Research Article

Gene Coexpression and Evolutionary Conservation Analysis of the Human Preimplantation Embryos

Tiancheng Liu,[1] Lin Yu,[2] Guohui Ding,[1,3] Zhen Wang,[1] Lei Liu,[1] Hong Li,[1] and Yixue Li[1,3]

[1]*Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences,*
 320 Yueyang Road, Shanghai 200031, China
[2]*Key Laboratory of Contraceptive Drugs and Devices of National Population and Family Planning Commission of China,*
 Shanghai Institute of Planned Parenthood Research, 2140 Xietu Road, Shanghai 200032, China
[3]*Shanghai Center for Bioinformation Technology, 1278 Keyuan Road, Shanghai 201203, China*

Correspondence should be addressed to Hong Li; honglibio@gmail.com and Yixue Li; yxli@sibs.ac.cn

Evolutionary developmental biology (EVO-DEVO) tries to decode evolutionary constraints on the stages of embryonic development. Two models—the "funnel-like" model and the "hourglass" model—have been proposed by investigators to illustrate the fluctuation of selective pressure on these stages. However, selective indices of stages corresponding to mammalian preimplantation embryonic development (PED) were undetected in previous studies. Based on single cell RNA sequencing of stages during human PED, we used coexpression method to identify gene modules activated in each of these stages. Through measuring the evolutionary indices of gene modules belonging to each stage, we observed change pattern of selective constraints on PED for the first time. The selective pressure decreases from the zygote stage to the 4-cell stage and increases at the 8-cell stage and then decreases again from 8-cell stage to the late blastocyst stages. Previous EVO-DEVO studies concerning the whole embryo development neglected the fluctuation of selective pressure in these earlier stages, and the fluctuation was potentially correlated with events of earlier stages, such as zygote genome activation (ZGA). Such oscillation in an earlier stage would further affect models of the evolutionary constraints on whole embryo development. Therefore, these earlier stages should be measured intensively in future EVO-DEVO studies.

1. Introduction

Evolutionary developmental biology (EVO-DEVO) studies how the dynamics of development affect the phenotypic variation arising from genetic variation and its correlation with phenotypic evolution. In this subject there is a central issue: which is the most conserved period or the crucial section during the entire developmental process of an organism. While it is unarguable that the later stages of embryogenesis are not conserved among species, two major models have been proposed: the "funnel-like" model, in which the earliest embryo shows the most conserved pattern, and the "hourglass" model, in which the middle point of development is imposed with the most evolutionary constraints [1]. The "hourglass" model, which assumes the midembryonic stage (phylotypic), which shows developmental constraints and

functional importance, was originally proposed due to the expression of Hox genes in middle point of vertebrate development [2] and has been preferred in comparative transcriptomic studies nowadays [3, 4]. In addition to the transcriptomic similarity of phylotypic stages between different species, transcriptome age index (TAI) based methods, which address the total evolutionary ages of expressed genes in each developmental stage, show convergent evolution matching an hourglass pattern of embryogenesis in animals and plants [5, 6].

Mammalian preimplantation embryonic development (PED) starts from fertilization and ends at implantation of the embryo in the endometrial lining of the uterus [7]. After fertilization, the major genetic substances in the transcriptome of the zygote are the maternally deposited transcripts. After 2-3 rounds of cell divisions, maternally inherited transcripts

are degraded gradually and new transcripts of zygote are produced by the new diploid nucleus. This process is termed zygote genome activation (ZGA) [8]. These changes are not easily captured by traditional gene expression microarray techniques, as the sensitivity of microarray technology is limited when detecting low expressed genes or expression in a single cell [5, 9]. With the development of single cell RNA sequencing technology [10], we were able to identify precisely gene expression changes during the embryo developmental process which are unapparent in the microarray analysis [5]. In order to illustrate the earliest developmental gene expression fluctuation of PED which contain the crucial ZGA process and may further affect the later developmental stages, it is meaningful to look into these PED stages and identify the genetic modules regulating in each period of PED [11].

From an EVO-DEVO viewpoint, inspection of the selective constrains in PED is interesting because the trend in this period would further influence the tendency of evolutionary constrains in the middle stages of embryo development such as the phylotypic stage. Despite the importance of the expression profile of the PED stage, previous comparative transcriptome research has yet to characterize it [3, 5]. The lack of understanding of these PED stages has led past researchers to conclude that selective constraints during the earlier developmental stages increase continuously in the "funnel-like" model while they decrease in the "hourglass" model. Analysis of the selective constrains of genes in each PED stage could aid in distinguishing between the formation mechanism of the "hourglass" model or the "funnel-like" model and also consummate the whole pattern of selective constrains that act on embryonic development.

Based on the single cell RNA sequencing results of human preimplantation embryos from the oocyte stage to late blastocyst stage [12]. Applying weighted gene coexpression network analysis (WGCNA) [13], we were able to identify representative genes in each stage and summarized selective pressure on these genes to clarify the selective trend in earlier developmental stages. We found certain patterns of the evolutionary constraints that imposed on different stages of human preimplantation embryos; therefore we illustrated selective constraints on PED stages, which also presented fluctuation properties, considering that these earlier stages should be included for studying the constraints on the whole embryo development.

2. Results

2.1. Coexpression Modules for Stages in Human Preimplantation Embryos.
In the course of evolution, most biodiversity is due to alterations in gene regulation relationships rather than the sequence mutations on genes [14]. Coexpression gene modules tend to evolve together so as to share evolutionary patterns [15]. Therefore, we used gene coexpression analysis rather than differential expression to identify genes that may have close regulation relationships [16].

In order to study selective constrains in preimplantation embryonic development, we analyzed the transcriptome profiles of human preimplantation embryos (including oocyte, zygote, 2-cell, 4-cell, 8-cell, morulae, and late blastocyst stages) that were obtained by single cell RNA sequencing. Stage-specific coexpression modules were selected by the gene coexpression network analysis (WGCNA) (Figure 1) which is an unsupervised clustering method to group genes which have coexpression patterns into distinct modules [13]. This is a reliable gene coexpression analysis tool and is wildly adopted by many investigators [17–20]. After merging correlated modules with a stringent threshold, we assigned 27 out of 41 modules into a specific preimplantation developmental stage according to the correlation of eigengene of every module with each stage indicator ($r > 0.6$, $P < 0.001$). Some modules might correlate with two adjacent developmental stages because of the similarity of these two adjacent developmental stages. To remove the bias of stage comparison, we assigned this kind of modules to stage in which they had the highest correlation coefficient. After that, each module was classified into a specific developmental stage and most of the genes in each module showed consistent overexpressed behavior in corresponding developmental stage (Figure 1).

Genes in multiple modules of the same stage were merged together. In total, we obtained 2 coexpression modules for the oocyte stage, 1 module for the zygote stage, 1 module for the 2-cell stage, 5 modules for the 4-cell stage, 4 modules for the 8-cell stage, 5 modules for the morulae stage, and 9 modules for the late blastocyst stage (Figure 2). We obtained 1409, 583, 481, 1494, 1731, 1720, and 3132 specific genes for each stage, respectively. The large number of genes in the oocyte showed a complicated regulation mechanism that involved the expression of maternal genes. The number of coexpression modules and genes gradually increased with the progress of zygote development, which implied the formation of embryo complexity and modularity.

2.2. Validation of the Biological Function for Modules.
We further investigated the biological functions of genes in each specific stage by using DAVID software [21]. Gene ontology biological process (GOBP) enrichment analysis showed that genes from each stage were enriched in the relevant functions of corresponding developmental process. We also verified the function of genes in each stage by comparing them with the known function categories that were identified by Xue et al. on a different dataset of human preimplantation embryos [11]. And we compared them with the functional term identified by different methods on same datasets [12] (Table 1). The zygote gene activation (ZGA) process, which is the principal transformation of the pre-implantation period, was endorsed by significant overrepresentation of genes involved in transcription and transcription regulation process from 4-cell stage to morulae stage. In the late blastocyst stage, genes were significantly enriched in protein translation and function-associated pathways such as protein localization, transport, and phosphorylation.

