scientific reports

Check for updates

OPEN

Alternative splicing dynamics during gastrulation in mouse embryo

Wei Wang[1,3], Yu Zhang[1,3], Yuanyuan Zhai[1], Wuritu Yang[2]✉ & Yongqiang Xing[1]✉

Alternative splicing (AS) plays an essential role in development, differentiation and carcinogenesis. However, the mechanisms underlying splicing regulation during mouse embryo gastrulation remain unclear. Based on spatial-temporal transcriptome and epigenome data, we detected the dynamics of AS and revealed its regulatory mechanisms across primary germ layers during mouse gastrulation, spanning developmental stages from E6.5 to E7.5. Subsequently, the dynamic expression of splicing factors (SFs) during gastrulation was characterized, while the expression patterns and functions of germ layer-specific SFs were identified. The results indicate that AS and differential alternative splicing events (DASEs) exhibit dynamic changes and are significantly abundant during the late stage of gastrulation. Similarly, SFs demonstrate stage-specific expression, with elevated levels observed during the middle and late stages of gastrulation. Epigenetic signals associated with SFs and AS sites demonstrate significant enrichment and undergo dynamic changes throughout gastrulation. Overall, this study offers a systematic analysis of AS during mouse gastrulation, identifies primary germ layer-specific AS events, and characterizes the expression patterns of SFs and the associated epigenetic signals. These findings enhance the understanding of the mechanisms underlying the formation of the three germ layers during mammalian gastrulation, with a focus on pre-mRNA AS.

Keywords Mouse gastrulation, Alternative splicing, Splicing factors, Epigenetic regulation, Embryonic development

During mammalian embryonic development, a single totipotent zygote, formed by the fusion of an oocyte and a sperm, undergoes multiple rounds of cell fate commitment to establish the foundation of all future body plans[1–3]. During the preimplantation stage, the inner cell mass (ICM) and the trophectoderm (TE) form and separate, representing the first lineage specification event. Upon implantation, the ICM differentiates into the epiblast (EPI) and primitive endoderm (PrE)[4]. Gastrulation, a critical early postimplantation process, is fundamental to generating the three primary germ layers: endoderm, ectoderm, and mesoderm. During gastrulation, EPI cells differentiate into lineage-specific precursors, establishing the basic body plan of the mammalian embryo[5,6]. Advances in single-cell multi-omics technologies and in *vitro* studies of embryogenesis have enhanced our understanding of the characteristics and mechanisms underlying mammalian gastrulation[7]. The spatially resolved transcriptomes of germ layers during development from pre- to late-gastrulation stages in mouse, human, and cynomolgus monkey have been reported[6,8–10]. In *vitro* embryo culture of cynomolgus monkeys beyond 14 days post-fertilization has been established, advancing the characterization of primate gastrulation[11]. Additionally, epigenetic information is essential for the formation of primary germ layers and subsequent organ development[12]. Collectively, these findings highlight the significance of gene expression and epigenetic regulation in mammalian gastrulation.

Despite significant advancements, the dynamics of pre-mRNA alternative splicing (AS) during mouse gastrulation, particularly in the generation of spatial-temporal germ layers, remains unclear. AS is a critical process in the expression of most eukaryotic genes, involving the removal of introns and the precise joining of exons in different combinations to produce mature mRNA. Splicing reactions are catalyzed by the core spliceosome and auxiliary splicing factors (SFs). Common patterns of AS include: (i) alternative 5′ splice site (A5SS); (ii) alternative 3′ splice site (A3SS); (iii) exon skipping (SE); (iv) intron retention (RI); (v) mutually exclusive exons (MXE); (vi) alternative first exon (AFE); and (vii) alternative last exon (ALE)[13]. Up to 95% of human genes undergo AS, where a single gene generates multiple distinct mature mRNAs, thereby increasing

[1]Inner Mongolia Key Laboratory of Life Health and Bioinformatics, School of Life Science and Technology, Inner Mongolia University of Science and Technology, Baotou, China. [2]Computer Department, Hohhot Vocational College, Hohhot, China. [3]Wei Wang and Yu Zhang contributed equally to this work. ✉email: yang@imu.edu.cn; xingyongqiang1984@163.com

mRNA complexity and proteome diversity[14]. Additionally, AS serves as an essential layer in cell differentiation, proliferation, signal transduction, embryonic and organ development, and carcinogenesis[15–17]. Lu et al.[18] reported that AS extensively occurs in embryonic 7.5 day of mouse primary germ layers and found that numerous SFs are differentially expressed and alternatively spliced across the three germ layer. Aich et al.[19] revealed that TOBF1 maintains mouse embryonic stem cell (ESC) identity through the regulation of AS. The AS profiles and their functions during early embryonic development in human, mouse, and zebrafish were investigated, revealing that embryonic genome activation is associated with splicing disruption[16,20,21]. AS has been characterized during mammalian organ development, while dynamic AS events have been identified across developmental stages[22]. Spliceosomal repression in mouse ESCs drives the pluripotent-to-totipotent state transition[23,24]. Additionally, AS is involved in the sex determination of vertebrates[25]. However, the detailed spatial-temporal patterns and regulatory roles of AS during gastrulation in mammalian embryonic development remain unclear.

Therefore, this study aims to investigate the AS dynamics across the primary germ layers from E6.5 to E7.5 during mouse gastrulation by integrating spatial-temporal transcriptomic and epigenomic data. First, we constructed an AS atlas detailing the development of the extraembryonic, mesodermal, mesendodermal, and ectodermal layers from E6.5 to E7.5. The differential alternative splicing events (DASEs) across different spatial-temporal developmental lineages were identified separately. Next, the dynamic expression of SFs was analyzed, revealing layer-specific expression patterns. Finally, the epigenetic regulation of AS events across different developmental lineages was characterized by examining the enrichment of various histone modifications and DNA methylation. This study offers a systematic AS landscape during mouse gastrulation, identifying AS events specific to each primary germ layer and profiling SF expression patterns and associated epigenetic modifications. These findings could enhance the understanding of the mechanisms underlying ectoderm, mesoderm, and endoderm formation during mammalian gastrulation.

Materials and methods

Dataset construction

Spatial-temporal transcriptome and epigenome datasets across primary germ layers from E6.5 to E7.5 days during mouse gastrulation were downloaded from NCBI GEO under accession numbers GSE98101 and GSE104243, deposited by Jing's group[12,26]. Overall, 18 transcriptome samples from the primary germ layers during this period, along with corresponding epigenome samples (including H3K4me1, H3K4me3, H3K27ac and DNA methylation), were curated (Supplementary Table S1).

Collection of splicing factors

First, we manually curated 509 and 455 SFs from human and mouse, respectively, by reviewing relevant literature. Subsequently, homology analysis between human and mouse SFs was conducted. Consequently, 598 SFs in mousee, including several SR proteins and hnRNP proteins, were collected. Supplementary Table 2 presents detailed information about the mouse SFs. The interaction network of SFs during gastrulation was constructed using STRING (https://string-db.org/) with a confidence score threshold set to >0.7.

RNA-seq data processing and analysis

Quality control of raw RNA-seq data was performed using FastQC (v0.11.8)[27] and Trimmomatic (v0.38)[28]. The GRCm38 genome version was used in this study. Transcript and gene expression quantification were conducted using Salmon (v0.12.0)[29], with expression values reported as TPM (transcripts per million). Principal component analysis (PCA) was used to visualize the similarities between the 18 samples. Additionally, differentially expressed genes (DEGs) were analyzed using the R package DESeq2 (v1.30.1)[30]. DEGs were considered statistically significant if $|\log2FC| \geq 1$ and the p-value was <0.05 [31–33].

Analysis of alternative splicing events

SUPPA2 (v2.3) was used to identify AS events from RNA-seq data[34]. The mouse annotation file (Mus_musculus.GRCm38.102.chr.gtf) was input into the generateEvents function of SUPPA2 to generate AS events. Subsequently, the TPM values of transcripts were used as input for the psiPerEvent function to calculate the inclusion level (percentage spliced inclusion, PSI) of each AS event. Furthermore, the diffSplice function was used to calculate the PSI change (ΔPSI) and p-value across different spatial-temporal developmental stages based on the TPM values of transcripts. The threshold for determining DASEs in the contrast group was set as follows: (1) a splicing change $|\Delta PSI| > 0.1$ between two spatial-temporal developmental stages, (2) a significant difference in ΔPSI with a p-value <0.05. The dynamics of DASEs across spatial-temporal developmental stages of the mouse gastrula were visualized using the R package ggalluvial (v0.12.4)[35]. Additionally, the sequence characteristics of AS events during mouse gastrulation, including the analysis of splicing site signals, were analyzed using ASTK (https://huang-sh.github.io/astk-doc/)[36].

