

# Single-Cell Analysis of Alternative Splicing and Gene Regulatory Network Reveals Remarkable Expression and Regulation Dynamics During Early Embryonic Development

**Jiwei Chen**

East China Normal University School of Life Sciences <https://orcid.org/0000-0002-8783-8134>

**Yunjin Li**

East China Normal University School of Life Sciences

**Geng Chen**

East China Normal University School of Life Sciences

**Tieliu Shi** (✉ [tieliushi@yahoo.com](mailto:tieliushi@yahoo.com))

East China Normal University School of Life Sciences

---

## Research

**Keywords:** single cell RNA-seq, alternative splicing, major isoform switching, differential expression, gene regulatory network, early embryonic development

**Posted Date:** February 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-190947/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Single-cell RNA-seq (scRNA-seq) technologies greatly revolutionized our understanding of cell-to-cell variability of gene expression. Although several studies investigated the expression profile of early embryos, they mainly focused on the expression changes at gene level. Here we systematically explored the gene expression dynamics of human early embryonic development from expression level, alternative splicing, isoform switching and expression regulatory network.

## Results

We found that the genes involved in significant changes of these three aspects are all gradually decreased along embryonic development from E3 to E7 stage. Moreover, these three types of variations are complementary for profiling expression dynamics and they vary greatly across embryonic development as well as between different sexes. Strikingly, only a small number of genes exhibited prominent expression level changes between male and female embryos in E3 stage, whereas many more genes showed variations in alternative splicing and major isoform switching. Additionally, we identified functionally important specific gene regulatory modules for each stage and revealed dynamic usage of transcription factor binding motifs (TFBMs).

## Conclusions

Collectively, our study gain insights into the expression dynamics of early embryonic development from expression level, alternative splicing, isoform switching and gene regulatory networks, which could benefit the understanding of underlying mechanism of embryonic development.

## Background

The advent of single-cell RNA-seq (scRNA-seq) technologies enables the exploration of gene expression at single-cell resolution. To date, scRNA-seq has been widely applied to explore cell subtypes, gene expression dynamics and heterogeneity of diverse tissues in many species [1-8], which have revealed various meaningful results. By contrast, traditional bulk RNA-seq technologies mainly reflect the averaged gene expression of a large number of cells, which are inappropriate to study gene expression heterogeneity especially early embryonic development since only a small number of cells are contained. With the innovation of scRNA-seq, an increasing number of different scRNA-seq protocols have been developed, but only the full-length transcript capturing scRNA-seq technologies (such as Smart-seq2 [9]) allow the analysis of alternative splicing rather than the 3'-end capturing protocols (e.g. drop-seq [10]) [8]. However, the great majority of the published scRNA-seq studies mainly focused on the expression changes at gene level, which ignored the expression dynamics at isoform and exon levels.

Previously, Yan et al. first investigated the transcriptome landscape of human early embryos from oocyte to blastocyst, where they revealed dynamic expression patterns across different developmental stages [11]. Then Xue et al. further compared the gene expression profiles between humans and mice from oocyte to morula, and highlighted the evolutionary conservation of some expression networks in early embryonic development [12, 13]. Recently, Petropoulos et al. provided a single-cell transcriptome profiling of a number of human preimplantation embryos across five developmental stages (E3-E7) [14], but they did not explore the alternative splicing and gene regulatory profiles of early embryos. Both of the changes in gene expression level (gene-level) and alternative splicing (exon-level) are crucial for understanding the gene expression dynamics of early embryonic development. Moreover, isoform switching (isoform-level) is another important aspect of gene expression since most of human genes could encode multiple different isoforms, but none of these studies investigated and compared these different-level changes. Besides, a network view of dynamic gene expression regulations during embryonic development for unraveling developmental regulatory changes is still lacking. Consequently, it is necessary to further explore the expression of genes from different aspects (gene-level, isoform-level, exon-level and regulatory network) to better and more comprehensively understand the gene expression dynamics and the underlying mechanisms of embryonic development.

Here we further comprehensively analyzed the gene expression profile of human early embryonic development process from distinct levels of gene, isoform, exon and regulatory network based on the scRNA-seq data of 1,529 cells from 88 male and female embryos. Besides differential gene expression analysis, we also investigated the changes of alternative splicing and major isoform switching along embryonic development as well as the sex differences between male and female embryos. Furthermore, we compared the genes and related pathways involved in differential expression, significantly alternative splicing and major isoform switching, and revealed large inconsistency among these aspects of gene expression changes. Finally, we constructed gene expression regulatory network of embryonic development and identified stage-specific regulatory modules as well as enriched motifs bound by corresponding transcription factors. Our systematic analyses revealed interesting gene expression dynamics from diverse aspects, which deepens the understanding of underlying processes and mechanisms of human early embryonic development.

## Results

### **Gene expression level, alternative splicing and major isoform switching are very dynamic across embryonic development**

To systematically investigate the gene expression changes from different levels during embryonic development, we first explored the changes of gene expression level, alternative splicing and major isoform switching between adjacent embryonic stages based on 1,529 scRNA-seq data of human early embryos from E3 to E7 stages. The software of SCDE was employed to carry out differential gene expression calling for a total of 58,302 genes (Ensembl gene annotation). Using the threshold of FDR < 0.05, we detected thousands of differentially expressed genes (DEGs) in each group of adjacent stage

comparisons (Figure 1B). Interestingly, the number of stage-related DEGs were gradually decreased along embryonic development, where the largest number of genes were shown significant expression changes between E3 and E4 stages, while the lowest count of DEGs was identified in E6 versus E7 (Figure 1B). Moreover, more genes were changed in expression level in male embryos compared to that of female embryos during embryonic development except E6 versus E7. In adjacent stage comparisons, a large portion of DEGs shared between male (47.8% ~ 72.4%) and female (27.7% ~ 76.3%) embryos. Surprisingly, only 127 DEGs were common among those neighbouring stage comparisons, leaving most of the DEGs were different between any two groups (Figure 1C). Gene functional enrichment analysis showed that these 127 DEGs were mainly enriched in the pathways/biological processes of oxidative phosphorylation, autophagy-animal, aminoacyl-tRNA biosynthesis, aldosterone-regulated sodium reabsorption and lysosome (see Supplementary Figure 2A).

We further explored the alternative splicing changes between adjacent developmental stages using BRIE. Similar with the trend of stage-related DEGs, the amount of differential alternative splicing genes (DASGs) was also gradually reduced during embryonic development (Figure 1D). Specifically, the counts of DASGs between neighbouring stages range from 2,415 (E3 versus E4) to 1,183 (E6 versus E7). More genes were also exhibited differential alternative splicing in male embryos compared to that of female embryos in each neighbouring stage comparison. We also observed that a large portion of stage-related DASGs were common between male (28.8% ~ 57.3%) and female (24.2% ~ 51.8%) embryos, while the percentages of shared DASGs were smaller than that of corresponding stage-related DEGs. Those stage comparison groups only have 96 common DASGs, while each group possesses hundreds of unique DASGs (Figure 1E). These 96 genes are enriched in the pathways/biological processes of autophagy-animal, ErbB signaling pathway, adipocytokine signaling pathway, prolactin signaling pathway and mTOR signaling pathway (see Supplementary Figure 2B).

