database will help with further study in assisted reproductive technology and regenerative medicine.

2. Material and method

2.1. Derivation of development-related genes

The current release of EmExplorer was designed to provide early embryonic development gene-related information regarding the pathway, basic information, function annotation and the most significant temporal expression value. As shown in figure 1, the KEGG database was chosen as the original pathway data source. It was searched with specific key words, by inputting 'embryonic development', for different species. We finally got an average of 300 relevant pathways for each organism after filtering the pathway dataset manually. The current version contains five mammals, and the detailed numbers of pathways for each species are illustrated in table 1. In total, the five mammals contain 306 functional pathways. By mining the genes under the pathways for each species, we collected 13 000+ development-related genes (include 3200+ common genes and species-specific genes) for five species, including 5098 *Homo sapiens* (human) genes, 5874 *Mus musculus* (mouse) genes, 6073 *Bos taurus* (cattle) genes, 5565 *Sus scrofa* (wild boar) genes and 4290 *Macaca mulatta* (rhesus monkey) genes. The Latin name for organisms in EmExplorer corresponds to the Latin name in the National Center for Biotechnology Information (NCBI) and KEGG databases, which are specific strains for each species. The detailed information for all genes was referenced against the information obtained from NCBI. The genes were further annotated by the corresponding terms stored in Gene Ontology. The integration of function annotations provides more comprehensive information. In addition, it may give users insights into the way a gene actually functions.

2.2. Manual annotation at the gene level

We chose PubMed as the original literature data source. It was searched with the keywords for different early development
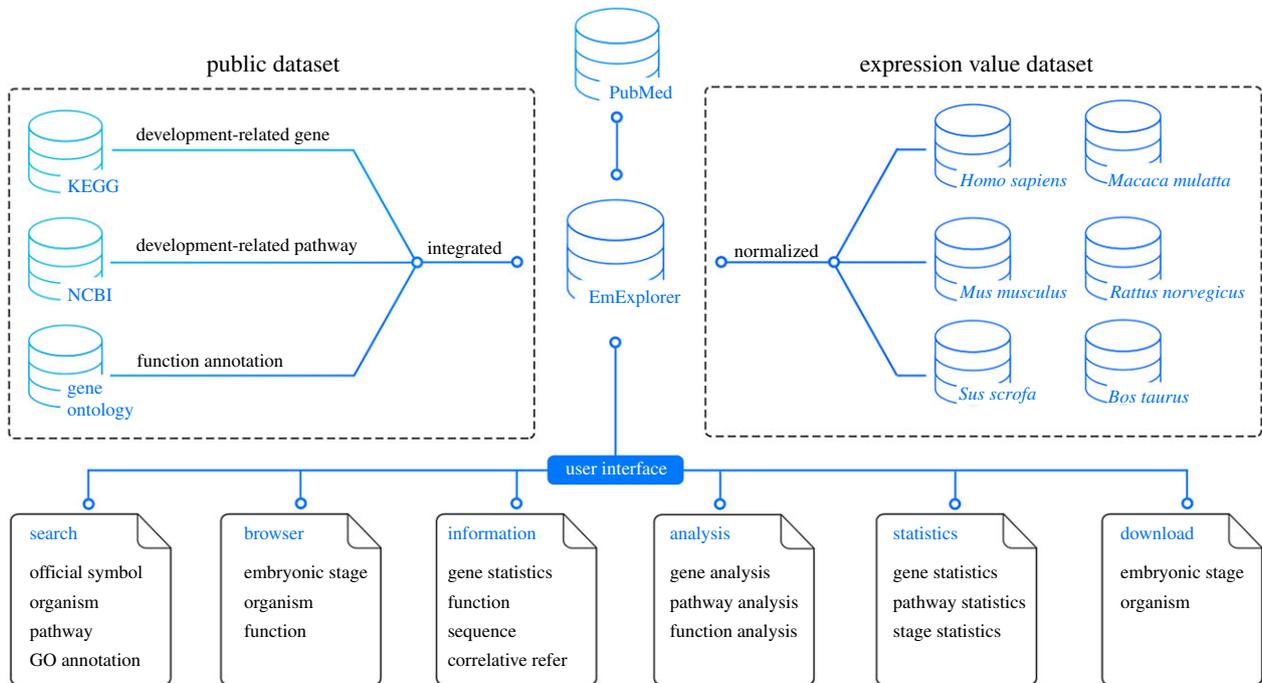
Figure 1. Overview of the establishment process and workflow of EmExplorer. EmExplorer integrates development-associated genes from public resources and sorts them into functional pathways. Users can input key words to the query engine, and the relevant information will be extracted from the database. The analysis tools enable users to make comparisons between genes on different levels. All search and analysis results will be helpful for further analysis.