Identification of alternative splicing switching events

Spliced isoform switch events occur when the relative abundance of two alternatively spliced isoforms are reversed under different conditions. The TSIS tool was used to analyze spliced isoform switches during mouse gastrulation[37]. Spliced isoform switches were identified based on the following criteria: (1) a switch probability >0.5, (2) a sum of the average difference between the two isoforms >1, (3) a p-value <0.001 indicating a significant difference before and after the switch, and (4) a minimum interval duration of ≥ 2-time points.

Epigenetic data processing and analysis

ChIP-Seq data analysis

Quality control of raw FASTQ data was performed using FastQC (v0.11.8) and Trimmomatic (v0.38). Cleaned FASTQ datasets were aligned to the GRCm38 reference genome using Bowtie2 (v2.1.0). SAM files were converted to sorted BAM files, and PCR duplicates were removed using Samtools (v1.10)[38]. ChIP-seq signal quantification was performed through peak calling with MACS2 (v2.1.4)[39]. All intersection calculations were performed using the 'intersect' module of Bedtools (v2.27.1)[40]. The scale-regions function in the 'computeMatrix' module of deepTools (v3.5.0) was used to extract scores for regions of interest into a matrix format[41]. Profiles and heatmaps were visualized using the 'plotProfile' and 'plotHeatmap' functions of deepTools. Finally, peak positions in the genome were annotated using the R packages ChIPseeker and ChIPpeakAnno[42,43].

DNA methylation data analysis

Quality control of DNA methylation data was performed using FastQC (v0.11.8) and Trimmomatic (v0.38). Cleaned FASTQ datasets were aligned to the GRCm38 genome using Bowtie2 (v2.1.0). Subsequently, the clean data were further aligned to the GRCm38 reference genome using Bismark[44]. BAM files from the two replicates ($n = 2$) were merged. The 'computeMatrix' module of deepTools (v3.5.0) was used to extract the scores for regions of interest in matrix format. Subsequently, profiles were visualized using the 'plotProfile' function. Differential methylation analysis, including the identification of differentially methylated genes (DMGs) and differentially methylated regions (DMRs), was conducted using the methylKit R package. Furthermore, the getMethylDiff function was used to extract the results of differential analysis.

Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses, along with the visualization of DEGs, SFs, and genes undergoing differential alternative splicing (DAS), were conducted using the ClusterProfiler (v4.1.4) package in R[45]. Mouse annotation data were obtained from Bioconductor (org.Mm.eg.db, v3.12.0). Statistical significance thresholds for all GO and KEGG enrichment analyses were set at $p.adjust < 0.05$.

Results

Construction and analysis of the alternative splicing atlas during mouse gastrulation

AS events during gastrulation were identified using SUPPA2 based on the annotation file. Figure 1A demonstrates that the frequency of AS events varies dynamically across different lineages during mouse embryonic development. Supplementary Table 3 presents detailed information on the frequency of AS events. Most AS events are SE and AFE, while MXE and ALE represent the smallest proportions among the seven AS types. The total AS frequency reaches its peak at the E7.5_A and E7.5_AM among the nine regional samples across three developmental stages. This indicated that the transcriptomes of the ectodermal lineage (E6.5_Epi, E7.0_A, E7.5_A), mesodermal lineage (E6.5_Epi, E7.0_P, E7.5_AM), and mesendodermal lineage (E6.5_Epi, E7.0_P, E7.5_P) exhibited greater diversity. In contrast, the extraembryonic lineage (E6.5_ExE, E7.0_ExE, E7.5_ExE), particularly at the E7.0_ExE stage, demonstrated the fewest AS events and the least transcriptomic complexity. Additionally, the frequency of AS events gradually increases within the same developmental lineage as development progresses. This may correlate with the differentiation potential of the various developmental lineages. These findings suggested that AS may play a lineage-specific role during mouse gastrulation. Compared to the extraembryonic lineage, the formation of the three germ layers potentially involves more protein interactions, particularly in the development of the ectoderm and mesoderm.

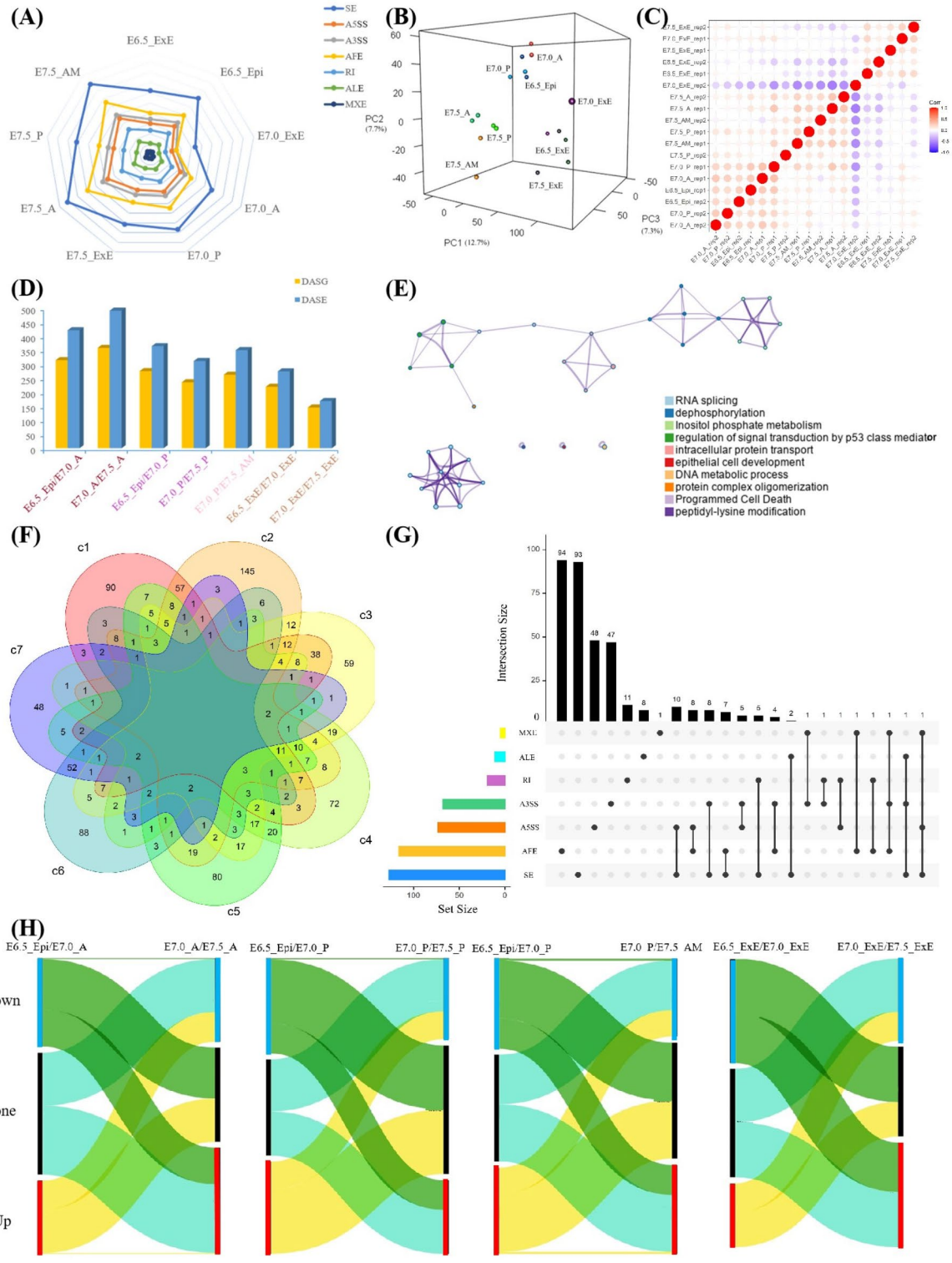
Conventionally, sample heterogeneity across developmental stages has been assessed primarily through gene expression levels. However, given the specific distribution of AS frequencies across different developmental lineages, we aim to conduct this analysis based on AS events. The results showed that using the PSI value, representing the inclusion level of AS events as input features for PCA effectively cluster samples of the nine regional samples. Samples from extraembryonic and embryonic lineages are distinctly separated (Fig. 1B). Furthermore, the temporal characteristics of samples during the formation of the three germ layers are clearly evident. These findings suggested that AS may play a significant regulatory role, and the PSI value of AS event can serve as a reliable indicator for sample clustering and the inference of developmental trajectories during mammalian gastrulation.