2.3. Various Selective Pressures on Gene Sequence.
The nonsynonymous to synonymous substitution ratio (dN/dS) is a widely used method to measure gene sequence conservation [22]. We used dN/dS ratio for genes in each module to quantify the selective pressure on the corresponding developmental stage. The dN/dS ratios were calculated between
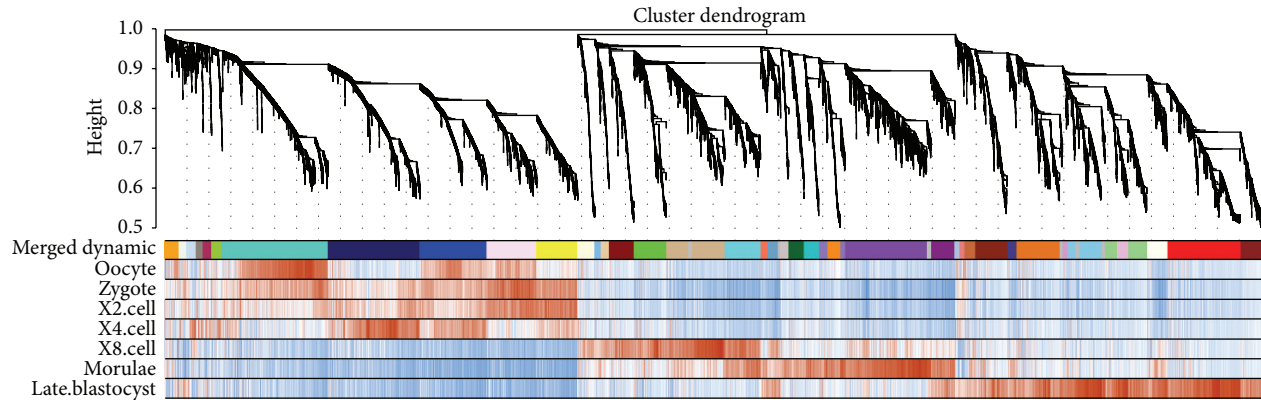
FIGURE 1: Coexpression gene modules in human preimplantation embryonic development. Hierarchical cluster tree shows the coexpression modules identified by WGCNA. The panels from top to bottom are merged dynamic modules labeled with different colors and genes correlation with indicators of each stage. The red means positive correlation while blue means negative correlation, and correlation coefficients are in direct proportion to the color depth.

mouse and human, as we intended to measure the pressure acting on sequence of genes in the mammalian species. Then their distributions were illustrated in Figure 3(a). Next we randomly sampled same number genes within each stage and calculated the median of the dN/dS distribution for the random dataset, which stood for the background. Figure 3(b) shows the median of dN/dS for genes in coexpression modules and the median of dN/dS for randomly selected genes.

The dN/dS ratios of stage-specific genes gradually decreased until the 4-cell stage, which may be caused by the consumption of the maternal genes and the expression of new genes of the zygote itself as shown by previous studies of preimplantation embryos [11, 12, 23–26]. Oocytes and zygotes had a higher median of dN/dS ratios relative to the median dN/dS for all genes whereas the dN/dS ratios of genes belonging to the 4-cell stage were significantly lower than the median dN/dS for all genes. From the zygote stage to the turning point 4-cell stage, the decreasing trend of dN/dS ratios was parallel with the process that the maternal genes expended and zygote genes emerged. The genes of the 4-cell stage were more inclined to be expressed by zygotes and had low dN/dS ratios. At the same time, genes regulated in the zygote or oocyte stages were left by maternal source and these genes had high dN/dS ratios. So the decreasing trend from maternity to zygote might suggest more striking selective pressure acting on the genes produced by zygote than selective pressure effecting genes inherited from maternity [27]. After the 4-cell stage we detected a pattern of increasing dN/dS ratios, which shows these stages expressed genes with selective pressure not as strong as 4-cell stage.

2.4. Stage-Specific Genes Were Born in Different Ancient Roots. The ages of stage-specific genes have been used as indices of evolutionary constraint [5]. We traced the root of every gene expressed in human preimplantation embryo in the phylogeny and used the ancient level of the root to represent the conservation of the gene. Based on the phylogenetic taxonomy of their roots, genes were separated into four groups: (1) Opisthokonta-Bilateria, (2) Sarcopterygii-Amniota, (3)

Chordata-Euteleostomi, and (4) Mammalia-Eutheria. For each gene set, the number of genes in each of the 4 groups was calculated to represent the age distribution. Next we marked every preimplantation developmental stage with a specific age distribution and used the age distribution of all genes as background. To detect the difference of gene age during different development stages, the age distribution of genes in each stage was compared with the background distribution of all genes (Figure 4).

We detected a clear changing trend for the Opisthokonta-Bilateria genes, with their proportion decreasing from the zygote to the 8-cell stage and then increasing until the late blastocyst. In particular, the genes belonging to the 8-cell stage were significantly depleted in Opisthokonta-Bilateria and overrepresented in Mammalia-Eutheria, which implied most genes expressed in this stage are recently born in the Mammalia-Eutheria lineage compared to other stages. In other words, these new genes, which were expressed and regulated as modules in 8-cell stage, were products of developmental evolution in the Mammalia-Eutheria lineage. This suggested that genes expressed in the 8-cell stage had a crucial function for the ZGA process of organism in Mammalia-Eutheria lineage [11, 28, 29].

As Figure 3 shows, after the 8-cell stage there was an opposite trend of increasing Opisthokonta-Bilateria genes from depletion to overrepresentation and decreasing Mammalia-Eutheria genes from overrepresentation to depletion. Finally the late blastocyst stage showed the opposite pattern—the stage was significantly depleted genes belonging to Mammalia-Eutheria and Chordata-Euteleostomi groups and it was overrepresented of genes in Opisthokonta-Bilateria group. This sort of opposite pattern illustrated that the late blastocyst stage was conserved as it tended to express the oldest genes.

2.5. Genes in Each Stages Present Diverse Duplicated States. Gene duplication state is also an indicator of selective pressure [30]. In order to evaluate the conservation of genes more widely, we chose the zebra fish, an evolutionary distant