Considering that the major isoform (the isoform with the highest expression of a gene) switching is also an important aspect of gene expression changes, we then examined the isoform switching events between adjacent embryonic stages. Strikingly, the number of major isoform switching genes (MISGs) was gradually decreased with embryonic development as well, which was consistent with that of DEGs and DASGs (Figure 1F). Except the comparison of E3 versus E4, a bigger number of genes were switched their major isoforms in male embryos than that of female in other stage comparisons. Notably, only four MISGs (*SLC39A7*, *TMC7*, *CLDN6* and *DDR1*) shared among distinct comparison groups (Figure 1G). *SLC39A7* is zinc transporter, plays a critical role in regulating cell growth and death. Yan et al. found *SLC39A7* is required for eye, brain, and skeleton formation during early embryonic development in zebrafish [15]. The function of *TMC7* is not very clear. Recent study of Cheng et al. revealed that *TMC7* is overexpressed in pancreatic carcinoma and contribute to tumor progression and metastasis [16]. *CLDN6* is one of the earliest molecules expressed in embryonic stem cells committed to epithelial differentiation in both murine and human tissues [17, 18]. *DDR1* has been proved to be functionally important in differentiation, cell motility, collagen synthesis and signaling [19]. Consequently, our results show that the changes of gene expression level, alternative splicing and major isoform switching are all decreased

during early embryonic development, and the genes involved in these three types of changes varied greatly across different embryonic stages.

### **Male and female embryos have large differences in gene expression, alternative splicing and major isoform usage**

To further check the expression changes associated with sex differences during early embryonic development, we further explored the DEGs, DASGs and MISGs between male and female embryos in each stage. E6 stage had the greatest amount of DEGs, while E3 stage exhibited the fewest number (only 47) of genes changed in expression for gender comparison (Figure 2A). Notably, 32 of those 47 DEGs (68.1%, 20 X chromosome genes and 12 Y chromosome genes) were from sex chromosomes, suggesting that most of the autosome genes do not varied in expression level in E3 stage (see Supplementary Table 1). In contrast, a number of genes were significantly changed in alternative splicing and major isoform switching between male and female embryos of E3. Specifically, 1,038 sex-related DASGs were detected in E3 which is the second largest compared to other stages (Figure 2B), and the largest number of sex-related MISGs (1,248) were identified in E3 (Figure 2C). Thus although only a small fraction of genes exhibited significant changes in expression level between male and female embryos of E3 stage, great differences exist in alternative splicing and major isoform switching. Only 17 DEGs, 13 DSAGs and 5 MISGs shared among different stages of sex comparison and a large portion of unique DEGs, DSAGs and MISGs were identified in each stage (Figure 2D-F). Interestingly, all these 17 DEGs are from sex chromosomes (six of them come from X chromosome and the remaining 11 are from Y chromosome), and 12 of the 13 DASGs are encoded by autosomes, and four of the 5 MISGs come from sex chromosomes (see Supplementary Table 1). Accordingly, the genes involved in expression level changes are largely distinct from those ones significantly changed in alternative splicing and isoform switching for gender comparison in early embryonic development.

Moreover, we found that a significant portion of sex-related DEGs are from sex chromosomes, but the great majority of sex-related DSAGs and MISGs are expressed by autosomes. Only 47 DEGs were detected in E3 stage between male and female, 68.1% of these sex-related DSAGs were from sex chromosomes (20 X chromosome genes and 12 Y chromosome genes) and E4 stage had 52.75% of sex-related DEGs encoded by sex chromosomes, while much smaller portion of sex-related DEGs were generated from sex chromosomes for E5 (17.49%), E6 (10.81%) and E7 (13.16%) stages, indicating that many autosome genes mainly exhibited gene expression level changes between male and female embryos after E4 stage. By contrast, lots of autosome genes showed significant alternative splicing variations and major isoform switching between different sexes from E3 stage, where only 2.28% ~ 4.07% of DASGs and 4.36% ~ 5.95% of MISGs were from sex chromosomes (see Supplementary table 1). Therefore, expression level profile of genes are much different from that of alternative splicing and major isoform switching between male and female embryos during early embryonic development.

### **Gene expression level, alternative splicing and isoform switching are complementary in profiling gene expression variations**

Then we compared the DEGs, DASGs and MISGs resulted from the comparisons of embryonic development and different sexes. Strikingly, only a small portion of genes were common between any two types of DEGs, DASGs and MISGs for stage comparisons (Figure 3A-D), which is similar with the result of sex comparisons for distinct stages (Figure 3E-I). This could be explained by that DEGs, MISGs and DASGs mainly reflect the expression changes at levels of gene, isoform and exon, respectively. Exon level variation in expression may not affect the whole expression at gene and isoform levels, and isoform level changes do not necessarily result in the expression change at gene level as well. Moreover, for the genes shared among DEGs, DASGs and MISGs, the percentages for stage comparisons ranged from 3.9% to 11.4%, while the ratios were smaller for sex comparisons (0.5% ~ 5.3%). Combining the three types of genes, DEGs occupied the largest portion of genes followed by DASGs and MISGs along the embryonic development (Figure 3A-D). However, distinct results were observed between male and female embryos. The DEGs detected from sex comparisons took up the smallest fractions among these three types of genes in both E3 and E4 stages, but occupied the largest portions in sex comparisons of E5, E6 and E7 stages (Figure 3E-I). Therefore, the genes involved in changes of expression level, alternative splicing and major isoform switching are largely different across distinct embryonic developmental stages as well as between disparate sexes. Our results also show that DEGs, DASGs and MISGs reflect the expression changes of genes from distinct aspects, which are complementary for dissecting gene expression dynamics.

### **DEGs, DASGs and MISGs are involved in distinct but important pathways related to embryos**

To gain insights into the functions of those DEGs, DASGs and MISGs, we first conducted gene functional enrichment analysis for stage-related genes using clusterProfiler [20]. As expected, a notable portion of the significantly enriched pathways were distinct among stage-related DEGs, DASGs and MISGs. Specifically, the stage-related DEGs are mainly involved in thermogenesis, oxidative phosphorylation and adherens junction (Figure 4A), whereas stage-related DASGs are mainly enriched in autophagy, ubiquitin mediated proteolysis and phosphatidylinositol signaling system (Figure 4B), and the top enriched pathways for stage-related MISGs are ribosome, cell cycle and HIF-1 signaling pathway. Intriguingly, we found that most of the significantly enriched pathways for stage-related MISGs and DASGs are common with that of stage-related DEGs (Figure 4D,  $P$ -value < 0.05). But only 10 significantly enriched pathways shared among the DEGs, DASGs and MISGs of stage comparisons (such as endocytosis, autophagy- animal, mitophagy, mTOR signaling pathway, Fc gamma R-mediated phagocytosis, endometrial cancer, choline metabolism in cancer, tight junction, AMPK signaling pathway and ubiquitin mediated proteolysis), and each type of stage-related genes enriched in some unique pathways (Figure 4D,  $P$ -value < 0.05).