Additionally, we calculated the sample correlation coefficient based on the PSI values of AS events (Fig. 1C). The results indicated that the sample correlation coefficient is significantly higher within the same developmental stage or lineage. The extraembryonic samples are distinctly separated from the intraembryonic samples. The correlation coefficient between the E7.0_ExE sample and the other 17 samples is anomalous, potentially due to the low quality of the sample. Overall, these results indicated that the quality of the 18 samples used in this study is reliable, making them suitable for downstream analyses of gene expression and AS in germ layer lineage differentiation.

Analysis of differential alternative splicing events during mouse gastrulation

Identification and functional analysis of differential alternative splicing events

Using the diffsplice module of SUPPA2, DASEs and differentially alternatively spliced genes (DASGs) between consecutive developmental stages in germ cell specification were identified (Fig. 1D, Supplementary Table S4-S5). The results indicated that the number of DASEs and DASGs is highest between E7.0_A and E7.5_A while lowest between E7.0_ExE and E7.5_ExE. Compared to the extraembryonic lineage, the embryonic lineages (ectodermal, mesodermal, and mesendodermal) exhibit a greater abundance of DASEs and DASGs, consistent with the observed distribution of splicing events. This indicated that the development of the nervous system



and sensory organs may rely more heavily on DASEs. Enrichment analysis was performed on DASGs between the E7.0_A and E7.5_A stages (Fig. 1E). Biological processes such as RNA splicing, dephosphorylation, inositol phosphate metabolism, regulation of signal transduction by p53, and protein complex oligomerization are significantly enriched (Supplementary Table S6). These findings indicated that splicing regulation may play a crucial role in the development of the nervous system and sensory organs from the ectodermal lineage. Additionally, no gene exhibits DAS across all consecutive stages of the three developmental lineages (Fig. 1F). A higher number of overlapping DASGs was observed between consecutive stages within the same developmental lineage, while a lower number of overlapping DASGs was observed between embryonic and extraembryonic developmental lineages. These findings indicated that different developmental lineages may require different sets of DASGs during their development.

◀**Fig. 1**. Global analysis of AS during mouse gastrulation. (**A**) Radar plot showing the frequency of splicing events in different lineages of mouse gastrulation. E6.5_ExE: E6.5_extraembryonic, E6.5_Epi: E6.5_epiblast, E7.0_ExE: E7.0_extraembryonic, E7.0_A: E7.0_anterior epiblast, E7.0_P: E7.0_posterior epiblast, E7.5_ExE: E7.5_extraembryonic, E7.5_A: E7.5_anterior epiblast, E7.5_AM: anterior mesoderm part, E7.5_P: E7.5_posterior. (**B**) PCA plot denoted by different colors during mouse gastrulation employing PSI value of AS events as input features of nine regional samples across three developmental stages. (**C**) Heatmap showing the sample correlation based on AS events of different lineages during mouse gastrulation. The red and purple respectively represent positive and negative correlation. The characterization of DASEs in mouse gastrulation. (**D**–**H**) The characterization of DASEs in mouse gastrulation. (**D**) Bar chart showing the number of DASEs and DASGs between consecutive developmental lineages. The different colors of x-axis label represent different developmental lineages. (**E**) Enrichment analysis plot of DASGs between E7.0_A and E7.5_A. The size of the dots represents the number of genes, and the thickness of the lines represents the strength of the interactions between genes. (**F**) Venn diagram showing the overlap of DASG between different stages of three developmental lineages. c1-c7 corresponds to 7 consecutive stages shown in Fig. 1d. (**G**) Upset plot of interaction sets among all AS event patterns between E7.0_A and E7.5_A. (**H**) Sankey diagram of DASEs in ectodermal lineage (E6.5_Epi, E7.0_A, E7.5_A), mesendodermal lineage (E6.5_Epi, E7.0_P, E7.5_P), mesodermal lineage (E6.5_Epi, E7.0_P, E7.5_AM) and extraembryonic lineage during mouse gastrulation. Up (red color) denotes significantly up-regulated DASEs (dPSI > 0.1 and $p < 0.05$); Down (blue color) denotes significantly down-regulated DASEs (dPSI < -0.1 and $p < 0.05$); None (black color) denotes DASEs without significant difference (|dPSI| <0.1 or $p > 0.05$).

An upset plot was presented to illustrate the distribution patterns of DASEs between E7.0_A and E7.5_A, along with their intersections across multiple dimensions (Fig. 1G). Among these, AFE and SE events are the most abundant. Most genes exhibit only one type of AS, while a few genes demonstrate up to three distinct DASEs occurring within a single gene. The different combinations of these DASEs may contribute to the diversity in gene expression regulation and transcriptomic variation during embryonic layer differentiation.

Dynamic distribution of differential splicing events

To better visualize the dynamic changes of DASEs across the four developmental lineages during mouse gastrulation, the dPSI and *p*-value of DAS were used as input for the Sankey diagram, which characterizes the changes in splicing patterns (Fig. 1H). Most DASEs exhibit dynamic changes in embryonic and extraembryonic developmental lineages. For example, significantly downregulated DASEs (dPSI < -0.1 and $p < 0.05$) transition into significantly upregulated DASEs and non-DASEs between the E7.0_A and E7.5_A stages. Furthermore, significantly upregulated DASEs (dPSI > 0.1 and $p < 0.05$) transition into significantly downregulated DASEs and non-DASEs between the E7.0_A and E7.5_A stages. Non-DASEs (|dPSI| < 0.1 or $p \geq 0.05$) transition into significantly upregulated and downregulated DASEs between the E7.0_A and E7.5_A stages. This finding may indicate that DASEs are stage-specific during lineage differentiation. However, the extent of transformation between DAS types varies across the four developmental lineages, particularly in the extraembryonic developmental lineages. Furthermore, only a few DASEs exhibit a consistent dPSI trend across different developmental stages. For example, the dPSI variation trends of one downregulated and one upregulated DASEs in the ectodermal lineage remain unchanged (Fig. 1H). The genes corresponding to these DASEs are Pdp1 and Amd1. Pdp1 may enhance metal ion binding and phosphoprotein phosphatase activities. Pdp1 may play a role in positively regulating serine-threonine dephosphorylation and pyruvate dehydrogenase activity. Amd1 may promote adenosine methionine decarboxylase activity, lyase activity, and protein binding.

Expression profile of splicing factors and their regulatory function during mouse gastrulation

The precise removal of introns and ligation of exons is coordinated by cis-acting RNA sequence elements, trans-acting SFs, and epigenetic modifications, including chromatin structure, DNA methylation, histone modifications, and RNA modifications. Additionally, the elongation rate of RNA polymerase directly affects exon skipping. Furthermore, SFs play crucial regulatory roles in disease onset, tissue differentiation, and development. Dysfunction of SFs may lead to abnormal cell differentiation and apoptosis. In this study, the expression dynamics of SFs and their regulatory functions during mouse gastrulation were investigated.

Tissue-specific expression of splicing factors

DEG analysis revealed lineage-specific gene expression patterns[26]. For example, we identified 3,669 DEGs in the ectodermal lineage (E6.5_Epi, E7.0_A, E7.5_A) with |log2(Fold change)| ≥ 1 and *p*-value < 0.05. The number of DEGs changes dynamically across the three developmental stages, peaking between E7.0_A and E7.5_A (Supplementary Figure S1). Furthermore, the number of highly expressed genes at the E7.5_A stage is 436, which is significantly higher than that at the E6.5_Epi and E7.0_A stages of ectodermal lineage. The expression profile of 119, 73, and 436 highly expressed genes at the E6.5_Epi, E7.0_A, and E7.5_A stages can be observed, respectively (Fig. 2A).