Module-trait relationships

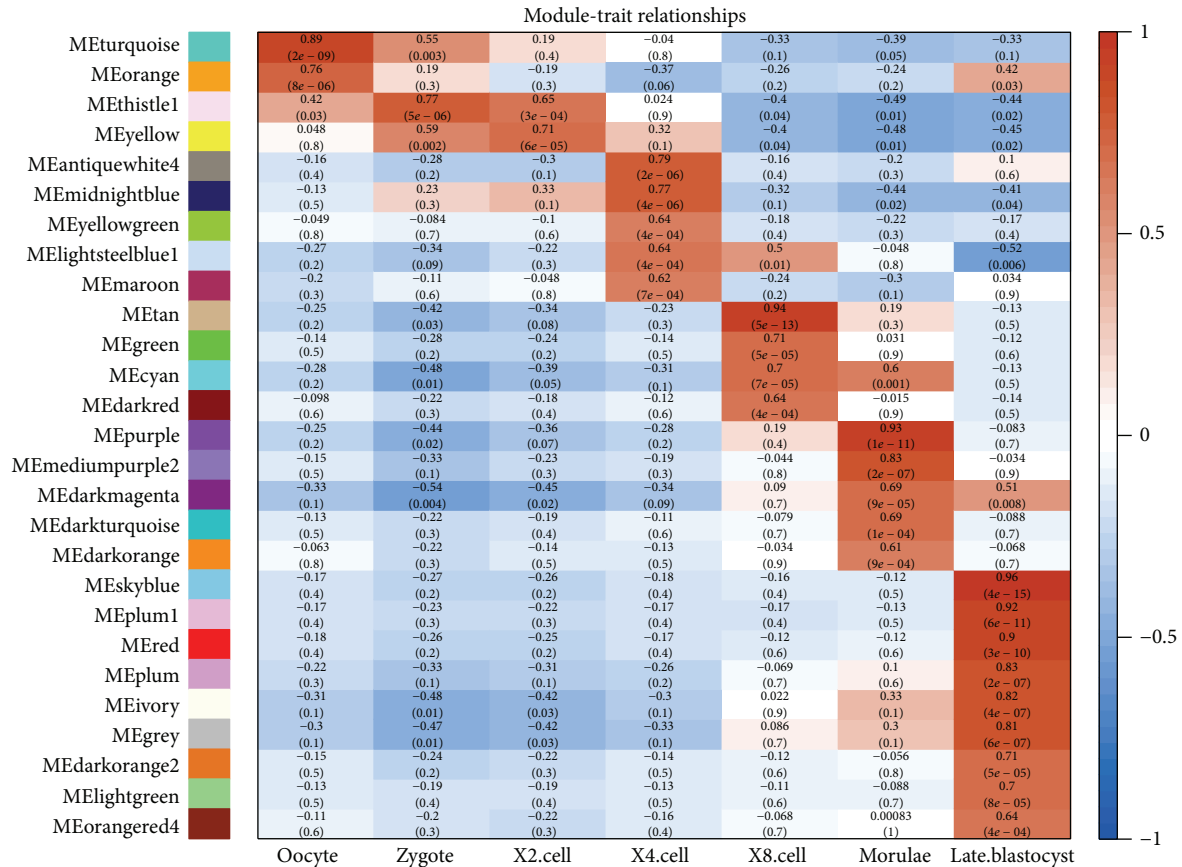| Module	Oocyte	Zygote	X2.cell	X4.cell	X8.cell	Morulae	Late.blastocyst
MEturquoise	0.89 (2e−09)	0.55 (0.003)	0.19 (0.4)	−0.04 (0.8)	−0.33 (0.1)	−0.39 (0.05)	−0.33 (0.1)
MEorange	0.76 (8e−06)	0.19 (0.3)	−0.19 (0.3)	−0.37 (0.06)	−0.26 (0.2)	−0.24 (0.2)	0.42 (0.03)
MEthistle1	0.42 (0.03)	0.77 (5e−06)	0.65 (3e−04)	0.024 (0.9)	−0.4 (0.04)	−0.49 (0.01)	−0.44 (0.02)
MEyellow	0.048 (0.8)	0.59 (0.002)	0.71 (6e−05)	0.32 (0.1)	−0.4 (0.04)	−0.48 (0.01)	−0.45 (0.02)
MEantiquewhite4	−0.16 (0.4)	−0.28 (0.2)	−0.3 (0.1)	0.79 (2e−06)	−0.16 (0.4)	−0.2 (0.3)	0.1 (0.6)
MEmidnightblue	−0.13 (0.5)	0.23 (0.2)	0.33 (0.1)	0.77 (4e−06)	−0.32 (0.1)	−0.44 (0.02)	−0.41 (0.04)
MEyellowgreen	−0.049 (0.8)	−0.084 (0.7)	−0.1 (0.6)	0.64 (4e−04)	−0.18 (0.4)	−0.22 (0.3)	−0.17 (0.4)
MElightsteelblue1	−0.27 (0.2)	−0.34 (0.09)	−0.22 (0.3)	0.64 (4e−04)	0.5 (0.01)	−0.048 (0.8)	−0.52 (0.006)
MEmaroon	−0.2 (0.3)	−0.11 (0.6)	−0.048 (0.8)	0.62 (7e−04)	−0.24 (0.2)	−0.3 (0.1)	0.034 (0.9)
MEtan	−0.25 (0.2)	−0.42 (0.03)	−0.34 (0.08)	−0.23 (0.3)	0.94 (5e−13)	0.19 (0.3)	−0.13 (0.5)
MEgreen	−0.14 (0.5)	−0.28 (0.2)	−0.24 (0.2)	−0.14 (0.5)	0.71 (5e−05)	0.031 (0.9)	−0.12 (0.6)
MEcyan	−0.28 (0.2)	−0.48 (0.01)	−0.39 (0.05)	−0.31 (0.1)	0.7 (7e−05)	0.6 (0.001)	−0.13 (0.5)
MEdarkred	−0.098 (0.6)	−0.22 (0.3)	−0.18 (0.4)	−0.12 (0.6)	0.64 (4e−04)	−0.015 (0.9)	−0.14 (0.5)
MEpurple	−0.25 (0.2)	−0.44 (0.02)	−0.36 (0.07)	−0.28 (0.2)	0.19 (0.4)	0.93 (1e−11)	−0.083 (0.7)
MEmediumpurple2	−0.15 (0.5)	−0.33 (0.1)	−0.23 (0.3)	−0.19 (0.3)	−0.044 (0.8)	0.83 (2e−07)	−0.034 (0.9)
MEdarkmagenta	−0.33 (0.1)	−0.54 (0.004)	−0.45 (0.02)	−0.34 (0.09)	0.09 (0.7)	0.69 (9e−05)	0.51 (0.008)
MEdarkturquoise	−0.13 (0.5)	−0.22 (0.3)	−0.19 (0.4)	−0.11 (0.6)	−0.079 (0.7)	0.69 (1e−04)	−0.088 (0.7)
MEdarkorange	−0.063 (0.8)	−0.22 (0.3)	−0.14 (0.5)	−0.13 (0.5)	−0.034 (0.9)	0.61 (9e−04)	−0.068 (0.7)
MEskyblue	−0.17 (0.4)	−0.27 (0.2)	−0.26 (0.2)	−0.18 (0.4)	−0.16 (0.4)	−0.12 (0.5)	0.96 (4e−15)
MEplum1	−0.17 (0.4)	−0.23 (0.3)	−0.22 (0.3)	−0.17 (0.4)	−0.17 (0.4)	−0.13 (0.5)	0.92 (6e−11)
MEred	−0.18 (0.4)	−0.26 (0.2)	−0.25 (0.2)	−0.17 (0.4)	−0.12 (0.6)	−0.12 (0.6)	0.9 (3e−10)
MEplum	−0.22 (0.3)	−0.33 (0.1)	−0.31 (0.1)	−0.26 (0.2)	−0.069 (0.7)	0.1 (0.6)	0.83 (2e−07)
MEivory	−0.31 (0.1)	−0.48 (0.01)	−0.42 (0.03)	−0.3 (0.1)	0.022 (0.9)	0.33 (0.1)	0.82 (4e−07)
MEgrey	−0.3 (0.1)	−0.47 (0.01)	−0.42 (0.03)	−0.33 (0.1)	0.086 (0.7)	0.3 (0.1)	0.81 (6e−07)
MEdarkorange2	−0.15 (0.5)	−0.24 (0.2)	−0.22 (0.3)	−0.14 (0.5)	−0.12 (0.6)	−0.056 (0.8)	0.71 (5e−05)
MElightgreen	−0.13 (0.5)	−0.19 (0.4)	−0.19 (0.4)	−0.13 (0.5)	−0.11 (0.6)	−0.088 (0.7)	0.7 (8e−05)
MEorangered4	−0.11 (0.6)	−0.2 (0.3)	−0.22 (0.3)	−0.16 (0.4)	−0.068 (0.7)	0.00083 (1)	0.64 (4e−04)

FIGURE 2: Modules are distributed to each stage according to the correlation between eigengenes and stage indicators. The red means positive correlation while blue means negative correlation, and correlation coefficients are in direct proportion to the color depth as shown in right side. Each element of the matrix denotes the correlation coefficients between module eigengenes (row) and stage (column); then the significant level of correlation is marked below the correlation coefficients.

species, as reference to check the gene-duplicated situation of human genes in each development stage. Genes were separated into four groups based on the gene duplication states: (1) one-to-many, (2) one-to-one, (3) many-to-many, and (4) new gene (no ortholog in the zebra fish genome). We removed the many-to-many gene pairs because it is difficult to evaluate their conservation. As stated above, we compared the observed distribution of genes in each stage with the expected distribution that was recorded by distributing all genes into these 3 groups (Figure 5).