Next, we carried out functional enrichment analysis for those DEGs, DASGs and MISGs identified in gender comparisons. The enriched pathways differ greatly among these three types of genes, which shows larger differences with that of stage comparisons. The top enriched pathways for sex-related DEGs are oxidative phosphorylation, protein processing in endoplasmic reticulum and thermogenesis (Figure 4E), while sex-related DASGs are mainly involved in ubiquitin mediated proteolysis, autophagy-

animal, and phosphatidylinositol signaling system (Figure 4F). The sex related MISGs are mainly enriched in spliceosome, ribosome, cell cycle, and RNA transport (Figure 4G). Surprisingly, no significantly enriched pathways are common among sex related DEGs, DASGs and MISGs, suggesting that these three types of genes mainly function in distinct pathways.

### Functionally important regulatory modules are identified in early embryonic development

To explore the gene expression regulations during early embryonic development, we further constructed the gene expression regulatory network in early embryos by employing SCENIC [21]. A total of 506 regulons that the modules are with significant motif enrichment of upstream regulating TFs were identified. Each regulon contains a transcription factor (TF) and the target genes likely regulated by this TF. Among these 506 TFs of regulons, 277 were stage-related DEGs, 80 were stage-related DASGs and 70 were stage-related MISGs. Thus most of those TFs were involved at least one level changes of gene expression, alternative splicing and major isoform switching. Interestingly, the embryonic developmental process from E3 to E7 can be clearly revealed based on those 506 regulons in t-SNE (t-Distributed Stochastic Neighbor Embedding), suggesting that the gene regulatory networks vary greatly among different developmental stages (Figure 5A). The male and female cells from the same embryonic stage mixed together in t-SNE, indicating that male and female embryos share similar regulatory network in the same stage. In order to identify stage specific regulons, we further examined the genes with significantly enriched expression in each stage using the method of analysis of variance (ANOVA). With the criteria of fold change > 2 and adjusted *P*-value < 0.01, E3 had the largest number (2,977) of genes with enriched expression (EEGs), whereas E6 possessed only 16 EEGs, which was the smallest. For E4, E5 and E7 stages, 530, 50 and 193 EEGs were identified, respectively (Figure 5B). Then we compared these EEGs with those 506 regulons and found that 56, 9, 1, and 9 specific regulons were separately detected E3, E4, E5 and E7 stages. No specific regulons were identified in E6 stage, which could be resulted from the fact that the gene expression profile in E6 was similar with that of E7 since that the smallest amount of DEGs, DASGs and MISGs were detected in E6 versus E7 compared to other adjacent stage comparisons.

The regulon of TF *CREB1* is one of the specific regulons in E3 stage, where 602 genes potentially regulated by *CREB1* were involved (Figure 5C). *CREB1* is a critical TF that regulates ~25% of the eukaryotic genome and plays an important role in the synchronization of circadian rhythmicity, the differentiation of adipose cells, initial development of the nervous system, memory formation, and neuronal protection [22-24]. Some studies indicate that *CREB1* mediates signals essential for maintaining cell viability during early embryonic development of mouse [25, 26]. Gene functional enrichment analysis showed that those 602 genes were mainly involved in the biological processes of histone modification, covalent chromatin modification and chromosome segregation (*P*-value < 0.05). The regulon of TF *ATF3* and its 461 target genes were mainly expressed in E4 stages (Figure 5D). *ATF3* is induced by a variety of signals, including cell cycling, neutrophil migration and sexual differentiation [27-29]. Cheng et al. revealed that *ATF3* has an important role in regulating human endometrial receptivity and embryo attachment in vitro via up-regulation of leukemia inhibitory factor [30]. Those 461 genes were enriched in the biological processes of anion transmembrane transport, response to corticosteroid and response to

glucocorticoid ( $P$ -value < 0.05). The only one specific regulon of E5 stage contains TF *NFKB2* and its 42 target genes (Figure 5E). A central role for the *NFKB2* gene is the maintenance of the peripheral B-cell population, humoral immune responsiveness, and splenic architecture [31]. But no previous studies show the association between *NFKB2* and embryonic development, suggesting that *NFKB2* is a novel TF specifically expressed E5 stage. Those 42 targeting genes of *NFKB2* were mainly involved in epithelial tube morphogenesis, protein autophosphorylation and regulation of smooth muscle cell proliferation ( $P$ -value < 0.05). In E7 stage, TF *GCM1* coupled with 38 genes formed a specific regulon (Figure 5F). Proteins encoded by *GCM1* are crucial for mediating the differentiation of trophoblast cells along both the villous and extravillous pathways in placental development [32]. Those 38 targeting genes were mainly enriched in leukocyte migration and cell chemotaxis ( $P$ -value < 0.05).

### **The usage of transcription factor binding motifs for TFs is very dynamic along embryo development**

Since the binding sequences of TFs are usually conserved, we further investigated the transcription factor binding motifs (TFBMs) for those 506 regulons by employing RcisTarget [21]. Interestingly, the counts of enriched TFBMs for those 506 regulons varied tremendously (range from 1 to 2,601). It is notable that TF *GABPA* has the largest number (2,601) of TFBMs. *GABPA* belongs to the E-twenty-six (ETS) family of DNA-binding factors and regulates a broad range of genes involved in cell cycle control, apoptosis, differentiation, hormonal regulation, and other critical cellular functions [33]. Most of the TFs in those regulons possessed 1~50 TFBMs (see Supplementary Figure 3). Over 91% (461 out of 506) of TFs had at least two TFBSs, suggesting that these TFs exhibit dynamic TFBS usage in early embryonic development. Specifically, the aforementioned stage-specific TFs *CREB1*, *ATF3*, *NFKB2* and *GCM1* for E3, E4, E5 and E7 stages possessed 6, 223, 48 and 8 TFBMs, respectively. Some of those TFBMs for the stage-specific TFs are shown in Figure 5C-F. Thus, our result show that the TFBM usage for those 506 TFs of regulons are very dynamic in early embryonic development and distinct TFBMs could be used in different developmental stages.