Table 1. EmExplorer data content and detailed statistics .

species	pathway	gene	function annotation	gene with dynamic expression value	gene with function annotation
Homo sapiens	306	5098	131	3663	399
Macaca mulatta	292	4290	14	3123	24
Mus musculus	302	5874	132	3333	356
Sus scrofa	302	5565	106	3180	182
Bos taurus	302	6073	99	5253	250
total	306	13 040	195		484

stages and species. The search returned in total more than 200 000 published articles and the relevant articles were further filtered by using keywords as follows: 'preimplantation development', 'early embryo', 'pathway', 'genome activation', etc. Finally, we found more than 17 500 relevant manuscripts as the final source, from which we extracted the popular development-related genes. From this study, we further enriched the genes associated with early development and inputted them into EmExplorer. We simultaneously counted the significance and attention of academic studies to these genes, and we took the top 50 genes as the most popular genes and incorporated them in the analysis module (mono-gene analysis tool).

2.3. Data processing

The dataset used in this research originates from papers published previously [12,23–25] and the experimental results from our laboratory (the expression data of pig early embryo), the sources are listed in the references and are also shown in the download module of EmExplorer. However, the expression value of *Sus scrofa* derives from the latest study in our

laboratory, but we have currently decided not to publish it. The previous work matched the gene with its pathways and used functions to annotate it. We then combined the experimental expression data with their corresponding genes.

Directly making comparisons between several genes by using the raw data in the dataset is unjustified. Take housekeeping genes as a case in point, with a low expression level a housekeeping gene can significantly influence the development process [26]. However, most genes with a much higher expression level than housekeeping genes play an insignificant role during early development [27]. In the case of using raw data, the comparison between different species is also impossible. Thus, all these expression data underwent the same normalization procedure, so that they can be compared at the same level [18]. The normalization process is simple but effective

$$\text{norm}_{\text{val}} = \frac{\text{val}_{\text{stage}}}{\sum \text{val}_s},$$

where norm_{val} is the normalized expression value of a gene at a particular stage, $\text{val}_{\text{stage}}$ denotes the gene expression value at a

certain stage and Σval_s represents the sum of the expression values of a certain gene during the entire early development process. Data processing cannot be carried out at a cross-species level. However, y using a normalizing procedure, common genes across different species can be compared at the same level. The dataset was finally stored in MySQL database.

2.4. Identification of a specific pathway

Before discussing figure 2, two concepts need to be clarified. First, we defined the high expression genes (HEGs) at the pathway level as being those genes whose expression value is higher than the average level of all genes in a particular pathway. We called the percentage of HEGs in all genes the high expression gene ratio (HEGr). Second, the mean value of these HEGs was abbreviated as HEGm. We assumed that the specific stage of a pathway shows high HEGr and HEGm, which means the specific stage shows its specificity in these two quantities.

We defined HEGs as having expression values higher than the mean level

$$HEGr = \frac{\sum_{k \in G_p} Z_k}{\sum k \in G_p},$$

where HEGr is the percentage of HEGs in all genes in the same pathway. We denote G_p as the set of genes under a particular functional pathway. Z_k is a binary variable which takes the value of 1 when the gene expression level is above the median of overall genes. We further ranked expression genes in pathways according to their expression value from small to large, and then introduced HEGm, which represents the median of the HEG expression level [28]

$$HEGp = \frac{3\Sigma f}{4}$$

$$\beta = \{HEGp\}, \quad G_L = [HEGp], \quad G_N = [HEGp] + 1,$$
$$HEGm = \beta \cdot Val_{GN} + (1 - \beta) \cdot Val_{GL},$$

where HEGp is the position of the HEG median and f represents the total number of genes in a certain pathway. β is the fractional part of HEGp. G_L is the integer part of HEGp, which corresponds to the ranking number of the gene whose expression is equal to the average in the sorted expression gene list. We, respectively, defined Val_{GN} and Val_{GL} as the expression value of the GN^{th} gene and the GL^{th} gene in a certain development stage. Finally, we can calculate HEGm, which is the median of HEGs. The stage which has a high value of both HEGr and HEGm is denoted as the specific expression stage of a particular pathway.