Genes falling in one-to-many orthologs revealed that they were single copy in human and their orthologs had duplications in zebra fish. Knowing that constrained developmental stages should display less change in gene family size [31], the genes, which duplicate in other species but keep singleton in human developmental stages, should be considered to be conserved specifically in *Homo sapiens*. Otherwise, the one-to-one orthologs retain the functions of ancestral gene since the last shared common ancestor, and left no duplication in the human or zebra fish lineage. Therefore genes of one-to-one orthologs also should be subject to functional constraints. Just as in the above age analysis of genes, the new genes, which were new products during the evolutionary

process of *Homo sapiens*, were considered to be under less constraint. At last, many-to-many orthologs showed duplication events in both species and their conservation patterns were complicated; thus we ignored many-to-many orthologs in the further analysis.

As Figure 5(b) shows, the one-to-one (single-copy) orthologs and new genes exhibited opposite trends in the preimplantation period which implied the transformation of evolutionary constraints on different stages during the developmental process. In particular, the 4-cell stage showed significant depletion of the one-to-one genes but overrepresentation of the one-to-many genes that signify genes of this stage is under strong functional constraints on their sequence in *Homo sapiens*. It accorded with the dN/dS result showing that genes expressed in the 4-cell stage had significantly lower human-mouse dN/dS ratios and lends further evidence to the hypothesis of conservation of genes belonging to the 4-cell stage. Moreover, the 8-cell stage showed overrepresentation of the newborn genes and depletion of the one-to-one genes, which is consistent with the gene age analysis. The large number of new genes in the 8-cell stage offers further evidence for human-specific embryonic development occurring in this stage [11, 28, 29].

TABLE 1: Enriched biological process terms for stage-specific genes in preimplantation embryonic development. Similar function annotation clustering was presented by one typical function term. The last column stands for the validation of corresponding term in other studies of human preimplantation embryonic development.

Stage	Functional term	Bonferroni P value	Consistence with previous report
Oocyte	DNA repair	$3.35E - 4$	
	Cell cycle	$2.92E - 3$	[11, 12]
	DNA metabolic process	$7.30E - 3$	
Zygote	Cell cycle	$3.33E - 4$	[12]
2-cell	Posttranscriptional regulation of gene expression	$3.56E - 2$	
4-cell	Transcription	$1.76E - 4$	[11, 12]
	Regulation of transcription	$6.34E - 3$	[11, 12]
	Small GTPase mediated signal transduction	$3.14E - 2$	[11]
8-cell	DNA packaging	$8.09E - 6$	
	RNA processing	$9.27E - 5$	[11]
	Transcription	$1.02E - 4$	[11, 12]
	Regulation of transcription	$3.08E - 4$	[11, 12]
	Translational elongation		
	Ribonucleoprotein complex biogenesis	$5.89E - 4$	[12]
	RNA elongation	$2.09E - 3$	
Morulae	RNA processing	$2.38E - 8$	[11]
	Transcription	$7.33E - 8$	[11, 12]
	Regulation of transcription	$5.38E - 7$	[11, 12]
	Ribonucleoprotein complex biogenesis	$9.81E - 6$	[12]
Late blastocyst	Generation of precursor metabolites and energy	$7.78E - 18$	[12]
	Translation	$4.11E - 12$	[12]
	Oxidative phosphorylation	$9.24E - 12$	[12]
	Cellular respiration	$1.69E - 8$	[12]
	Membrane organization	$9.28E - 6$	
	Protein localization	$2.74E - 5$	
	Cofactor metabolic process	$3.64E - 5$	
	Protein transport	$6.17E - 5$	
	Monosaccharide metabolic process	$3.67E - 3$	

As with the above gene age analysis, we also detected conserved convergence from the 8-cell stage to late blastocyst stage reflected by the transition from an overrepresented state to a depleted state of the new genes and by the transition from depleted state to overrepresented state of the one-to-one genes. Finally, the late blastocyst stage reached a conserved state, which was significantly depleted of new genes and overrepresented for one-to-one genes.

2.6. Evolvability of Regulatory Regions in Upstream of Stage-Specific Genes. Conservation of *cis*-regulatory sequences is also a critical standard for measuring the selective pressure on genes [14, 32], and highly conserved noncoding elements (HCNEs) are often considered to be associated with developmental regulatory genes or transcription factors (TFs) [23, 33]. Therefore, we determined the transcriptional importance of stage-specific genes by analyzing their potential to become TFs and the distribution of HCNEs in their promoter regions (Figure 6).

We found that promoter regions of genes in the 2-cell stage were significantly enriched for HCNEs, and there are more TFs in 2-cell stage than expected. These enriched transcriptional factors and transcriptional regulatory elements may promote effective gene transcripts in the 2-cell embryo and launch the progress of zygote genome activation (ZGA). TFs were significantly enriched in 4-cell, 8-cell, and morulae stages, which indicated that the gene expression and regulation network became more sophisticated during the zygote gene activation (ZGA) process. Our finding of a relatively desolate transcriptional scenario in the late blastocyst stage accords with the findings of Piasecka et al. [31], who proposed that the cleavage/blastula modules of zebra fish development are not enriched with transcriptional devices. Finally, the gathering of these transcriptional elements during the ZGA
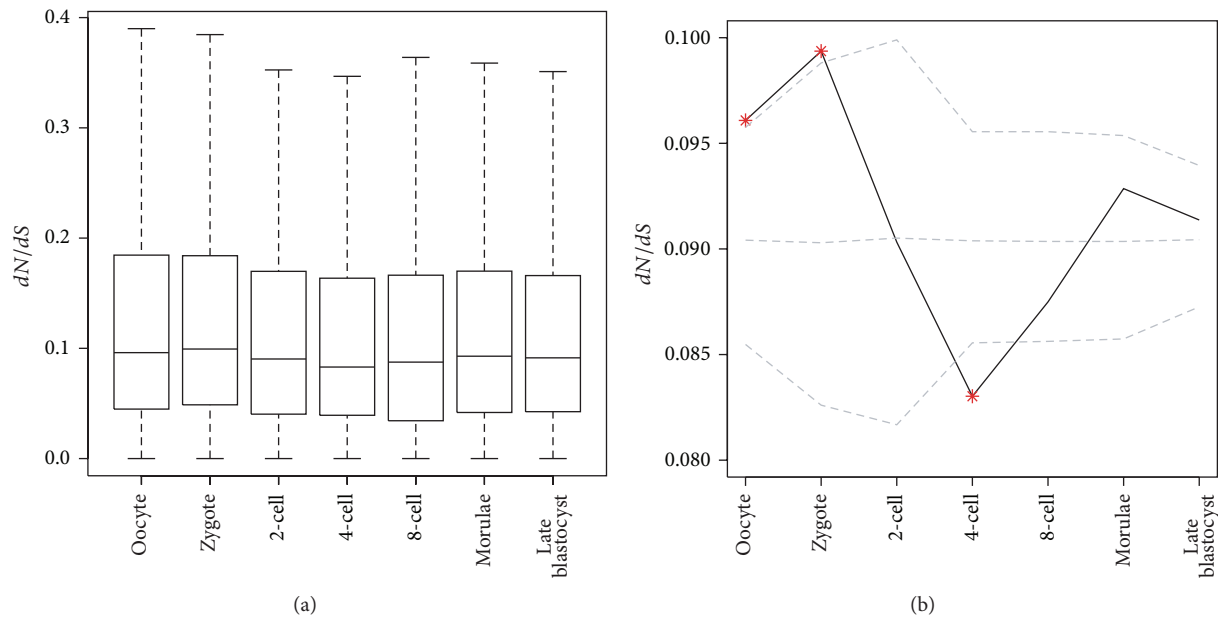
(a)

(b)

FIGURE 3: (a) Box plot of conserved index during different development stages. Conservation is measured by the nonsynonymous to synonymous substitution ratios (dN/dS). (b) The median of dN/dS ratio in each stage is represented by the solid line. The dash lines denote the median and the 95% confidence interval for the dN/dS ratio of randomly selected genes, which have same number of genes in corresponding stage. The asterisks denote that dN/dS are significantly different ($P < 0.05$) from the median of background.