## **Discussion**

In this study, we systematically explored the gene expression profile of human early embryos from the changes of gene-level (differential expression calling), isoform level (major isoform switching), exon-level (alternative splicing analysis) and expression regulatory network inference. Intriguingly, a common trend among DEGs, DASGs and MISGs is that their quantity is gradually decreased along the developmental stages, indicating that the expression changes is reduced during embryonic development. But for the comparison between male and female embryos of each stage, we did not observe similar trend as stage comparisons. The number of sex-related DEGs, DASGs and MISGs detected in different stages vary greatly as well. An interesting result is that only 47 sex-related DEGs were detected in E3 stage; however, 1,038 DASGs and 1,284 MISGs were identified between male and female comparison of E3. The result indicates that although the expression has change little between different sexes at gene level in E3 stage, great variations existed in alternative splicing and isoform switching of genes. Moreover, the stage-related DEGs, MISGs and DASGs had another similar tendency that the majority of each type of genes were

distinct among the comparisons of E3-E4, E4-E5, E5-E6 and E6-E7, suggesting that the expression changes varied widely across different levels of gene, isoform and exon. Gene functional enrichment analysis results showed that those DEGs, DASGs and MISGs were mainly enriched different pathways, which is reasonable since the large differences of distinct types of genes. Therefore, combing expression level, alternative splicing and isoform switching can reveal a broader range of gene expression changes.

Alternative splicing enables the multi-exon genes to generate different protein-coding and/or noncoding isoforms, which largely increased the diversity of transcriptome and can influence both of gene expression level and isoform switching events [34]. When a gene executes specific role/function in distinct conditions, it may employ disparate isoforms and change the expression levels of corresponding isoforms, resulting the switching of major isoform [34]. In fact, gene expression level reflects the total expression changes of all the isoforms encoded by this gene, while isoform switching and alternative splicing separately show the expression variations of isoforms and exons. Therefore, the genes involved in expression changes among the levels of gene, isoform and exon could vary tremendously. Currently, most scRNA-seq studies mainly focused on the expression level changes of genes, which missed a large number of genes that significantly changed in alternative splicing and isoform switching [35]. Our results indicate that the genes significantly changed in these three aspects are much different in early embryonic development, and they are complementary for elucidating the dynamic expression profile of genes. To the best of our knowledge, our study is the first to simultaneously explore the expression variations from gene expression level, alternative splicing and isoform switching.

Additionally, we identified 506 functional important regulons related to early embryonic development including some stage-specific regulons (such as E3: *MSX1*, E4: *SOX2* and E7: *TFEB*), and revealed dynamic usages of TFBMs for those involved TFs. *MSX1* encodes transcription factors that control organogenesis and tissue interactions during embryonic development [36]. Systemic deletion of *MSX1* in mice results in perinatal lethality [37]. *SOX2* is necessary for the maintenance of pluripotency in epiblast and embryonic stem cells and knockout of *SOX2* is early embryonic lethal [38, 39]. Later in development, *SOX2* is required in various tissue stem cells and early progenitors, especially in the nervous system [40]. Eiríkur et al. showed that *TFEB* plays a critical role in the signal transduction processes required for normal vascularization of the placenta [41]. Therefore, the stage-specific regulons identified by us are functionally important for embryonic development.

## Conclusions

Collectively, our study gains insights into the dynamic changes of gene activities during early embryonic development from gene expression level, alternative splicing, isoform switching and gene expression regulatory network. The systematic analyses of gene expression profile from different aspects can deepen the understanding of gene expression variations and the underlying molecular mechanisms.

## Methods

## Single-cell RNA-seq data processing and sex determination of early embryos

A total of 1,529 scRNA-seq data of 88 human male and female early embryos (from stage E3 to E7) were first downloaded from ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>) with accession number E-MTAB-3929. Then the scRNA-seq data were mapped to the human reference genome GRCh38 using HISAT2 [42] (version 2.1.0) with default parameters. Expression of genes and transcripts were quantified in Transcripts Per Kilobase Million (TPM) by StringTie [43] (version 1.3.3b) with parameters of “-e -A” based on the human gene annotation file of Ensembl database in GTF format (version 91). In order to determine the sex of each cell, the sum of TPM values for seven Y chromosome genes (DDX3Y, EIF1AY, KDM5D, PRKY, RPS4Y1, UTY and ZFY) were used to distinguish male embryos (cut-off:  $\geq 60$  TPM) from female embryos (cut-off:  $\leq 40$  TPM) (see Supplementary Figure 1) [14].

## Differential expression calling of scRNA-seq data

Differential gene expression analyses between adjacent embryonic stages or between male and female embryos were conducted by employing the tool of Single-Cell Differential Expression (SCDE, version 2.8.0) [44]. Then differentially expressed genes (DEGs) were determined using the criteria of  $|cZ| > 1.96$  which corresponds to  $FDR < 0.05$ .

## Differential alternative splicing analysis of human early embryos

To identify the alternative splicing events that significantly differentially changed between distinct conditions, we used the software specifically designed for scRNA-seq data of BRIE [45] (version 0.2.0) to analyze the differential alternative splicing events between male and female cells or between adjacent stages. Then differential alternative splicing genes (DASGs) were selected with the threshold of Bayes factor  $> 10$ .

## Major isoform switching analysis

To investigate the genes switched the isoform with the highest expression (major isoform) in different conditions, the transcript expressions of each multi-isoform gene in each sample were ranked from large to small according to their expression level. Considering that each embryonic stage or each gender of a certain stage contains many single cells, we defined the transcript that has the highest expression among all isoforms of a gene in at least 60% of the cells for a given condition as the major isoform. Then the major isoform of each multi-transcript gene in each condition were determined. The genes switched major isoforms were denoted as MISGs (major isoform switching genes).

## Gene functional enrichment analysis of GO terms and pathways

Gene Ontology (GO) enrichment analysis of biological process and KEGG pathway enrichment analysis were carried out for corresponding gene set using R package clusterProfiler [20] (version 3.8.1). Only those GO terms and pathways with  $P$ -value  $< 0.05$  were considered as statistically significant.

## **Construction of gene expression regulatory network**

Gene expression regulatory network inference was conducted by SCENIC [21] (version 1.0.1.1) following the standard analysis pipeline. After the filtering process of SCENIC, the count matrix of genes was used as input to perform co-expression analysis based on GENIE3 [46] (version 1.4.0) and the co-expressed genes potentially regulated by transcription factors (TFs) were identified. The TF co-expression modules were then confirmed by RcisTarget [21] (version 1.2.0) using human GRCh38/hg38 database. The significant TF-modules identified by SCENIC were defined as regulons.

## **Identification of stage specific regulons and motifs**

In order to identify the regulons and motifs specific to each stage, we first performed gene expression enrichment analysis through the analysis of variance (ANOVA) based on the gene expression profile of human early embryos. ANOVA followed by Tukey's range test (Tukey's honest significance differences) was employed to detect the genes with significantly higher expression in one stage compared to other stages (fold change > 2 and adjusted *P*-value < 0.01). Then the TFs with significantly enriched expression in each stage were obtained. Combining with the regulons identified by SCENIC pipeline, the regulons of TFs that specifically expressed in each stage were determined and the motifs bound by those TFs were defined as stage-specific motifs.