These results indicated that the transcriptome is more diverse and functionally complex at the E7.5_A stage than at the E6.5_Epi and E7.0_A germ layers during mouse gastrulation. However, the expression patterns of SFs in this context remain unclear. To address this, we constructed a dataset of 598 mouse SFs, including the most common SR and hnRNP proteins, based on literature mining and homology analysis. The intersection between highly expressed genes in the ectodermal lineage and 598 SFs was analyzed (Fig. 2B). The results showed that Rbpms2 is a highly expressed SF in the E6.5_Epi germ layer, Tial1 and Zc3hav1 are highly expressed SFs in the

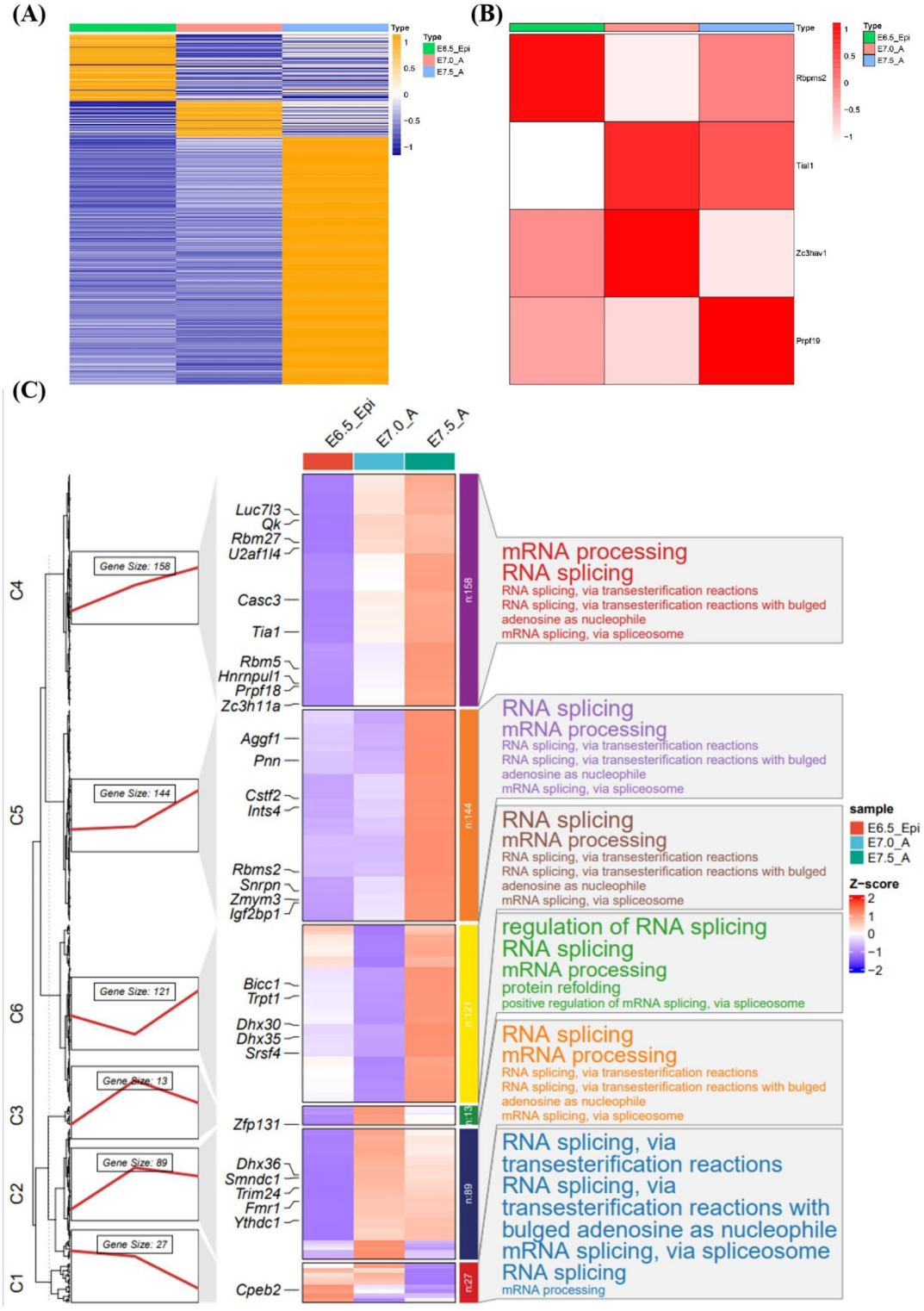


Fig. 2. Expression profiles of highly expressed genes and specific SFs in different development stages of mouse ectodermal lineage. (**A**) Heatmap of highly expressed genes in E6.5_Epi, E7.0_A and E7.5_A germ layer. (**B**) Heatmap of overlapped genes between SFs and highly expressed genes. (**C**) Expression patterns of SFs in ectodermal lineage of mouse gastrulation. Different color bars on the top represent different developmental stages. The color from red to blue of Z-score represents the expression level from high to low. The number in the left rectangular box indicates the number of genes corresponding to the cluster. The right rectangular box shows the top five enriched terms for corresponding cluster.

E7.0_A germ layer, and Prpf19 is a highly expressed SF in the E7.5_A germ layer. A previous study demonstrated that the RNA-binding protein RBPMS2 regulates gastrointestinal smooth muscle development and is essential for controlling smooth muscle plasticity[46]. Additionally, RBPMS2 is a cardiac-enriched splicing regulator critical for heart function[47]. TIAL1 tightly regulates RNA splicing and plays a crucial role in DNA damage sensing and repair during B cell development[48]. This finding suggested that Zc3hav1 and Prpf19 may mediate AS based on its high expression, thereby regulating ectodermal lineage development.

K-means clustering was used to analyze the specificity of SF expression in the ectodermal lineage (E6.5_Epi, E7.0_A, E7.5_A) (Fig. 2C). The clustering performance and interpretability of k-means are optimal at $k=6$. The results showed that SFs are clustered into six categories: 27 SFs in category 1, 89 SFs in category 2, 13 SFs in category 3, 121 SFs in category 4, 144 SFs in category 5, and 158 SFs in category 6. The expression levels of SFs in category 4 gradually increase from E7.0_A to E7.5_A. SFs in category 5 are highly expressed at E7.5_A, while SFs in category 6 exhibit high expression at both E6.5_Epi and E7.5_A. SFs in categories 2 and 3 are most highly expressed at E7.0_A. The SFs in category 1 are highly expressed at E6.5_Epi and E7.0. The expression of SFs is stage-specific, with most SFs showing a preference for expression during the mid to late stages of ectodermal lineage, particularly at E7.5_A. GO enrichment analysis revealed that the top five terms for each cluster are closely related to biological processes such as spliceosome assembly, mRNA processing, and mRNA splicing. Furthermore, we analyzed the expression profiles of the top 30 highly expressed SFs, with 24 of them exhibiting high expression at E7.5_A, 5 at E7.0_A, and only 1 at E6.5_Epi (Supplementary Figure S2). The high expression of SFs indicated that splicing may mediate mouse gastrulation, particularly during the late stage. Additionally, the specificity of SF expression in other developmental lineages was analyzed using k-means clustering (Supplementary Figure S3, S4, and S5).

Overall, the expression patterns of SFs exhibit lineage specificity and are closely related to key biological processes, such as spliceosome assembly and mRNA processing. Understanding these SF expression patterns and their regulatory mechanisms offers valuable insights into the complex processes of lineage differentiation and development.

Function analysis of splicing factors

To investigate the role of SFs in lineage development, 27 overlapping genes, identified between 598 SFs and 359 DASGs at E7.0_A and E7.5_A, were selected. This finding indicated that these SFs may contribute to lineage development through DAS. The 27 differentially spliced SFs were input into the STRING database to predict their interaction network. A subnetwork of SFs can be observed in central of Fig. 3A. It was indicated SFs Hnrnpa2b1, Hnrnph1, Hnrnph2, Hnrnpu, Syncrip, Sf3b1, Eftud2, and Ddx17 located in subnetwork are associated with ectodermal lineage development, with the hnRNP family having particularly significant association. Functional enrichment analysis of these 27 SFs revealed that they are closely related to biological processes, such as RNA splicing and spliceosome complex formation (Fig. 3B).

Most studies on germ layer differentiation have focused on gene expression regulation. However, differences in isoform abundance caused by AS cannot be detected at the gene-level measurements despite AS playing critical roles in embryogenesis. Therefore, we investigated isoform switch events using the TSIS program, which identifies instances where gene expression remains constant; however, the isoform abundances fluctuate during mouse gastrulation[37]. Ultimately, 212 significant isoform switch events ($p<0.001$) involving two transcript isoforms were identified during mouse gastrulation across 22 differentially alternatively SFs. Among the crucial SFs in the interaction network of these different SFs (Fig. 3A), Eftud2 and Ddx17 exhibit significant isoform switch events ($p<0.001$) (Fig. 3C and D).

Eftud2, known as 116 kDa, Snrp116, or U5-116kD, is essential for mRNA splicing regulation and participates in spliceosome assembly. It generates 13 different transcriptional isoforms with the following translation profiles: Eftud2-201 encodes a protein of 972 amino acids, Eftud2-202 encodes a protein of 971 amino acids, Eftud2-205 encodes a protein of 85 amino acids, Eftud2-206 encodes a protein of 30 amino acids, Eftud2-211 encodes a protein of 98 amino acids, and Eftud2-212 encodes a protein of 962 amino acids. In contrast, protein products of the Eftud2 isoforms Eftud2-203, Eftud2-204, Eftud2-207, Eftud2-208, Eftud2-209, Eftud2-210, and Eftud2-213 have not been detected. Studies have shown that Eftud2 is expressed throughout embryonic development, with its expression predominantly concentrated in the developing head and craniofacial regions. Eftud2 exhibits significant splicing switches across multiple developmental stages, including E6.5_Epi versus E6.5_ExE, E7.0_ExE versus E7.0_P, E7.5_A versus E7.5_ExE, and E7.5_P versus E7.5_ExE (Fig. 3C). These findings suggested that the SF translated from Eftud2-211 may play a role in extraembryonic lineage development, while Eftud2-202 may contribute to ectodermal lineage development.