(a)

(b)

(c)

(d)
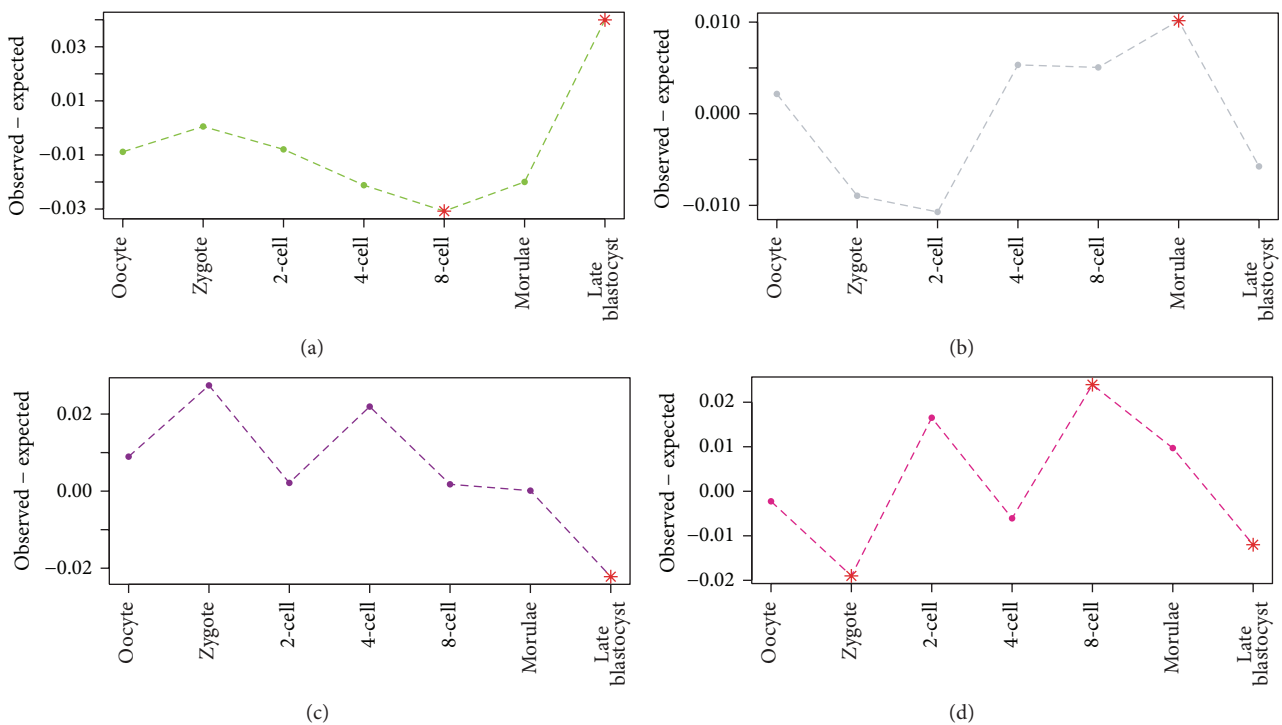
FIGURE 4: Distributions of gene ages. Genes were classified into four groups based on their first appearance in the phylogeny: (a) Opisthokonta-Bilateria, (b) Sarcopterygii-Amniota, (c) Chordata-Euteleostomi, and (d) Mammalia-Eutheria. For each stage, the vertical axis shows the observed frequencies minus expected frequencies of gene ages. The asterisks denote significant enrichment ($P < 0.01$) in a specific gene group for each stage.

(a)

(b)

(c)
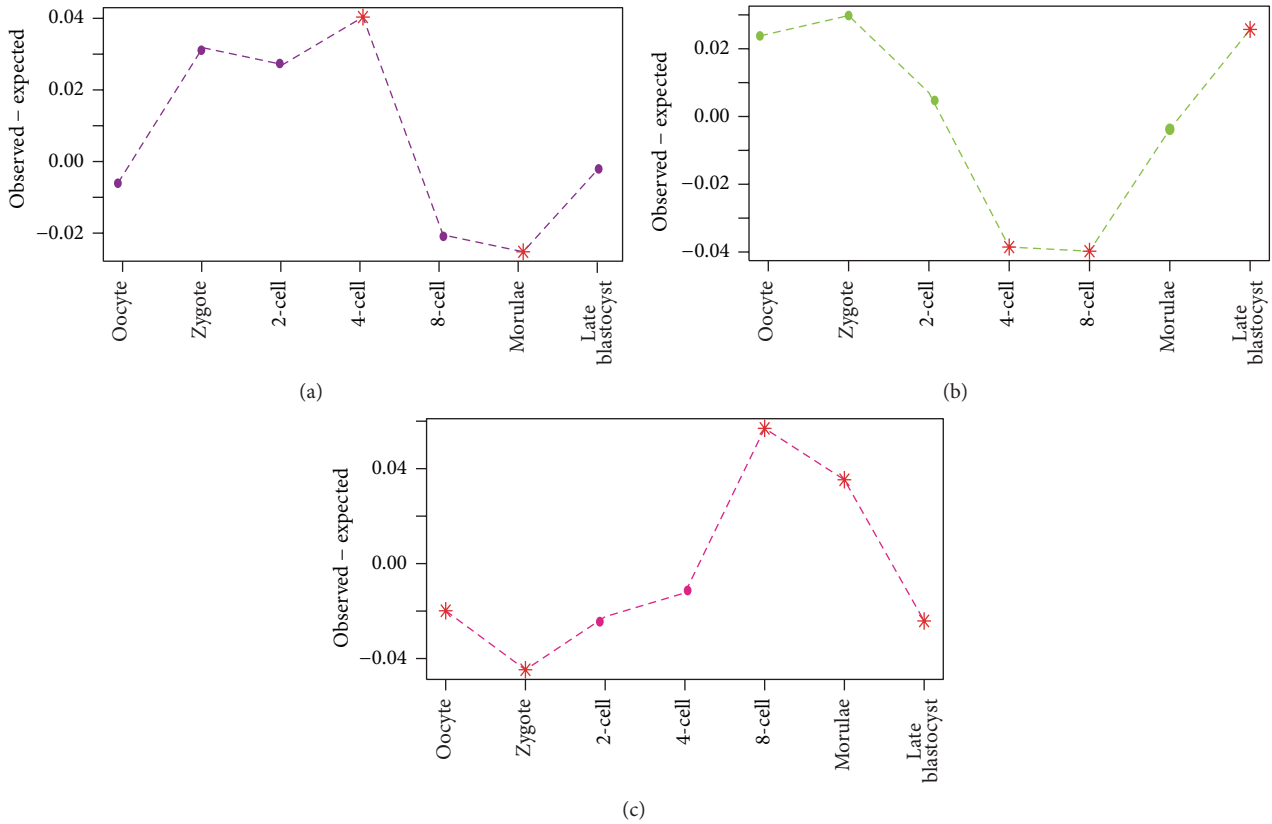
FIGURE 5: Distributions of different gene ortholog types comparing human with zebra fish: (a) one-to-many ortholog, (b) one-to-one ortholog, and (c) new genes. For each stage, the vertical axis shows the observed proportions minus expected proportions of genes in each ortholog type. The asterisks denote significant enrichment ($P < 0.01$) in a specific ortholog scenario for each stage.