## **Declarations**

### **Ethics approval and consent to participate**

Not applicable

### **Consent for publication**

Not applicable

### **Availability of data and materials**

All data analysed during this study are included in this published article "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos".

### **Competing interests**

The authors declare that they have no competing interests

### **Funding**

This work was supported by the China Human Proteome Project (Grant No.2014DFB30010, 2014DFB30030), National Natural Science Foundation of China (31771460, 31671377) and Shanghai 111 Project (B14019).

## Authors' contributions

JC and YL analyzed the data. JC, GC and TS wrote the manuscript. GC and TS conceived and supervised the study. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable

## References

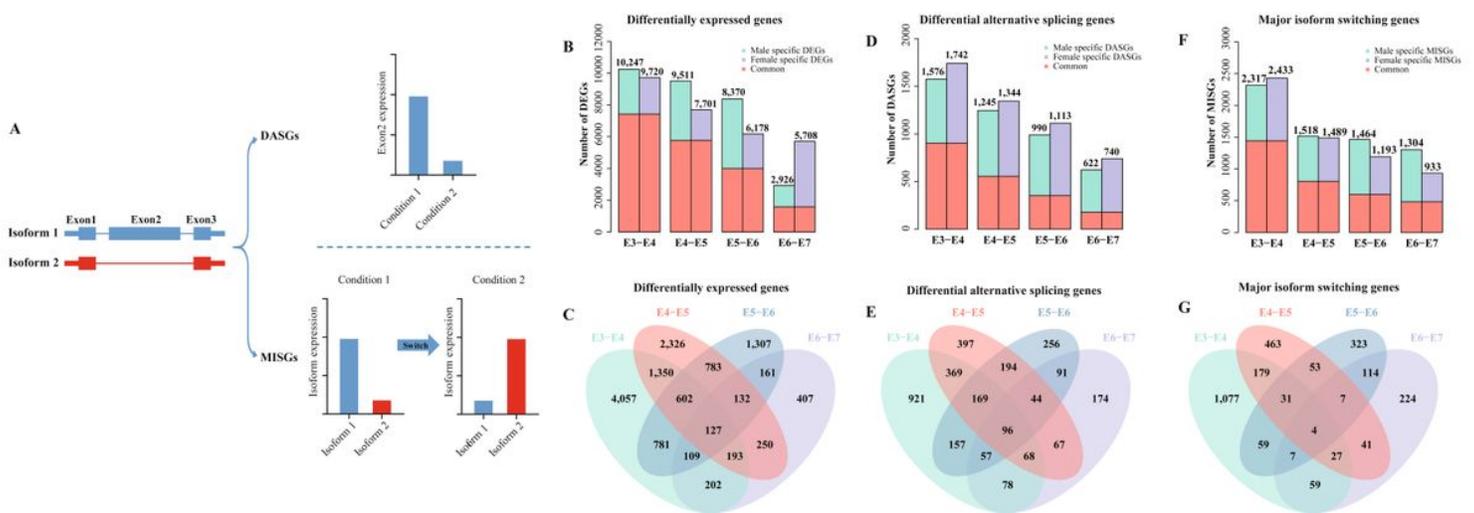
1. Loo L, Simon JM, Xing L, McCoy ES, Niehaus JK, Guo J, Anton ES, Zylka MJ: Single-cell transcriptomic analysis of mouse neocortical development. *Nature communications* 2019, 10(1):134.
2. Hammond TR, Dufort C, Dissing-Olesen L, Giera S, Young A, Wysoker A, Walker AJ, Gergits F, Segel M, Nemesh J *et al*: Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* 2019, 50(1):253-271 e256.
3. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, Koopmans F, Barres B, Smit AB, Sloan SA *et al*: Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nature biotechnology* 2018.
4. Paik DT, Tian L, Lee J, Sayed N, Chen IY, Rhee S, Rhee JW, Kim Y, Wirka RC, Buikema JW *et al*: Large-Scale Single-Cell RNA-Seq Reveals Molecular Signatures of Heterogeneous Populations of Human Induced Pluripotent Stem Cell-Derived Endothelial Cells. *Circulation research* 2018, 123(4):443-450.
5. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft L, Aibar S, Makhzami S, Christiaens V, Bravo Gonzalez-Blas C *et al*: A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* 2018, 174(4):982-998 e920.
6. Sebe-Pedros A, Saudemont B, Chomsky E, Plessier F, Mailhe MP, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S *et al*: Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* 2018, 173(6):1520-1534 e1520.
7. Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, Zhang H, Li L, Sun L, Pan N *et al*: A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* 2018, 555(7697):524-528.
8. Chen G, Ning B, Shi T: Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Frontiers in Genetics* 2019, 10(317).
9. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R: Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* 2014, 9(1):171-181.
10. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM *et al*: Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 2015, 161(5):1202-1214.

11. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J *et al*: Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology* 2013, 20(9):1131-1139.
12. Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE *et al*: Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 2013, 500(7464):593-597.
13. Xue L, Cai JY, Ma J, Huang Z, Guo MX, Fu LZ, Shi YB, Li WX: Global expression profiling reveals genetic programs underlying the developmental divergence between mouse and human embryogenesis. *BMC genomics* 2013, 14:568.
14. Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Reyes AP, Linnarsson S, Sandberg R, Lanner F: Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 2016, 167(1):285.
15. Yan G, Zhang Y, Yu J, Yu Y, Zhang F, Zhang Z, Wu A, Yan X, Zhou Y, Wang F: Slc39a7/zip7 plays a critical role in development and zinc homeostasis in zebrafish. *PloS one* 2012, 7(8):e42939.
16. Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, Gong S, Zhu Y: Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma. *EBioMedicine* 2019, 40:382-393.
17. Morita K, Furuse M, Fujimoto K, Tsukita S: Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96(2):511-516.
18. Turksen K, Troy TC: Claudin-6: a novel tight junction molecule is developmentally regulated in mouse embryonic epithelium. *Developmental dynamics : an official publication of the American Association of Anatomists* 2001, 222(2):292-300.
19. Faraci-Orf E, McFadden C, Vogel WF: DDR1 signaling is essential to sustain Stat5 function during lactogenesis. *Journal of cellular biochemistry* 2006, 97(1):109-121.
20. Yu G, Wang LG, Han Y, He QY: clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* 2012, 16(5):284-287.
21. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J *et al*: SCENIC: single-cell regulatory network inference and clustering. *Nature methods* 2017, 14(11):1083-1086.
22. Dworkin S, Heath JK, deJong-Curtain TA, Hogan BM, Lieschke GJ, Malaterre J, Ramsay RG, Mantamadiotis T: CREB activity modulates neural cell proliferation, midbrain-hindbrain organization and patterning in zebrafish. *Developmental biology* 2007, 307(1):127-141.
23. Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH: Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* 2004, 119(7):1041-1054.
24. Zhang X, Odom DT, Koo SH, Conkright MD, Canettieri G, Best J, Chen H, Jenner R, Herbolsheimer E, Jacobsen E *et al*: Genome-wide analysis of cAMP-response element binding protein occupancy,