Ddx17, known as p72, Gm926, 2610007K22Rik, and A430025E01Rik, is a SF characterized by the conserved Asp-Glu-Ala-Asp (DEAD) motif. DEAD-box RNA helicases play essential roles in various cellular processes involving changes in RNA secondary structure, such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Ddx17 produces four different transcriptional isoforms: Ddx17-201 translates into a protein comprising 652 residues, Ddx17-202 translates into a protein comprising 407 residues, Ddx17-203 translates into a protein comprising 650 residues, and Ddx17-204 translates into a protein comprising 418 residues. Ddx17 exhibits significant splicing switches between E7.5_A, E7.5_AM, and E7.5_ExE. Ddx17-201 and Ddx17-203 are the dominant transcriptional isoforms. Furthermore, Ddx17-201 is correlated with ectodermal lineage development at the E7.5_A stage, while Ddx17-203 is correlated with developmental processes across multiple stages, including E6.5_ExE, E7.0_A, E7.0_ExE, E7.0_P, and E7.5_AM, during embryonic lineage development (Fig. 3D).

Analysis of isoform switch events revealed the potential regulation of mouse gastrulation development by different isoforms of SFs, such as Eftud2 and Ddx17. These findings highlight that developmentally specific SFs are associated with lineage development processes.

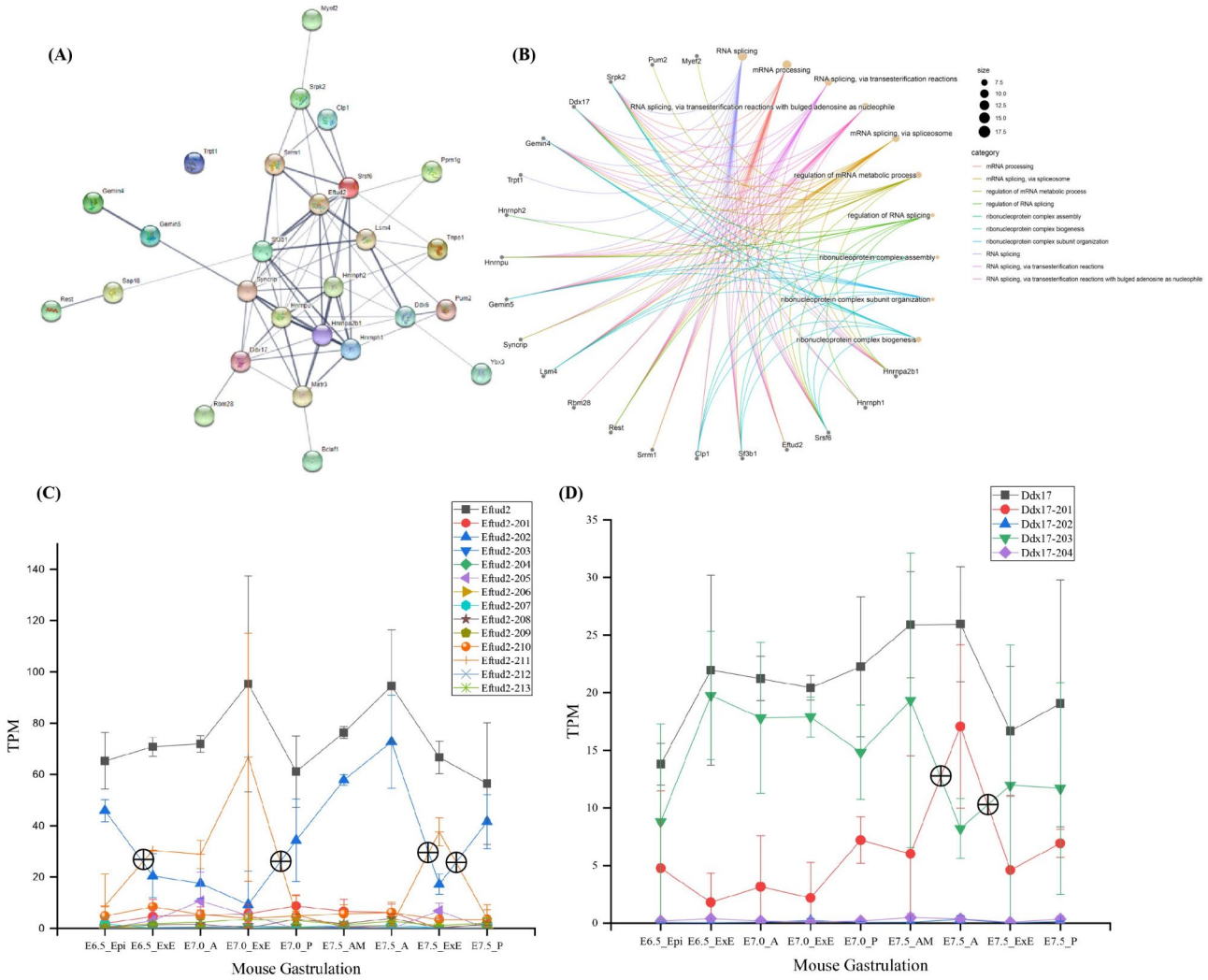


Fig. 3. Interaction network and enrichment analysis of 27 splicing factors which are differentially spliced between E7.0_A and E7.5_A. (**A**) The interaction network of splicing factors during mouse gastrulation. The thickness of the lines represents the strength of the interaction relationship. (**B**) Enrichment analysis of splicing factors. Different colors represent different biological processes. The size of the dots in each biological process represents the number of enriched genes. (**C**) Isoforms switch events identified in differentially alternatively spliced factors between E7.0_A and E7.5_A in ectodermal lineage. The y-axis represents the TPM of transcript isoforms, ⊕indicating the switching point. The black line represents the gene expression level, while the other colored lines represent the transcript expression level. Expression profile of the splicing factor Eftud2 in ectodermal lineage. (**D**) Expression profile of the splicing factor Ddx17 in ectodermal lineage.

Crosstalk between epigenetic events and alternative splicing during mouse gastrulation

AS of pre-mRNA represents a key mechanism for generating protein diversity. However, its regulatory processes remain incompletely understood. Epigenetic factors, such as DNA methylation and histone modifications, contribute to AS and are closely related to mammalian embryonic development. In this study, we systematically analyzed the epigenetic characteristics associated with AS during mouse gastrulation.

Identification and analysis of differential methylation regions adjacent to splicing factors

DNA methylation is crucial for various biological processes[49–51], such as embryonic development, cell fate determination, and the maintenance of genome stability. Increasing evidence suggests that DNA methylation regulates AS[52]. Specifically, DNA methylation regulates AS through various mechanisms, such as recruiting SFs, changing the chromatin structure, and modulating RNA polymerase II elongation rate. However, the precise regulatory mechanisms of DNA methylation in AS remain unclear.

In this study, the differential DNA methylation regions between the E7.0_A and E7.5_A germ layer of ectodermal lineage were identified. After annotation, The genes near SF-encoding genes with DNA methylation regions were obtained, and the biological functions were analyzed. The results showed that genes located in highly methylated regions adjacent to SFs are primarily involved in the regulation of adipocyte differentiation. Specific genes include Smad3, Metml, Zfpm1, Tmem64, Cebpb, and Ncor2 (Fig. 4A). In contrast, genes located