We attempted to determine the specific expression developmental stage of a gene. The development stages can be divided into three categories: (i) before transition (oocyte, zygote, cell-2 and cell-4); (ii) transitioning (cell-8, called MZT); (iii) after transition (morula and blastocyst, called embryonic genome activation) [21,29]. As only a few genes are activated during the before transition stages, the embryo mostly relies on the maternal substance deposited in the oocyte cytoplasm [30]. In the after transition stage, the embryo can maintain its crucial biological process. A previous study suggested that genes that are highly expressed in this period are closely related to morphological development

[31]. The obvious changes occur during the transition stage, when the majority of oogenic products are lost due to degradation [13,32]. Previous studies have indicated that the expression tendency of the functional pathways, which can be obtained by calculating the overall trend of genes in a particular pathway, also presents temporal specificity [1,33–36]. Based on the literature collected from PubMed, we finally extracted 36 experimentally validated stage-specific pathways, which are presented in the home page of EmExplorer (figure 2a), and, respectively, classified them into seven development stages (table 2). Finally, data collation, analysis of the expression value and the pathways of individual genes in the dataset were completed.

2.5. Functional enrichment analysis

Genes expressed specifically at a certain development stage may strongly relate to a biological function. Deeply mining the common functional effects of a group of specific expression genes may reveal that some development-related functions were enriched in a certain pathway in each stage. This part of the analysis is based on the hypergeometric test and test at the pathway level

$$p = \sum_{k=0}^{K-1} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n},$$

where N is the total number genes of a function pathway; M represents the number of specific expression genes at a particular stage; and n is the total number of gene functions in this pathway. k is defined as the number of specific genes during a stage which has a certain function. The p-value was finally corrected to obtain the q-value. If the q-value of a function is less than 0.05, we could assert that these stage-specific genes of the pathway are significantly related to performing a particular biological function. The analysis results are shown in the analysis page for each pathway, as this helps users to understand what the aim of these specific expression genes is.

2.6. System design and implementation

EmExplorer is based on LAMP architecture: Linux operating system, Apache HTTP server, MySQL database management system, and the Web platform based on the PHP background scripting language [37]. The thinkPHP framework was used for web applications, as its strong PHP subordinate code can guarantee the stability and performance of EmExplorer Web services. The client is built on the bootstrap front-end framework to implement a user-friendly interface, and can also implement any functionality on the PC side as on the same end of a mobile device. Java Script implements visualization of statistical data in Highcharts, and users can export images in different formats. EmExplorer can be accessed at http://bioinfor.imu.edu.cn/emexplorer/.