(a)

(b)
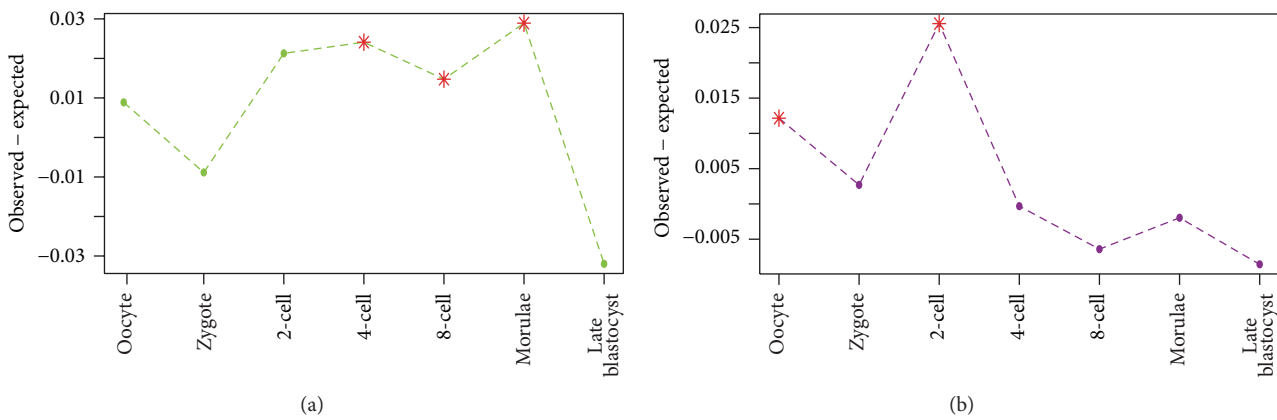
FIGURE 6: (a) Distributions of transcription factors (TFs) in stage-specific modules. (b) Distributions of genes with highly conserved noncoding elements (HCNEs) in their *cis*-regulatory regions. For each stage, the vertical axis shows the observed proportions minus expected proportions of TF and genes with HCNEs, respectively. The asterisks denote significant enrichment ($P < 0.01$).

process could not be disregarded and this might further influence the evolutionary model or regulation mechanism of the whole developmental schedule.

2.7. Patterns of Evolutionary Constraints in Preimplantation Embryonic Development. Based on WGCNA, we clustered the genes of human preimplantation embryonic development into modules and linked these modules to specific stages of this developmental process. Next, we checked four conservation properties for stage-specific genes, including gene sequences, gene ages, gene orthologs, and regulatory elements. All of these indices implied several features during
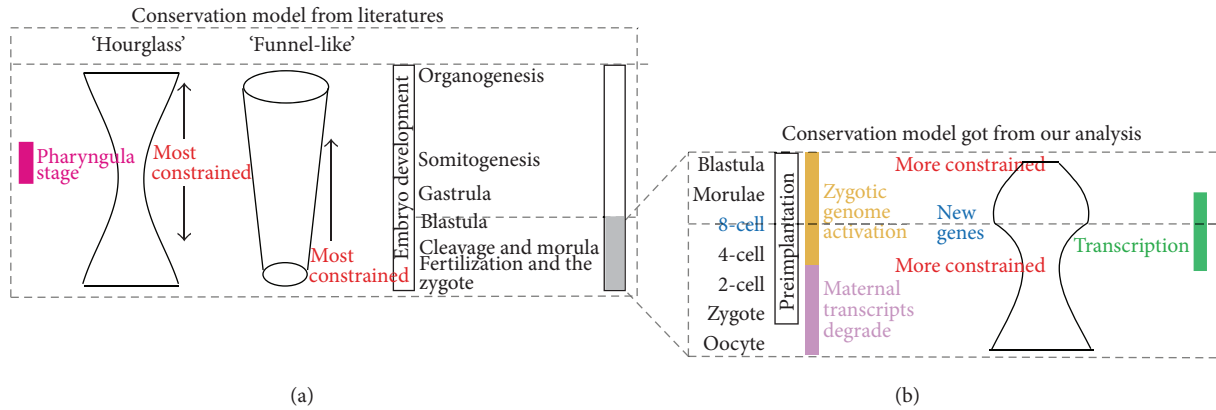
Figure 7: (a) Two models—the "funnel-like" model and the "hourglass" model, revealed by the evolutionary constraint on the whole developmental process of organism. Corresponding stages are labeled on the right side and earlier stages to late stages are in bottom to up sequence. Pharyngula stage (phylotypic stage) is displayed by pink rectangle on the left side. (b) The fluctuation of evolutionary constraints on the stages of human preimplantation embryonic development. In the left side, preimplantation stages are ranked from bottom to up. Maternal transcripts degrade and ZGA are labeled with purple rectangle and yellow rectangle, respectively.

the process of human preimplantation embryonic development.

First, we observed that maternal genes were under less selective constraints while there were strong selective constraints effecting on earlier zygote-activated genes. This was verified via the reduction of the dN/dS ratio accompanied by the consumption of maternal mRNAs and ZAG expressing from zygote to 4-cell stage (Figure 3), and the overrepresentation of the conserved one-to-many orthologs in the 4-cell stage (Figure 5). Secondly, we discovered a switch of the evolutionary constraints at 8-cell stage in which the embryo tended to express new genes. This trend is reflected by the fact that dN/dS begin elevating after the 8-cell stage (Figure 3). Meanwhile, genes in 8-cell stage present depletion of oldest Opisthokonta-Bilateria genes and show overrepresented of newest Mammalia-Eutheria genes (Figure 4). The burst of new genes in the 8-cell stage was further demonstrated by the depletion of one-to-one zebra fish orthologs genes and overrepresentation of human specific genes in this stage (Figure 5). Lastly, the selective pressure on late blastocyst stage tended to increase again. The late blastocyst stage was overrepresented in the oldest Opisthokonta-Bilateria genes (Figure 4) and one-to-one orthologs of zebra fish (Figure 5). The phylotypic stage in middle development was specifically enriched for transcriptional elements so that the transcriptional factory was subtly working in this stage [31]. Our work revealed that ZGA in early stages also showed the enrichment of transcriptional elements (Figure 6), which indicated the ZGA process is under precise regulation as phylotypic stage.

In summary, we found that, in the earlier developmental stages of the human embryo, the conservation indices presented the sequence of increasing—decreasing—increasing (Figure 7), rather than increasing or decreasing monotonically. And this trend is potentially correlated with the maternal transcripts degrade and ZGA. As part of the developmental process, earlier embryo development turns out to be a complicated process which also involves the fluctuation of selective pressure.