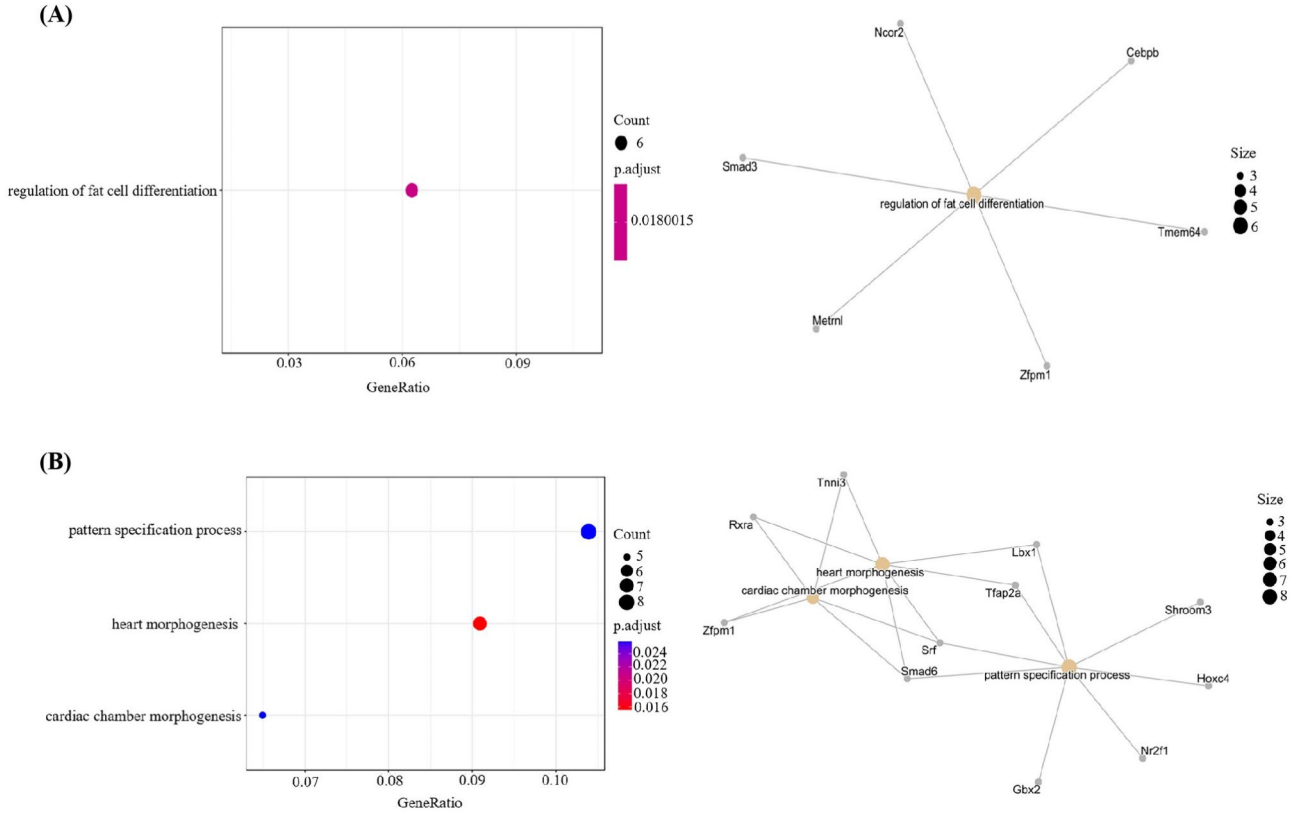


Fig. 4. Enrichment analysis of genes enriched in high methylated regions (**A**) and low methylated regions (**B**) between E7.0_A and E7.5_A during mouse gastrulation. The *x*-axis shows the GeneRatio, which represents the ratio of enriched genes to the total input genes, and the y-axis shows the description of enriched biological process terms. The size of a dot represents the number of genes associated with the GO term, and the color of dots represents the p-adjusted value.

in regions with low DNA methylation near SFs are primarily associated with heart morphogenesis, pattern specification, and cardiac chamber morphogenesis. These three biological processes are interconnected through the key genes Srf and Smad6 (Fig. 4B). These findings suggested that AS may mediate lineage development and differentiation through differential methylation near SFs.

Analysis of histone modifications in alternative splicing events during lineage differentiation

Histone modifications can influence AS outcomes by recruiting SFs through chromatin-binding proteins. However, the colocalization of AS and histone modifications during embryonic development remains unclear. In this study, we offered a systematic analysis of the distribution of histone modifications during mouse gastrulation. The binding characteristics of H3K4me1 in SE events during ectodermal lineage differentiation were exemplified (Fig. 5A). Overall, 1,860 shared SE events were identified across E6.5_Epi, E7.0_A, and E7.5_A. The H3K4me1 signal profiles exhibit consistent patterns, with significant changes observed at both the transcription start sites and end sites of genes associated with the SE events (Fig. 5B). E6.5_Epi, E7.0_A, and E7.5_A each exhibit 227, 295, and 320 specific SE events, respectively. The H3K4me1 signal profiles demonstrate significant specificity (Fig. 5C). At the E6.5_Epi stage, the H3K4me1 signal for specific SE events demonstrates an initial increase, followed by a stable signal across the gene body. At the E7.0_A stage, the H3K4me1 signal for specific SE events exhibits a high signal pattern toward the 3' end of the gene body. The H3K4me1 signal for specific SE events at the E7.5_A stage exhibits a more uniform pattern at the center of the gene body. In summary, the splicing event profiles of shared and specific epigenetic modification patterns reveal significant differences, indicating differential regulation of histone modifications and AS across different lineages.

To confirm whether the genes annotated with co-occurring peaks overlap with SFs, we performed an overlap analysis between the 1,860 genes annotated by H3K4me1 peaks and 598 SFs. This analysis revealed 38 overlapping genes (Fig. 5D). Enrichment analysis revealed that these genes are predominantly associated with biological processes related to AS regulation, mRNA processing, mRNA metabolism, and the spliceosome pathway (Fig. 5E, F).

Epigenetic modification signal patterns near splice sites

Precise identification of splice sites is a crucial step in RNA splicing. The epigenetic information surrounding splice sites plays a crucial regulatory role in their recognition and the binding of SFs. In this study, we used the