3. Database use and access

3.1. Overview of the usage of EmExplorer

The database supports five basic operations: home, search, browser, analysis and download. As shown in figure 3, users can obtain detailed information on
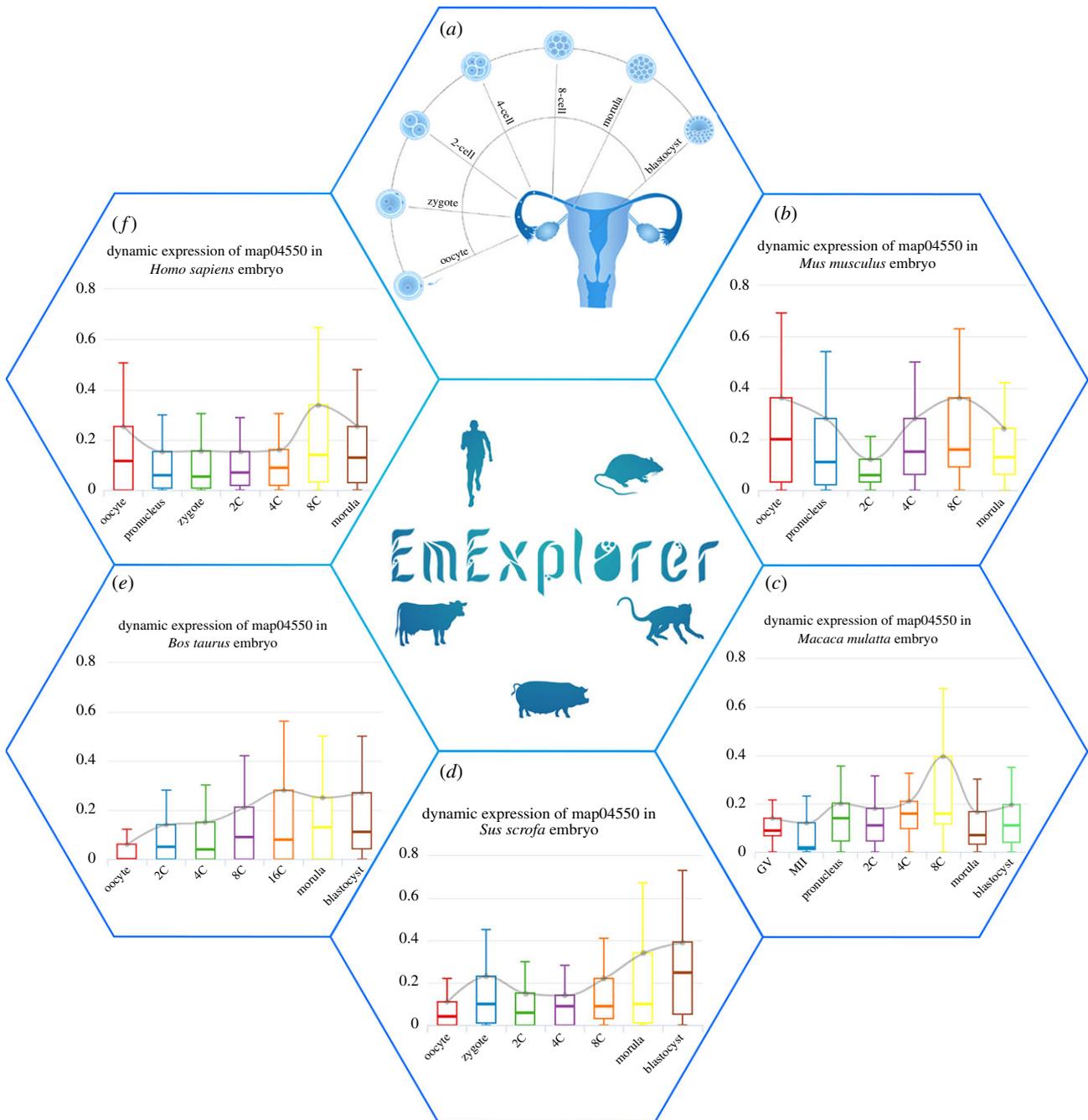
Figure 2. The analysis of the pluripotent stem cell (PSC) regulation pathway at the pathway level. (*a*) The experimentally validated information is provided on the home page. Each developmental stage contains several stage-specific functional pathways. We selected one functional pathway, which was confirmed by previous experiments and was the same as the results we obtained after data processing, at each stage. The current mainstream browsers can completely support the special effects of the home page. (*b*–*f*) The HEGs are defined as the group of genes whose expression values in a stage are higher than the median of the overall genes in a certain pathway. The median expression value of HEGs represents the overall level of HEGs. As the figure shows, the analysis result of the PSC regulation pathway for multiple species indicates that the common feature of these five organisms is that genes in the PSC pathway are rapidly raised at the eight-cell stage and decreased when it comes to the morula stage. The eight-cell phase may a key node for cellular pluripotency activation. Except for human and mouse, the temporal expression in the other three organisms reaches its lowest level at the oocyte stage. Such a difference between human and mouse and the other three species relates to the differences in biological functions performed in this stage.

development-related genes by exercising user-specific options. The search results are presented in a table, and users are able to click the genes in the table and be led to the website for comprehensive information. The browser was designed for new users or users who do not have a clear searching target to have a quick view of EmExplorer. Three analysis tools can be used to make comparisons at the gene level, pathway level and species level. All of the dataset hosted in EmExplorer can be downloaded to users' PC for further analysis.