- phosphorylation, and target gene activation in human tissues. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(12):4459-4464.
25. Jin XL, O'Neill C: cAMP-responsive element-binding protein expression and regulation in the mouse preimplantation embryo. *Reproduction* 2007, 134(5):667-675.
  26. Bleckmann SC, Blendy JA, Rudolph D, Monaghan AP, Schmid W, Schutz G: Activating transcription factor 1 and CREB are important for cell survival during early mouse development. *Molecular and cellular biology* 2002, 22(6):1919-1925.
  27. Liu B, Agras K, Willingham E, Vilela ML, Baskin LS: Activating transcription factor 3 is estrogen-responsive in utero and upregulated during sexual differentiation. *Hormone research* 2006, 65(5):217-222.
  28. Boespflug ND, Kumar S, McAlees JW, Phelan JD, Grimes HL, Hoebe K, Hai T, Filippi MD, Karp CL: ATF3 is a novel regulator of mouse neutrophil migration. *Blood* 2014, 123(13):2084-2093.
  29. Thompson MR, Xu D, Williams BR: ATF3 transcription factor and its emerging roles in immunity and cancer. *Journal of molecular medicine* 2009, 87(11):1053-1060.
  30. Cheng X, Liu J, Shan H, Sun L, Huang C, Yan Q, Jiang R, Ding L, Jiang Y, Zhou J *et al*: Activating transcription factor 3 promotes embryo attachment via up-regulation of leukemia inhibitory factor in vitro. *Reproductive biology and endocrinology : RB&E* 2017, 15(1):42.
  31. Heusch M, Lin L, Gelezianas R, Greene WC: The generation of nfkb2 p52: mechanism and efficiency. *Oncogene* 1999, 18(46):6201-6208.
  32. Baczyk D, Drewlo S, Proctor L, Dunk C, Lye S, Kingdom J: Glial cell missing-1 transcription factor is required for the differentiation of the human trophoblast. *Cell death and differentiation* 2009, 16(5):719-727.
  33. Perdomo-Sabogal A, Nowick K, Piccini I, Sudbrak R, Lehrach H, Yaspo ML, Warnatz HJ, Querfurth R: Human Lineage-Specific Transcriptional Regulation through GA-Binding Protein Transcription Factor Alpha (GABPa). *Molecular biology and evolution* 2016, 33(5):1231-1244.
  34. Chen G, Shi T, Shi L: Characterizing and annotating the genome using RNA-seq data. *Science China Life sciences* 2017, 60(2):116-125.
  35. Peng G, Jing N: The genome-wide molecular regulation of mouse gastrulation embryo. *Science China Life sciences* 2017, 60(4):363-369.
  36. Alappat S, Zhang ZY, Chen YP: Msx homeobox gene family and craniofacial development. *Cell research* 2003, 13(6):429-442.
  37. Satokata I, Maas R: Msx1 deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nature genetics* 1994, 6(4):348-356.
  38. Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA *et al*: Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology* 2007, 9(6):625-635.

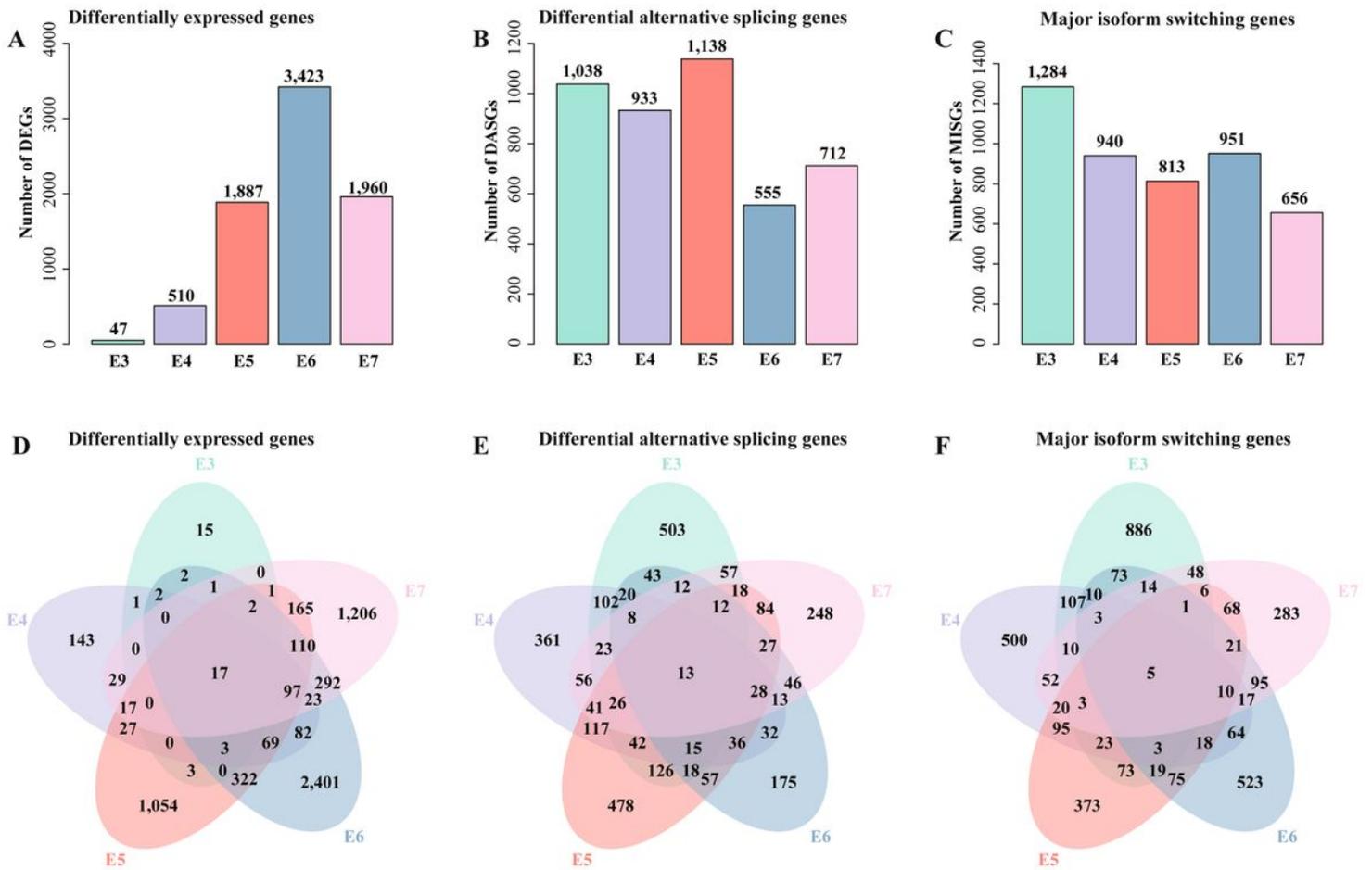
39. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R: Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* 2003, 17(1):126-140.
40. Pevny LH, Nicolis SK: Sox2 roles in neural stem cells. *The international journal of biochemistry & cell biology* 2010, 42(3):421-424.
41. Steingrimsson E, Tessarollo L, Reid SW, Jenkins NA, Copeland NG: The bHLH-Zip transcription factor Tfeb is essential for placental vascularization. *Development* 1998, 125(23):4607-4616.
42. Kim D, Langmead B, Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 2015, 12(4):357-360.
43. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* 2015, 33(3):290-295.
44. Kharchenko PV, Silberstein L, Scadden DT: Bayesian approach to single-cell differential expression analysis. *Nature methods* 2014, 11(7):740-742.
45. Huang Y, Sanguinetti G: BRIE: transcriptome-wide splicing quantification in single cells. *Genome biology* 2017, 18(1):123.
46. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: Inferring regulatory networks from expression data using tree-based methods. *PloS one* 2010, 5(9).