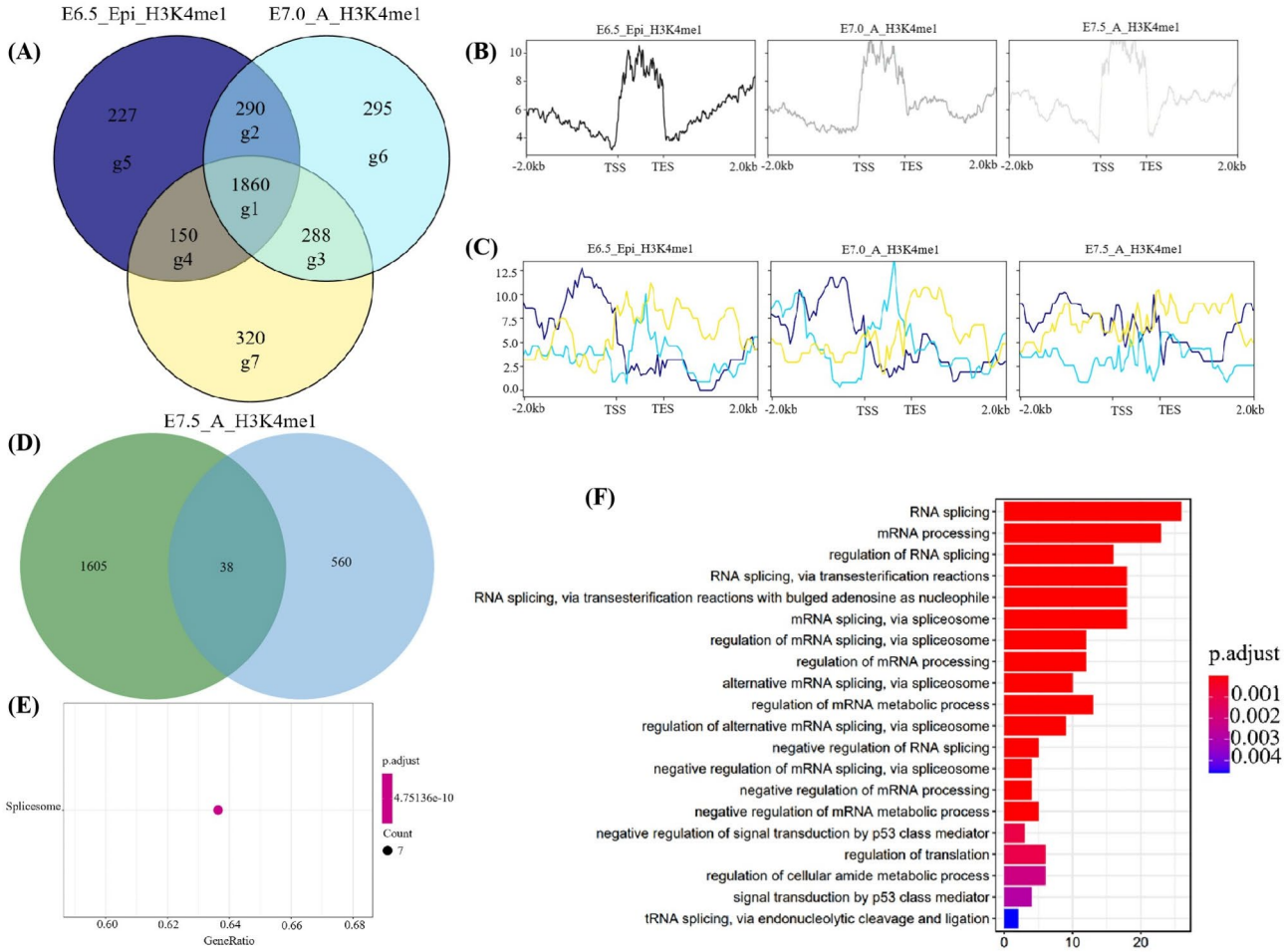


Fig. 5. H3K4me1 profiles in SE events during ectodermal lineage. (**A**)Venn diagram of SE event sites with H3K4me1 modifications in E6.5_Epi, E7.0_A and E7.5_A of ectodermal lineage. The group g1 denotes the SE event sites with same H3K4me1 modification profiles in E6.5_Epi, E7.0_A and E7.5_A. The groups g5, g6 and g7 denote SE event sites with specific H3K4me1 modification profile in E6.5_Epi, E7.0_A and E7.5_A, respectively. The groups g2, g3 and g4 denote SE event sites with same H3K4me1 modification profiles between E6.5_Epi and E7.0_A, E7.0_A and E7.5_A, E6.5_Epi and E7.5_A, respectively. (**B**) H3K4me1 profile of 1,860 shared SE events in E6.5_Epi, E7.0_A and E7.5_A. (**C**) The H3K4me1 modification profile of specific SE events in E6.5_Epi, E7.0_A, and E7.5_A. The dark blue, light blue and yellow respectively denote H3K4me1 signal of 227, 295 and 320 specific SE events in E6.5_Epi, E7.0_A and E7.5_A, respectively. (**D**) The overlap between the genes annotated by 1860 H3K4me1 peaks and 598 SFs. **(E–F)** GO and KEGG enrichment analysis of 38 overlapping SFs[53].

ASTK tool to analyze the distribution of epigenetic marks near the SE events during the E6.5_Epi stage of mouse gastrulation. This analysis included H3K4me1, H3K4me3, H3K27ac, and DNA methylation.

The results showed that histone modification signals are significantly correlated with the inclusion levels of SE events during the E6.5_Epi germ layer stage. The H3K4me1 signal is significantly enriched in the A1 region of SE events with high inclusion levels and in the A4 region of SE events with low inclusion levels (Fig. 6A). The H3K4me3 signal exhibits significant enrichment in the A1 and A4 regions of SE events with high inclusion levels and in the A1 region of SE events with low inclusion levels (Fig. 6B). The H3K27ac signal is enriched in all A1 regions of SE events, regardless of inclusion level (Fig. 6C). Histone modification signals are significantly higher for SE events with low inclusion levels. Additionally, we observed an association between the inclusion levels of SE events and the distribution of DNA methylation near splice sites (Fig. 6D). These findings suggested that epigenetic information directly may regulate mammalian embryonic development and mediate AS events during this process.

Discussion

Mouse gastrulation is a critical stage in embryonic development, involving complex biological processes such as cell migration, differentiation, and organ formation. Recent advancements in omics technologies have offered novel insights into the molecular mechanisms regulating this process. AS, a key form of post-transcriptional regulation, plays a vital role in gastrulation. However, research on AS during the lineage differentiation in mouse

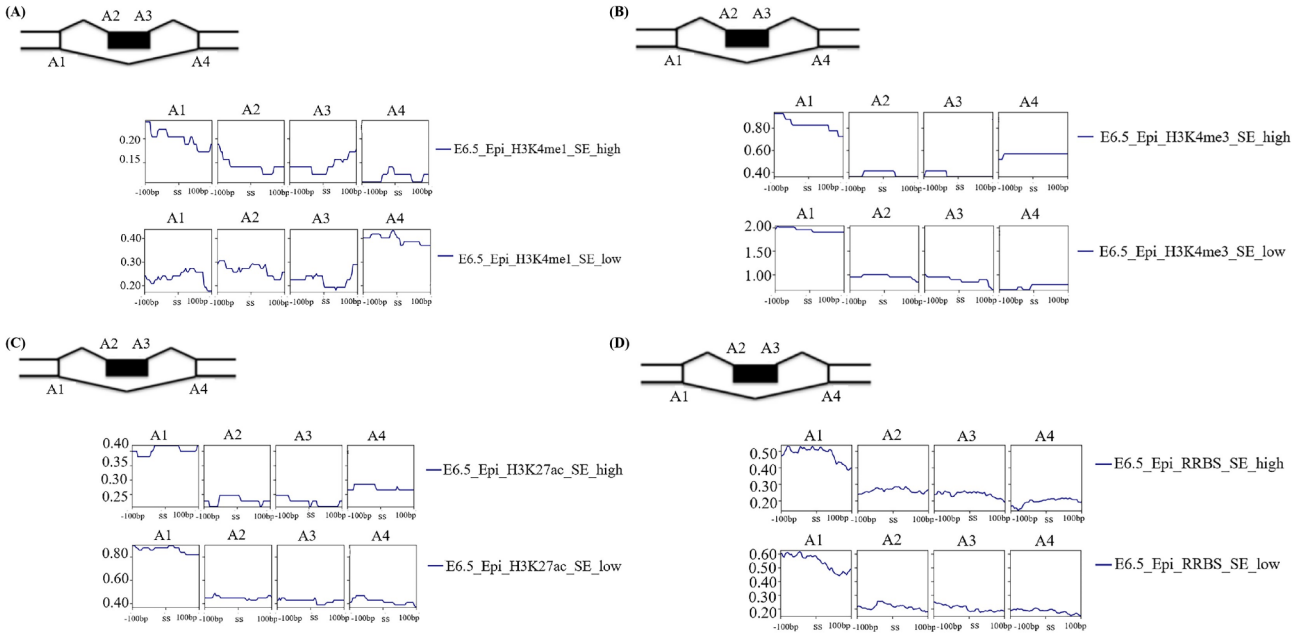


Fig. 6. The distribution of epigenetic modification signal near the splice site of SE events in E6.5_Epi stage of mouse gastrulation. (**A–D**) represents the distribution of H3K4me1, H3K4me3, H3K27ac and DNA methylation, respectively. The y-axis represents the normalized signal intensity; The "ss" denotes splice site; SE high denotes 0.8< PSI<1 of SE event; SE low denotes 0< PSI<0.2 of SE event; A1 denotes upstream and downstream 100 bp regions of the 3' ss of the upstream intron. A2 denotes upstream and downstream 100 bp regions of the 5' ss of the skipping exon. A3 denotes upstream and downstream 100 bp regions of the 3' ss of the skipping exon. A4 denotes upstream and downstream 100 bp regions of the 5' ss of the downstream intron.

gastrulation remains limited. This study demonstrated an AS map for mouse gastrulation lineage development to investigate the role of AS at various lineage developmental stages. It explored the dynamic expression, function, and epigenetic characteristics of SFs during lineage development. The results highlighted the dynamic changes in AS throughout mouse gastrulation, offering insights into its regulatory mechanisms.

Despite the quality of one of the samples from E7.0_ExE is slightly suboptimal, it does not affect the accuracy of AS identification and quantification. Our results showed that the frequency and types of AS events are dynamic across different stages of mouse gastrulation. We employed the PSI as a key metric for AS events and performed PCA. The findings are consistent with those derived from gene expression data. This finding indicated that the inclusion level of AS events is a novel parameter for PCA analysis, which can be used for clustering and inferring developmental trajectories. The identification and analysis of DASEs and DASGs revealed that DASEs are dynamic, with the number of DASEs and DASGs being highest between the E7.0_A and E7.5_A lineages during mouse gastrulation. Functional enrichment analysis revealed that several DASGs are involved in RNA splicing metabolism and AS regulation, indicating that some genes may influence pre-mRNA AS through their own AS during lineage development. This finding highlighted the potential regulatory role of AS in development, particularly in key biological processes, such as cell fate determination and organ formation.

Systematic analysis of SF expression patterns revealed stage-specific expression profiles. Compared to the early stages of gastrulation (E6.5), SF expression levels are higher during the middle and late gastrulation stages (E7.0 and E7.5). Additionally, the dynamic SF expression demonstrated that AS events are stage-specific during lineage development. Analysis of isoform switch events revealed that these events are often detected in lineage-specific SFs. Different isoforms of SFs, such as Eftud2 and Ddx17, may play different regulatory roles during mouse gastrulation. Nevertheless, validation experiments should be carried out to demonstrate this. We systematically analyzed the epigenetic features adjacent to AS events. Specific and common modification patterns of AS events during gastrulation were identified, revealing different histone modification landscapes. Signal maps of modifications near splicing sites in different lineages were characterized. The results showed that the regions adjacent to AS sites exhibit distinct modification patterns correlating with the inclusion levels of AS events.