3.2. Search

EmExplorer provides users with a convenient two-layer search tool that allows users to search directly for the records of the genes. The flexible search engine, based on a MySQL back-end, provides several query parameters, such as database ID, official symbol, pathway name, function annotation, etc. Fuzzy match was masterly applied in the module, so users need not be concerned about being off-target when they use search tool. For example, if visitors input 'IGF' in the

Table 2. The experimentally validated content in each developmental stage.

stage		*Homo sapiens*	*Macaca mulatta*	*Mus musculus*	*Bos taurus*	*Sus scrofa*
oocyte	pathway	2	0	2	2	1
	gene	18	0	22	11	11
	function	21	0	22	14	16
zygote	pathway	10	4	9	9	8
	gene	77	3	81	55	42
	function	48	6	52	46	42
2-cell	pathway	3	2	3	3	2
	gene	51	6	34	20	18
	function	53	7	43	38	38
4-cell	pathway	15	6	15	15	15
	gene	107	4	107	76	59
	function	6	3	56	40	54
8-cell	pathway	4	1	5	4	4
	gene	38	1	43	30	23
	function	31	2	34	27	26
morula	pathway	7	0	7	4	5
	gene	11	0	11	5	6
	function	11	0	11	6	7
blastocyst	pathway	7	3	7	7	7
	gene	74	2	64	55	31
	function	47	4	46	39	35

query box, the result contains 'IGF2', 'IGF2r', 'IGF3', etc., which as much as possible covers all possibilities. The quick search results are shown in figure 3; the result table summarizes the basic information of the records in the dataset which conforms to the key words. The links on these genes would bring users to the detailed information for the corresponding development-related genes. The marks on each entry link to the analysis result of the corresponding gene in the analysis module. When a query contains a lot of information, a great number of entries may return, which really frustrates users who want to quickly retrieve the information they need. In this regard, the second layer of this tool can be used for further selecting to get the most useful terms quickly. If a gene has multiple entries in the returning table, it probably means the gene has pluripotency or the gene exists in different pathways during early development.

The links for each gene have the same format as shown in figure 3, with detailed information consisting of the official symbol, genome location, functional pathway, the biological functions involved, relevant articles and the dynamic expression value ordered in timing sequence. Based on the information offered above, users can have a comprehensive view of a gene in three aspects: (i) whether the gene is located in either a genome or functional pathway; (ii) at which stage its expressing ability is active or shows its strongest expressing activity; (iii) how it influences the whole development process. Understanding the three questions above may help researchers to have a profound understanding of early embryonic development.

3.3. Browser

A user-friendly and configurable browser was developed to view the query results for preimplantation embryonic genes intuitively. The browser connected to the MySQL back-end allows users to select multiple species and multiple function annotations in a particular developmental stage. The result in the returning table is the intersection of the options. The information being offered by the browser, which was manually annotated by referring to over 17 500 articles, is completely credible. The browser was designed for the users who are new to the database and users whose search purpose is not clear. The browser aims to provide the most comprehensive relevant genetic information to users.

3.4. Analysis tool

The database also provides a series of analysis tools, each of them designed for comparison in one aspect. All these analysis tools are supported by the normalized expression values we stored in the database. At the outset, the expression levels of the same genes between different species are not in the same order of magnitude, so they cannot be directly compared. Even in the same species, there is no direct comparison between the different genes. After standardization of the data processing, all these values were eventually put on the same level for further analysis. The analysis tool consists of three independent functional parts. By using these tools, researchers can get expression information from three

Figure 3. The basic operations in EmExplorer are shown above. Taking the significant gene POU5F1 as a case in point, the search browser results are presented in a table and clicking the corresponding linkage leads users to the detailed information website. Analysis tools enable users to make comparisons at the gene, organism, and pathway levels, and the results are intuitively visualized. Functional enrichment analysis shows the significance of biological functions at a certain stage. All these data can be downloaded for further analysis. The Statistics box shows the current content in EmExplorer.

aspects: at the mono-gene level, at the multi-gene level and at the pathway level. We hope this may give researchers inspiration in the following ways. Firstly, scientists can know about a stage-specific expression gene of a certain species. Secondly, the database compares dynamic expression of multiple genes in a particular functional pathway. Thirdly, the tool leads researchers to understand the overall expression level of the genes in each pathway. Researchers can have a profound understanding of a particular gene or a functional pathway through the information above. The hypergeometric test result present on each development presents the significant biological functions for an organism in a particular pathway. The series tools combine organisms, genes, pathways and functions for a comprehensive understanding of the development process.