## Figures



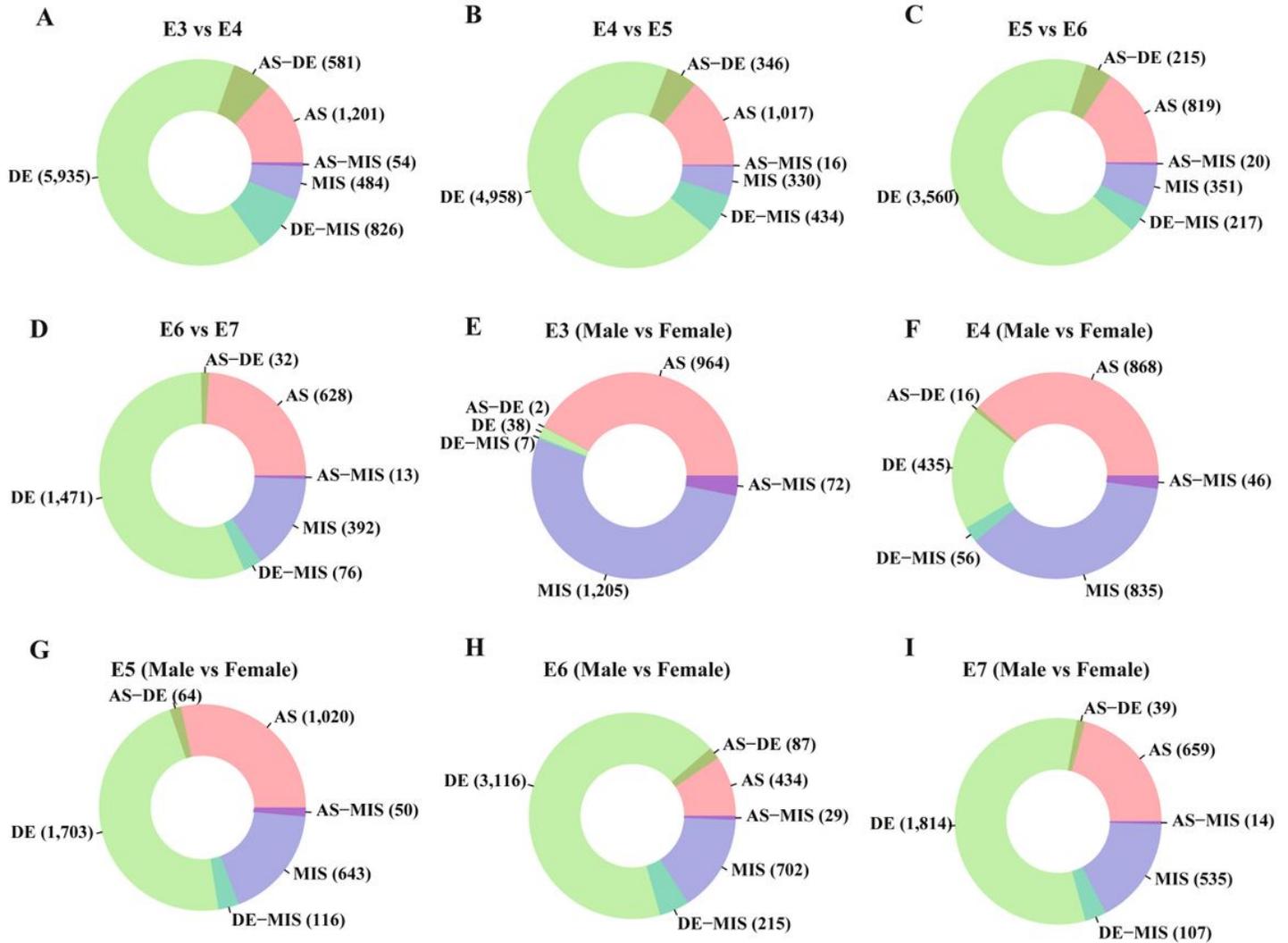
**Figure 1**

Distribution and comparison of DEGs, DASGs and MISGs for adjacent stage comparisons. A. Illustration of differential alternative splicing genes (DASGs) and major isoform switching genes (MISGs). B. Number distribution of DEGs for different stage comparisons. C. Venn graph of DEGs, DASG and MISGs for adjacent stage comparisons. D. Count distribution of DASG for neighbouring stage comparisons. E. Venn graph of DEGs, DASG and MISGs for different stage comparisons. F. Number distribution of MISGs for distinct stage comparisons. G. Venn graph of DEGs, DASGs and MISGs for disparate stage comparisons.