In summary, this study revealed global changes in AS during mouse gastrulation from a splicing perspective. The relationship between the epigenetic information and AS was characterized. Additionally, we found that SF expression is lineage-specific, and the inclusion rate of splicing events serves as a useful indicator of lineage differentiation. This study offers valuable insights into the regulatory mechanisms of mouse gastrulation. However, this study constitutes a bioinformatics analysis based on publicly available data. All results, findings, and conclusions lack experimental validation, which is a common limitation of bioinformatics work. In future experiments, we will functionally disrupt the expression of specific isoforms, similar to those performed by Lee et al.[54] or Xu et al.[55], to test their role on mouse gastrulation.

Data availability

The datasets analyzed for this study can be found in the Gene Expression Omnibus (GEO) repository. The accession numbers for the datasets are GSE98101 and GSE104243. No original data are generated in this study.

Received: 16 January 2025; Accepted: 26 March 2025
Published online: 31 March 2025

References

1. Zhai, J. et al. Primate gastrulation and early organogenesis at single-cell resolution. *Nature* **612**, 732–738 (2022).
2. Hu, B. et al. EmExplorer: A database for exploring time activation of gene expression in mammalian embryos. *Open. Biology*. **9**, 190054–190054 (2019).
3. Zheng, L. et al. EmAtlas: A comprehensive atlas for exploring Spatiotemporal activation in mammalian embryogenesis. *Nucleic Acids Res.* **51**, D924–d932 (2023).
4. Xiang, Y. et al. Epigenomic analysis of gastrulation identifies a unique chromatin state for primed pluripotency. *Nat. Genet.* **52**, 95–105 (2020).
5. Mittnenzweig, M. et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825–2842e2822 (2021).
6. Zeng, B. et al. The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell. Stem Cell.* **30**, 851–866e857 (2023).
7. Hong, Y. et al. An increment of diversity method for cell state trajectory inference of time-series scRNA-seq data. *Fundam Res.* **4**, 770–776 (2024).
8. Cui, G. et al. Spatial molecular anatomy of germ layers in the gastrulating cynomolgus monkey embryo. *Cell. Rep.* **40**, 111285 (2022).
9. Peng, G. et al. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528–532 (2019).
10. Zhai, J., Xiao, Z., Wang, Y. & Wang, H. Human embryonic development: From peri-implantation to gastrulation. *Trends Cell. Biol.* **32**, 18–29 (2022).
11. Zhai, J. et al. Neurulation of the cynomolgus monkey embryo achieved from 3D blastocyst culture. *Cell* **186**, 2078–2091 (2023). (2018).
12. Yang, X. et al. Silencing of developmental genes by H3K27me3 and DNA methylation reflects the discrepant plasticity of embryonic and extraembryonic lineages. *Cell Res.* **28**, 593–596 (2018).
13. Park, E. et al. The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* **102**, 11–26 (2018).
14. Pan, Q. et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
15. Bradley, R. K. & Anczuków, O. RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer*. **23**, 135–155 (2023).
16. Liu, Z. et al. Temporal dynamic analysis of alternative splicing during embryonic development in zebrafish. *Front. Cell. Dev. Biology*. **10**, 879795 (2022).
17. Liu, T. et al. CodLncScape provides a Self-Enriching framework for the systematic collection and exploration of coding LncRNAs. *Adv. Sci.* **11**, 2400009 (2024).
18. Lu, X. et al. Whole-transcriptome splicing profiling of E7.5 mouse primary germ layers reveals frequent alternative promoter usage during mouse early embryogenesis. *Biol. Open* **7** (2018).
19. Aich, M. et al. TOBF1 modulates mouse embryonic stem cell fate through regulating alternative splicing of pluripotency genes. *Cell. Rep.* **42**, 113177 (2023).
20. Torre, D. et al. Isoform-resolved transcriptome of the human preimplantation embryo. *Nat. Commun.* **14**, 6902 (2023).
21. Xing, Y. et al. Dynamic alternative splicing during mouse preimplantation embryo development. *Front. Bioeng. Biotechnol.* **8**, 35 (2020).
22. Mazin, P. V., Khaitovich, P., Cardoso-Moreira, M. & Kaessmann, H. Alternative splicing during mammalian organ development. *Nat. Genet.* **53**, 925–934 (2021).
23. Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859e2820 (2021).
24. Li, H. et al. Characterizing cellular differentiation potency and Waddington landscape via energy indicator. *Research* **6**, 0118 (2023).
25. Gómez-Redondo, I., Planells, B. & Navarrete, P. Gutiérrez-Adán, A. Role of alternative splicing in sex determination in vertebrates. *Sex. Dev.* **15**, 381–391 (2021).
26. Yang, X. et al. Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. *Cell Res.* **29**, 911–926 (2019).
27. Andrews, S. et al. FastQC: A quality control tool for high throughput sequence data. *Bioinformatics* **26**, 1968–1971 (2010).
28. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
29. Patro, R. et al. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14**, 417–419 (2017).
30. Love, M. I., Huber, W. & Anders, S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
31. Long, C. et al. Deciphering the decisive factors driving fate bifurcations in somatic cell reprogramming. *Mol. Therapy Nucleic Acids* **34**, 102044 (2023).
32. Liang, P. et al. HelPredictor models single-cell transcriptome to predict human embryo lineage allocation. *Brief. Bioinform.* 22. (2021).
33. Wang, H. et al. eHSCPr discriminating the cell identity involved in endothelial to hematopoietic transition. *Bioinformatics* **37**, 2157–2164 (2021).
34. Trincado, J. L. et al. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **19**, 40 (2018).
35. Brunson, J. C. Ggalluvial: Layered grammar for alluvial plots. *J. Open. Source Softw.* **5** (2020).
36. Huang, S. et al. ASTK: A machine learning-based integrative software for alternative splicing analysis. **6**, 2300594. (2024).
37. Guo, W., Calixto, C. P. G., Brown, J. W. S. & Zhang, R. TSIS: An R package to infer alternative splicing isoform switches for time-series data. *Bioinformatics* **33**, 3308–3310 (2017).
38. Li, H. et al. The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).
39. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137. (2008).
40. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
41. Ramírez, F. et al. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–191 (2014).

42. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: An R/Bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
43. Zhu, L. J. et al. ChIPpeakAnno: A bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **11**, 237 (2010).
44. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
45. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omics: J. Integr. Biol.* **16**, 284–287 (2012).
46. Notarnicola, C. et al. The RNA-binding protein RBPMS2 regulates development of gastrointestinal smooth muscle. *Gastroenterology* **143**, 687–697e689 (2012).
47. Akerberg, A. A. et al. RBPMS2 is a myocardial-enriched splicing regulator required for cardiac function. *Circ. Res.* **131**, 980–1000 (2022).
48. Osma-Garcia, I. C. et al. The splicing regulators TIA1 and TIAL1 are required for the expression of the DNA damage repair machinery during B cell lymphopoiesis. *Cell. Rep.* **41**, 111869 (2022).
49. Hasan, M. M. et al. i4mC-Mouse: Improved identification of DNA N4-methylcytosine sites in the mouse genome using multiple encoding schemes. *Comput. Struct. Biotechnol. J.* **18**, 906–912 (2020).
50. Manavalan, B. et al. Meta-4mCpred: A sequence-based meta-predictor for accurate DNA 4mC site prediction using effective feature representation. *Mol. Ther. Nucleic Acids* **16**, 733–744 (2019).
51. Pham, N. T. et al. ac4C-AFL: A high-precision identification of human mRNA N4-acetylcytidine sites based on adaptive feature representation learning. *Mol. Ther. Nucleic Acids* **35**, 102192 (2024).
52. Brown, S. J., Stoilov, P. & Xing, Y. Chromatin and epigenetic regulation of pre-mRNA processing. *Hum. Mol. Genet.* **21**, R90–96 (2012).
53. Kanehisa, M. et al. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–d677 (2025).
54. Lee, Y. J., Hong, K. H., Yun, J. & Oh, S. P. Generation of activin receptor type IIB isoform-specific hypomorphic alleles. *Genesis* **44**, 487–494 (2006).
55. Xu, X. et al. Murine fibroblast growth factor receptor 1alpha isoforms mediate node regression and are essential for posterior mesoderm development. *Dev. Biol.* **208**, 293–306 (1999).

Author contributions

WW: Investigation, data curation and analysis, writing-original draft. YZ: reviewing, formal analysis and editing, YZ: reviewing and editing, WY: conceptualization and Supervision. YX: funding acquisition, investigation, resources, supervision, visualization, writing-review and editing. All authors reviewed the manuscript.

Funding

This research was funded by The National Natural Science Foundation of China (62161039, 62371265), The Natural Science Foundation of Inner Mongolia (2024JQ10, 2024MS03054), Basic scientific research funding for universities directly under Inner Mongolia Autonomous Region (2023RCTD023), The 2025 Inner Mongolia Key Laboratory of Life Health and Bioinformatics Project.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96148-7.

Correspondence and requests for materials should be addressed to W.Y. or Y.X.