4. Result

Genes stored in EmExplorer are the basic functional entries in preimplantation embryonic development. Their accurate temporal and spatial expressions play important roles in biological research of embryonic development. Over recent years, many studies have produced large-scale data to detect dynamic changes in gene expression during mammalian embryonic development. We processed, analysed and visualized FPKM changes over time during the early embryo process, and constructed EmExplorer, the first database to provide researchers with convenient access to dynamic gene expression patterns in the developmental processes of multiple species.

The analysis tools are the highlights of the database. Since the data were processed, we firstly implemented the comparison across different species. The analysis of the pluripotent stem cell (PSC) regulation pathway at the pathway level is a case in point. PSCs are cell types unique to embryonic development and are the source of all mature tissues of the organism. The temporal regulation of PSC genes is a popular research field in PSCs. As shown in figure 2b–f, the common feature of these five organisms is that the activity of the genes in the PSC pathway rapidly increases at the eight-cell stage.

Except for human and mouse, the temporal expression in the other three organisms reaches its lowest level at the oocyte stage. Through the functional enrichment analysis, the result shows that genes specifically expressed in all five organisms at the oocyte stage and enriched in functions which negatively regulate the cell cycle and morphogenesis. However, the difference between human and mouse and the other three species is that genes with positive regulation of the cell cycle are also enriched at these stages (like TBX3). The result also indicates that the eight-cell stage may indicate nodes for cellular pluripotency activation.

Furthermore, multi-gene analysis tools are used to explore the temporal expression of pluripotency factors (e.g. NANOG, POU5F1 and SOX2) in embryonic development. The result suggested that these key genes expressed highly in the morula stage may be strongly relevant to the critical roles in initiating and maintaining pluripotency. This preliminary conclusion is consistent with previous studies [24,38,39], which means that EmExplorer is able to give researchers the right insight based on existing data. Although there are similar spatio-temporal expression patterns in various species, genes with significantly different expressive features are the focus of the researchers. For example, POU5F1 is a key factor for embryonic development and stem cell pluripotency [40,41], the expression of which rapidly increases at the four-cell stage in humans and mouse. On the contrary, it smoothly expresses in other species. BMP4, a key gene that works with POU5F1 to determine cell developmental fates [39], shows high expression at the eight-cell stage in all five species; however, in contrast to the other species, BMP4 in cattle shows continuously high expression activity at the morula stage. This result supports organism specificity

in embryo development regulation. Thus, EmExplorer provides a series of tools that greatly help users to understand the similarity or difference across diverse animals.

5. Conclusion

The significance of early embryo research may have further clinical values. For example, the recent study suggested that, during the four-cell stage in mouse embryos, heterogeneous gene expression has a strong association with cell fate decisions, the lower expression levels of Sox21 promoting extra-embryonic fate over pluripotency [30]. The characterization of the molecular mechanism in cell fate decisions may thus bring about new types of regenerative medicine therapeutics [42]. EmExplorer provides dynamic gene expression profiles to facilitate the molecular mechanism characterization of developmental processes. By using the database, experimental researchers can investigate gene expression profiles, screen development-associated genes, make comparison between species, and download relevant data for further research.

References

1. Richardson L et al. 2014 EMAGE mouse embryo spatial gene expression database: 2014 update. Nucleic Acids Res. 42, D835–D844. (doi:10.1093/nar/gkt1155)

2. Assou S. 2010 Dynamic changes in gene expression during human early embryo development: from fundamental aspects to clinical applications. Hum. Reprod. Update 17, 272–290. (doi:10.1093/humupd/dmq036)

3. Hamatani T et al. 2006 Global gene expression profiling of preimplantation embryos. Hum. Cell. 19, 98–117. (doi:10.1111/j.1749-0774.2006.00018.x)

4. Vassena R, Boué S, González-Roca E, Aran B, Auer H, Veiga A, Belmonte JCI. 2011 Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development 138, 3699–3709. (doi:10.1242/dev.064741)

5. Long CS, Li W, Liang PF, Liu S, Zuo YC. 2019 Transcriptome comparisons of multi-species identify differential genome activation of mammals embryogenesis. IEEE Access. 7, 7794–7802. (doi:10.1109/access.2018.2889809)