3. Discussion

Mammalian developments comprise three important processes: zygote genome activation (ZGA) at earlier stages [11], expression of Hox genes at middle stages [2], and morphological formation at late stages [34]. The evolutionary conservation of these three stages has been debated at length. Previous EVO-DEVO studies concerning development [1, 4, 5, 31] focused on conservation during the whole development process, while changes of selective pressure during earlier development were neglected. These studies typically used one stage (such as the zygote stage) to represent the earlier embryonic stages [5]. Most stages of the earlier embryo (such as 2-cell to 8-cell stages) were discarded as they are relatively short compared to the long time interval of stages in middle and late embryo. By monitoring these transient stages of earlier development, the exquisite regulation mechanism of the ZGA process could be revealed [11, 12]. Rather than monitoring once after certain time intervals [4], the detecting time points should be chosen according to developmental events and time interval of different stages should be specially selected. Therefore it is meaningful to set more observing points during the earlier stages [35]. Here we analyze eight stages in preimplantation embryonic development. Our results show that the conservation scenario of the earlier embryo has a degree of fluctuation different from the direct increase or decrease previously reported [1, 3, 5, 6, 31]. The fluctuation in PED stages was probably associated with the events occurring during these stages. For instance, more selective pressure was on early zygotic genes than maternal genes [27]; therefore the measure of conservation increased during the maternal transcripts degrading process. Then the embryo was building the infrastructure so that it expressed highly conserved genes during the 2-cell stage to 4-cell stage. After this, the embryo expressed some species specific genes to determine its fate inclination [36]. Finally, the genes expressed near middle embryonic stages presented to be conserved, which was accordance with the hourglass

model. Our work illustrates that the dynamics of evolutionary indices during these short-time early stages should also be taken into consideration in discussions of the "hourglass" model or the "funnel-like" model of embryo development. We believe a precise exploration of the evolutionary indices of earlier developmental stages will lead to the creation of a more sophisticated model of selective pressure on the whole development process.

4. Methods

4.1. Transcriptional Profiling of Preimplantation Embryos. The gene expression profilings of human preimplantation embryos were downloaded from NCBI's Gene Expression Omnibus [37] (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36552). It contained the whole transcriptomic RNA expression levels of oocyte stage, zygote stage, 2-cell stage, 4-cell stage, 8-cell stage, morulae stage, and late blastocyst stage, which were measured by RPKM (Reads Per Kilobase of transcript per Million mapped reads) via single cell RNA sequencing [38]. Each stage was composed of 4 biologically replicated samples except for oocyte and zygote stages, which had 3 samples. To eliminate the bias of genes which have zero or extremely low expression levels in many stages, genes with low expression in all stages (average RPKM < 0.5) were removed. Each gene symbol of the whole profile was mapped to its corresponding Ensembl gene ID and the gene symbols that have no corresponding Ensembl gene ID were discarded to reduce the potential noise. At last the expression profiles in each sample were processed by quantile normalization that accounts for different amounts of RNA present throughout embryo earlier development.

4.2. Weighted Gene Coexpression Network Analysis. The final expression matrix was proceed by the step by step WGCNA [13] method. First, we built a matrix which includes pairwise correlation coefficients between all pairs of genes [39]. Next, with the power of 12 which is the default value, the adjacency matrix was constructed. Depending on the resulting adjacency matrix, we calculated the topological overlap matrix, which measures the interconnectedness of the coexpression network [40]. And then this topological overlap matrix was used to perform hierarchical clustering in which genes with coexpression relationships were grouped together and formed a gene clustering tree. The primary modules were identified by Dynamic Hybrid Tree Cut algorithm [41] to cut the hierarchal clustering tree with coefficients deepSplit = 4. At last, we calculated the correlation coefficients of each pair of module eigengenes that stand for the first principal component of the module and merged highly similar modules by a stringent threshold (correlation > 0.9).

After identifying the coexpression modules, we associated these modules with specific embryo developmental stage and picked hub genes for each module. This process was based on correlating each module eigengene which represented each module with the stage indicator genes and all genes on the matrix. For genes with high correlation coefficients (correlation above 0.9 and *P* value < 0.01) with specific module, we treated them as the hub genes of this module. To associate these modules with developmental stage, we used a threshold (correlation coefficient > 0.7 and *P* value < 0.01) to pick up modules which belong to a certain stage. Modules that correlated with two stages were only kept in the stage with the highest correlation coefficients.

4.3. Gene Ontology Analysis. Functional annotation was performed with the DAVID Bioinformatics Resources. To correct multiple testing, the Bonferroni correction was applied. And the enriched GO biological process categories were picked up by the corrected *P* value (<0.01). Then we checked whether these enriched GO categories were also presented in the same stages of similar studies [11, 12].

4.4. dN/dS Analysis. We downloaded *dN* and *dS* values of all human genes using BioMart [42], which was calculated by the ortholog genes between human and mouse. After removing genes that were not presented on the expression matrix, we got 12865 *dN/dS* value.

We calculated the median *dN/dS* ratio of genes in each stage and evaluated if it was significantly higher (lower) than randomly selected genes. For each stage which has k genes, we generated 10000 sets of k randomly chosen genes from the background 12865 genes and calculated the median *dN/dS* ratio for each random set. *P* value was calculated as the tail probability of real *dN/dS* ratio in the distribution of randomly generated *dN/dS*.

4.5. Gene Age Analysis. Genes of *Homo sapiens* originated in different taxonomic root of the phylogeny so that genes have different age index. We could label every gene with an age index by its first appearance in the phylogeny. For each gene in our expression matrix the oldest node of its gene tree was retrieved from Ensembl release 75 [43] by Ensembl comparative genomics API. After that, each gene was marked with a unique age index from oldest Opisthokonta node to the latest human node.

In order to make subsequent test convincible we removed some genes falling in age interval from Eutheria to human because the number of genes in these interval is very rare (less than 5) which will obscure the statistical test. And the rest of the genes were merged into one of the following age intervals: Opisthokonta-Bilateria, Opisthokonta-Bilateria, Sarcopterygii-Amniota, and Mammalia-Eutheria. That made each category have sufficient number of genes to perform the statistical test.

Next, for every module we collected all the age indices of its k genes and counted the number of genes falling to each age interval. Then the age index distribution of expected background was estimated by classifying all genes (11919 genes) presented on the expression matrix into these categories. The number of genes in each age category was transformed to the proportion by dividing the number of all expressed genes in this stage. Then we plotted the "observed minus expected" proportions of each stage and performed Fisher's exact test to compare the observed and expected numbers of age indices in each stage. We picked up stages in which genes with certain age index were overrepresented

or underrepresented (P value < 0.01) and highlighted these stages on the plot.

4.6. Human-Zebra Fish Orthologous Genes. Based on an evolutional distant species zebra fish, homology information between human and zebra fish genes was retrieved from Ensembl release 75 [43] by BioMart [42]. 10919 of the 12865 genes presented in the expression matrix have human-zebra fish paired orthologs, including 7265 one-to-one orthologs, 2995 one-to-many orthologs, and 659 many-to-many orthologs. Then the remaining 1946 human genes which do not have ortholog relationship with zebra fish genome were labeled as new gene in human.

We calculated the observed number of stage-specific genes that were in the three kinds of ortholog types (one-to-one, one-to-many, and no orthology), and constructed the expected background distribution from all genes. For each category we plotted the "observed minus expected" proportions of each stage and performed Fisher's exact test to compare the observed and expected numbers.

4.7. Gene Transcriptional Region Analysis. Gene transcription analysis was based on the number of transcription factors (TFs) and highly conserved noncoding elements (HCNEs) in the promoter regions of genes. Genes with GO category annotation (GO: 0006355, regulation of transcription, DNA-dependent) were defined as TFs. Location data of HCNE between human and mouse with identity above 90% was downloaded from Ancora (http://ancora.genereg.net/downloads/hg19/vs_mouse/) [33]. For each of the 12865 genes considered in our analysis, we checked if there were HCNEs located in 500 base-pairs upstream from the transcription start site. Totally we annotated 1438 and 848 genes as TFs and HCNEs, respectively. For every stage we performed the hypergeometric test to assess if genes in this stage were significantly enriched in TFs and HCNEs.

Conflict of Interests

The authors declare they have no competing interests.

Authors' Contribution

Tiancheng Liu carried out the coexpression analysis and evolutionary analysis and then drafted the paper. Lin Yu provided the instruction of developmental biology and collected the datasets. Guohui Ding and Zhen Wang participated in understanding biological question and phylogenetic analysis. Hong Li revised the paper. Hong Li and Yixue Li conceived and designed the study. All authors read and approved the final paper. Tiancheng Liu and Lin Yu contributed equally to the study.