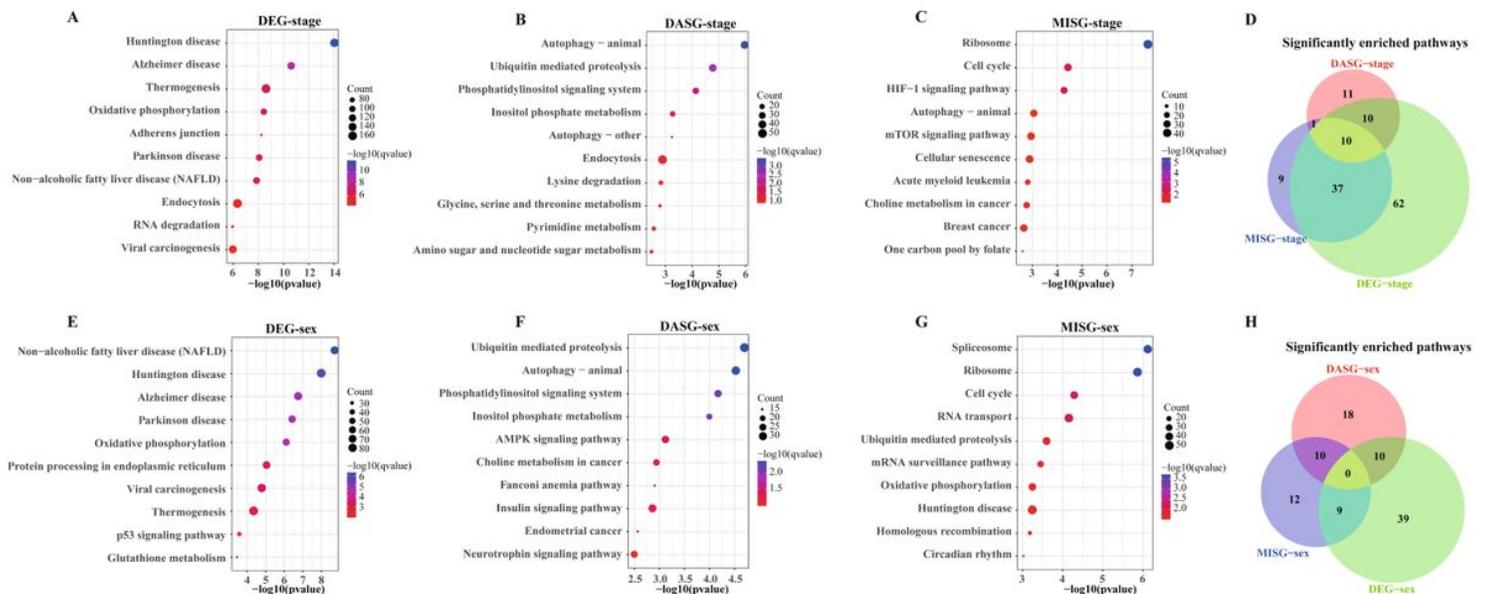
**Figure 2**

Distribution and comparison of DEGs, DASG and MISGs for sex comparisons of different stages. A-C. Number distribution of DEGs, DASGs and MISGs for sex comparisons of distinct stages. D-F. Venn graph of DEGs, DASGs and MISGs for sex comparisons of disparate stages.



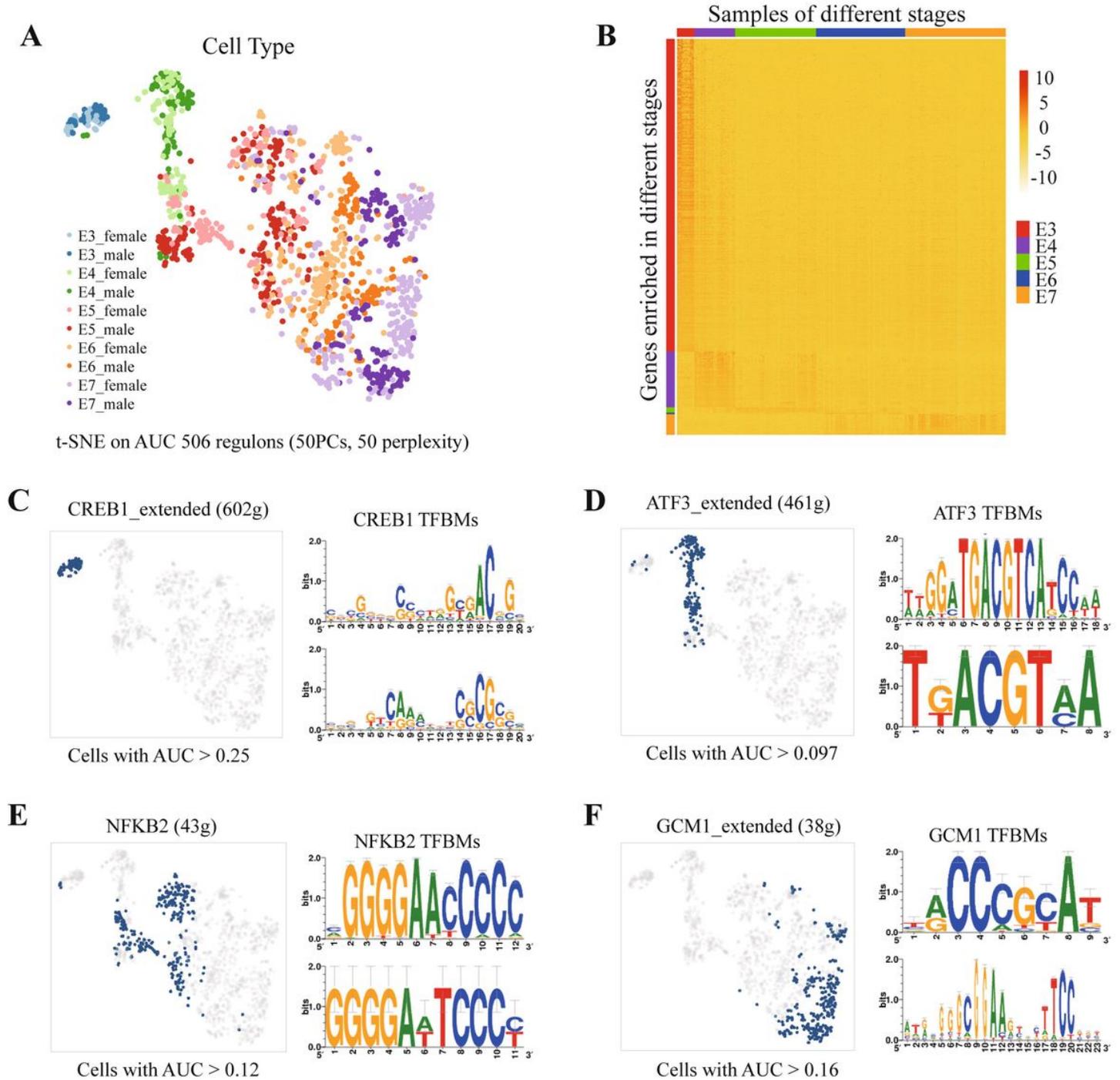
**Figure 3**

Comparison of stage and sex related DEGs, DASGs and MISGs. A-D. Comparison for stage related DEGs, DASGs and MISGs. E-I. Comparison of sex related DEGs, DASGs and MISGs.



## Figure 4

Enriched pathways for DEGs, DASGs and MISGs of stage and sex comparisons. A-C. Significantly enriched pathways for stage related DEGs, DASGs and MISGs. D. Comparison of significantly enriched stage related pathways. E-G. Significantly enriched pathways of sex related DEGs, DASGs and MISGs. H. Comparison of significantly enriched sex related pathways. P-value < 0.05.



## Figure 5

Gene expression regulatory network of early embryos. A. T-SNE for human early embryos of E3 to E7 stages based on 506 regulons. B. Genes with enriched expression for each stage. Fold change > 2 and

adjusted P-value < 0.05. C-F. Examples of stage specific regulons and transcription factor binding motifs for E3, E4, E5 and E7 stages.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.pdf](#)