6. Kojima Y, Tam OH, Tam PP. 2014 Timing of developmental events in the early mouse embryo.

Semin. Cell Dev. Biol. 34, 65–75. (doi:10.1016/j.semcdb.2014.06.010)

7. Lee MT, Bonneau AR, Giraldez AJ. 2014 Zygotic genome activation during the maternal-to-zygotic transition. Annu. Rev. Cell Dev. Biol. 30, 581–613. (doi:10.1146/annurev-cellbio-100913-013027)

8. Schultz RM. 2002 The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. Hum. Reprod. Update. 8, 323–331.

9. Zhu G, Yang H, Chen X, Wu J, Zhang Y, Zhao XM. 2017 CSTEA: a webserver for the Cell State Transition Expression Atlas. Nucleic Acids Res. 45, W103–W108. (doi:10.1093/nar/gkx402)

10. Smith ZD, Sindhu C, Meissner A. 2016 Molecular features of cellular reprogramming and development. Nat. Rev. Mol. Cell Biol. 17, 139–154. (doi:10.1038/nrm.2016.6)

11. Graf A, Krebs S, Heininen-Brown M, Zakhartchenko V, Blum H, Wolf E. 2014 Genome activation in bovine embryos: review of the literature and new insights from RNA sequencing experiments. Anim. Reprod. Sci. 149, 46–58. (doi:10.1016/j.anireprosci.2014.05.016)

12. Graf A, Krebs S, Zakhartchenko V, Schwalb B, Blum H, Wolf E. 2014 Fine mapping of genome activation

in bovine embryos by RNA sequencing. Proc. Natl Acad. Sci. USA 111, 4139–4144. (doi:10.1073/pnas.1321569111)

13. Braude P, Bolton V, Moore S. 1988 Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature 332, 459–461. (doi:10.1038/332459a0)

14. de Boer BA, Ruijter JM, Voorbraak FP, Moorman AF. 2009 More than a decade of developmental gene expression atlases: where are we now? Nucleic Acids Res. 37, 7349–7359. (doi:10.1093/nar/gkp819)

15. Frum T, Halbisen MA, Wang C, Amiri H, Robson P, Ralston A. 2013 Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. Dev. Cell. 25, 610–622. (doi:10.1016/j.devcel.2013.05.004)

16. Liu H et al. 2014 DevMouse, the mouse developmental methylome database and analysis tools. Database J. Biol. Databases Curation 2014, bat084.

17. Bai W, Yang W, Wang W, Wang Y, Liu C, Jiang Q, Hua J, Liao M. 2017 GED: a manually curated comprehensive resource for epigenetic modification of gametogenesis. Brief. Bioinform. 18, 98–104. (doi:10.1093/bib/bbw007)

18. Park SJ, Shirahige K, Ohsugi M, Nakai K. 2015 DBTMEE: a database of transcriptome in mouse

early embryos. *Nucleic Acids Res.* **43**, D771–D776. (doi:10.1093/nar/gku1001)

19. Wei Y *et al.* 2014 Metalmprint: an information repository of mammalian imprinted genes. *Development* **141**, 2516–2523. (doi:10.1242/dev. 105320)

20. Zuo Y, Su G, Wang S, Yang L, Liao M, Wei Z, Bai C, Li G. 2016 Exploring timing activation of functional pathway based on differential co-expression analysis in preimplantation embryogenesis. *Oncotarget* **7**, 74 120–74 131. (doi:10.18632/oncotarget.12339)

21. Zuo Y, Gao Y, Su G, Bai C, Wei Z, Liu K, Li Q, Bou S, Li G. 2014 Irregular transcriptome reprogramming probably causes the developmental failure of embryos produced by interspecies somatic cell nuclear transfer between the Przewalski's gazelle and the bovine. *BMC Genomics.* **15**, 1113. (doi:10.1186/1471-2164-15-1113)

22. Zuo Y, Su G, Cheng L, Liu K, Feng Y, Wei Z, Bai C, Cao G, Li G. 2017 Coexpression analysis identifies nuclear reprogramming barriers of somatic cell nuclear transfer embryos. *Oncotarget* **8**, 65 847–65 859. (doi:10.18632/oncotarget.19504)

23. Jiang Z *et al.* 2014 Transcriptional profiles of bovine in vivo pre-implantation development. *BMC Genomics* **15**, 756. (doi:10.1186/1471-2164-15-756)