Acknowledgments

References

[1] N. Irie and S. Kuratani, "Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis," *Nature Communications*, vol. 2, article 248, 2011.

[2] D. Duboule, "Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony," *Development*, supplement, pp. 135–142, 1994.

[3] Z. Wang, J. Pascual-Anaya, A. Zadissa et al., "The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan," *Nature Genetics*, vol. 45, no. 6, pp. 701–706, 2013.

[4] A. T. Kalinka, K. M. Varga, D. T. Gerrard et al., "Gene expression divergence recapitulates the developmental hourglass model," *Nature*, vol. 468, no. 7325, pp. 811–816, 2010.

[5] T. Domazet-Lošo and D. Tautz, "A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns," *Nature*, vol. 468, no. 7325, pp. 815–819, 2010.

[6] M. Quint, H.-G. Drost, A. Gabel, K. K. Ullrich, M. Bönn, and I. Grosse, "A transcriptomic hourglass in plant embryogenesis," *Nature*, vol. 490, no. 7418, pp. 98–101, 2012.

[7] M. S. H. Ko, "Embryogenomics of pre-implantation mammalian development: current status," *Reproduction, Fertility and Development*, vol. 16, no. 1-2, pp. 79–85, 2004.

[8] J. Kaňka, "Gene expression and chromatin structure in the pre-implantation embryo," *Theriogenology*, vol. 59, no. 1, pp. 3–19, 2003.

[9] R. Vassena, S. Boué, E. González-Roca et al., "Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development," *Development*, vol. 138, no. 17, pp. 3699–3709, 2011.

[10] A.-E. Saliba, A. J. Westermann, S. A. Gorski, and J. Vogel, "Single-cell RNA-seq: advances and future challenges," *Nucleic Acids Research*, vol. 42, no. 14, pp. 8845–8860, 2014.

[11] Z. Xue, K. Huang, C. Cai et al., "Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing," *Nature*, vol. 500, no. 7464, pp. 593–597, 2013.

[12] L. Yan, M. Yang, H. Guo et al., "Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells," *Nature Structural and Molecular Biology*, vol. 20, no. 9, pp. 1131–1139, 2013.

[13] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, article 559, 2008.

[14] S. B. Carroll, "Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution," *Cell*, vol. 134, no. 1, pp. 25–36, 2008.

[15] M. Sémon and L. Duret, "Evolutionary origin and maintenance of coexpressed gene clusters in mammals," *Molecular Biology and Evolution*, vol. 23, no. 9, pp. 1715–1723, 2006.

[16] A. de la Fuente, "From 'differential expression' to 'differential networking'—identification of dysfunctional regulatory networks in diseases," *Trends in Genetics*, vol. 26, no. 7, pp. 326–333, 2010.

[17] X. A. Chang, S. A. Liu, Y.-T. Yu, Y.-X. Li, and Y.-Y. Li, "Identifying modules of coexpressed transcript units and their

organization of *Saccharopolyspora erythraea* from time series gene expression profiles," *PLoS ONE*, vol. 5, no. 8, Article ID e12126, 2010.

[18] X. Chang, L. L. Shi, F. Gao et al., "Genomic and transcriptome analysis revealing an oncogenic functional module in meningiomas," *Neurosurgical Focus*, vol. 35, no. 6, article E3, 2013.

[19] C. R. Farber, "Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data," *Journal of Bone and Mineral Research*, vol. 25, no. 11, pp. 2359–2367, 2010.

[20] C. G. J. Saris, S. Horvath, P. W. J. van Vught et al., "Weighted gene co-expression network analysis of the peripheral blood from amyotrophic lateral sclerosis patients," *BMC Genomics*, vol. 10, p. 405, 2009.

[21] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.

[22] H. J. M. Aarts, E. H. M. Jacobs, G. van Willigen, N. H. Lubsen, and J. G. G. Schoenmakers, "Different evolution rates within the lens-specific β-crystallin gene family," *Journal of Molecular Evolution*, vol. 28, no. 4, pp. 313–321, 1989.

[23] R. Sanges, Y. Hadzhiev, M. Gueroult-Bellone et al., "Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development," *Nucleic Acids Research*, vol. 41, no. 6, pp. 3600–3618, 2013.

[24] E. M. Thompson, "Chromatin structure and gene expression in the preimplantation mammalian embryo," *Reproduction Nutrition Development*, vol. 36, no. 6, pp. 619–635, 1996.

[25] L. Li, X. Lu, and J. Dean, "The maternal to zygotic transition in mammals," *Molecular Aspects of Medicine*, vol. 34, no. 5, pp. 919–938, 2013.

[26] T. Hamatani, M. S. Ko, M. Yamada et al., "Global gene expression profiling of preimplantation embryos," *Human Cell*, vol. 19, no. 3, pp. 98–117, 2006.

[27] J. Mensch, F. Serra, N. J. Lavagnino, H. Dopazo, and E. Hasson, "Positive selection in nucleoporins challenges constraints on early expressed genes in *Drosophila* development," *Genome Biology and Evolution*, vol. 5, no. 11, pp. 2231–2241, 2013.

[28] A. Galán, D. Montaner, M. E. Póo et al., "Functional genomics of 5- to 8-cell stage human embryos by blastomere single-cell cDNA analysis," *PLoS ONE*, vol. 5, no. 10, Article ID e13615, 2010.

[29] Z. Jiang, J. Sun, H. Dong et al., "Transcriptional profiles of bovine in vivo pre-implantation development," *BMC Genomics*, vol. 15, no. 1, article 756, 2014.

[30] J. P. Demuth and M. W. Hahn, "The life and death of gene families," *BioEssays*, vol. 31, no. 1, pp. 29–39, 2009.

[31] B. Piasecka, P. Lichocki, S. Moretti, S. Bergmann, and M. Robinson-Rechavi, "The hourglass and the early conservation models—co-existing patterns of developmental constraints in vertebrates," *PLoS Genetics*, vol. 9, no. 4, Article ID e1003476, 2013.

[32] G. A. Wray, "The evolutionary significance of cis-regulatory mutations," *Nature Reviews Genetics*, vol. 8, no. 3, pp. 206–216, 2007.

[33] P. G. Engström, D. Fredman, and B. Lenhard, "Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes," *Genome Biology*, vol. 9, no. 2, article R34, 2008.

[34] A. C. Burke, C. E. Nelson, B. A. Morgan, and C. Tabin, "Hox genes and the evolution of vertebrate axial morphology," *Development*, vol. 121, no. 2, pp. 333–346, 1995.

[35] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, "Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells," *Science*, vol. 343, no. 6167, pp. 193–196, 2014.

[36] F. H. Biase, X. Y. Cao, and S. Zhong, "Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing," *Genome Research*, vol. 24, no. 11, pp. 1787–1796, 2014.

[37] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[38] G. P. Wagner, K. Kin, and V. J. Lynch, "Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples," *Theory in Biosciences*, vol. 131, no. 4, pp. 281–285, 2012.

[39] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, article 17, 2005.

[40] A. M. Yip and S. Horvath, "Gene network interconnectedness and the generalized topological overlap measure," *BMC Bioinformatics*, vol. 8, article 22, 2007.

[41] P. Langfelder, B. Zhang, and S. Horvath, "Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R," *Bioinformatics*, vol. 24, no. 5, pp. 719–720, 2008.

[42] D. Smedley, S. Haider, B. Ballester et al., "BioMart—biological queries made easy," *BMC Genomics*, vol. 10, article 22, 2009.

[43] P. Flicek, M. R. Amode, D. Barrell et al., "Ensembl 2014," *Nucleic Acids Research*, vol. 42, no. 1, pp. D749–D755, 2014.