24. Wang X, Liu D, He D, Suo S, Xia X, He X, Han J-DJ, Zheng P. 2017 Transcriptome analyses of rhesus monkey preimplantation embryos reveal a reduced capacity for DNA double-strand break repair in primate oocytes and early embryos. *Genome Res.* **27**, 567–579. (doi:10.1101/gr.198044.115)

25. Xue Z *et al.* 2013 Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597. (doi:10.1038/nature12364)

26. Ross PJ, Wang K, Kocabas A, Cibelli JB. 2010 Housekeeping gene transcript abundance in bovine fertilized and cloned embryos. *Cellular Reprogramming* **12**, 709–717. (doi:10.1089/cell.2010.0036)

27. Liu D, Li G, Zuo Y. 2018 Function determinants of TET proteins: the arrangements of sequence motifs with specific codes. *Brief. Bioinform.* **bby53**, 1–8. (doi:10.1093/bib/bby053)

28. Duren Z, Chen X, Jiang R, Wang Y, Wong WH. 2017 Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl Acad. Sci. USA* **114**, E4914–E4923. (doi:10.1073/pnas.1704553114)

29. Xie D *et al.* 2010 Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20**, 804–815. (doi:10.1101/gr.100594.109)

30. Ma JY *et al.* 2013 Maternal factors required for oocyte developmental competence in mice: transcriptome analysis of non-surrounded nucleolus (NSN) and surrounded nucleolus (SN) oocytes. *Cell Cycle* **12**, 1928–1938. (doi:10.4161/cc.24991)

31. Wang S *et al.* 2008 Dissecting signaling pathways that govern self-renewal of rabbit embryonic stem cells. *J. Biol. Chem.* **283**, 35 929–35 940. (doi:10.1074/jbc.M804091200)

32. Wang Y, Wang F, Sun T, Trostinskaia A, Wygle D, Puscheck E, Rappolee DA. 2004 Entire mitogen activated protein kinase (MAPK) pathway is present in preimplantation mouse embryos. *Dev. Dyn.* **231**, 72–87. (doi:10.1002/dvdy.20114)

33. Carter AC, Davis-Dusenbery BN, Koszka K, Ichida JK, Eggan K. 2014 Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Rep.* **2**, 119–126. (doi:10.1016/j.stemcr.2013.12.010)

34. Cormier S, Vandormael-Pournin S, Babinet C, Cohen-Tannoudji M. 2004 Developmental expression of the Notch signaling pathway genes during mouse preimplantation development. *Gene Express.* *Patterns* **4**, 713–717. (doi:10.1016/j.modgep.2004.04.003)

35. Natale DR, Paliga AJ, Beier F, D'Souza SJ, Watson AJ. 2004 p38 MAPK signaling during murine preimplantation development. *Dev. Biol.* **268**, 76–88. (doi:10.1016/j.ydbio.2003.12.011)

36. Zhang Y, Yang Z, Wu J. 2007 Signaling pathways and preimplantation development of mammalian embryos. *FEBS J.* **274**, 4349–4359. (doi:10.1111/j.1742-4658.2007.05980.x)

37. Zuo Y, Li Y, Chen Y, Li G, Yan Z, Yang L. 2017 PseKRAAC: a flexible web server for generating pseudo K-tuple reduced amino acids composition. *Bioinformatics* **33**, 122–124. (doi:10.1093/bioinformatics/btw564)

38. Goolam M, Scialdone A, Graham SJL, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M. 2016 Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell.* **165**, 61–74. (doi:10.1016/j.cell.2016.01.047)

39. Wang Z, Oron E, Nelson B, Razis S, Ivanova N. 2012 Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440–454. (doi:10.1016/j.stem.2012.02.016)

40. Ortega NM, Winblad N, Plaza Reyes A, Lanner F. 2018 Functional genetics of early human development. *Curr. Opin Genet. Dev.* **52**, 1–6. (doi:10.1016/j.gde.2018.04.005)

41. Suwinska A, Ciemerych MA. 2011 Factors regulating pluripotency and differentiation in early mammalian embryos and embryo-derived stem cells. In *Vitamins & hormones* (ed. G. Litwack), pp. 1–37. New York, NY: Academic Press.

42. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P. 2010 Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell.* **18**, 675–685. (doi:10.1016/j.devcel.2010.02.012)