Deciphering the Cell-Specific Transcript Heterogeneity and Alternative Splicing during the Early Embryonic Development of Zebrafish

Xiumei Lin[1*], Xue Wang[1*], Chang Liu[1,2*], Chuanyu Liu[2*], Tao Zeng[3*], Ziqi Yuan[1,5*], Meidi Hu[6], Rong Xiang[1], Kaichen Zhao[1,4], Jie Zhou[1], Shichen Yang[2], Yang Wang[1], Kaifeng Meng[6], Hui Wang[4], Guangli He[1], Rui Zhao[1], Jiaheng Liu[1], Yunqi Huang[1], Jingfang Pan[1], Jialu Wang[1], Junyi Chen[3], Fei Guo[3], Yuliang Dong[2,3#], Xun Xu[2#], Daji Luo[6#], Ying Gu[2#], Longqi Liu[1#], Zhiqiang Dong[4#], Liang Chen[1#]

[1]BGI Research, Hangzhou 310030, China.

[2]BGI Research, Shenzhen 518083, China.

[3]BGI Hangzhou CycloneSEQ Technology Co., Ltd, Hangzhou 310030, China.

[4]College of Biomedicine and Health, College of Life Science and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China.

[5]College of Life Sciences, Wuhan University, Wuhan 430072, China.

[6]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture, Institute of Hydrobiology, Hubei Hongshan Laboratory, Chinese Academy of Sciences, Wuhan 430072, China.

*These authors contributed equally to this paper.

[#]Correspondence and requests for materials should be addressed to Yuliang Dong (dongyuliang@genomics.cn), Xun Xu (xuxun@genomics.cn), Daji Luo (luodaji@ihb.ac.cn), Ying Gu (guying@genomics.cn), Longqi Liu (liulongqi@genomics.cn), Zhiqiang Dong (dongz@mail.hzau.edu.cn) or Liang Chen (chenliang8@genomics.cn).

Abstract

Cell fate determination during early embryonic development is a complex process modulated by gene expression. The intricate interplay of transcriptional and post-transcriptional regulation is integral to the developmental trajectory of embryogenesis, yet how RNA processing may contribute to early development programming is largely elusive. Leveraging recent technological advances in single-molecule nanopore sequencing, we developed a single-cell long-read transcriptome sequencing technology, allowing a clear view of transcript diversity during zebrafish embryogenesis during pre- and post-zygotic genome activation (ZGA). A closer examination of the dynamic transcript usage and potential alternative splicing revealed that abundant stage-specific transcripts with differential coding potentials are involved in distinct biological functions. Specifically, we identified two cell populations at the onset of ZGA based on isoform diversity instead of gene profiling, which followed divergent

developmental trajectories toward the ectoderm and the presumptive ectoderm. These two populations of cells were characterized by divergent splicing regulations linked to differential RNA-binding proteins, including SNRPA and SFPQ. Altogether, using the single-cell long-read transcriptome sequencing strategy, our work has revealed the cell-specific transcriptome dynamics contributing to the cell fate determination during embryogenesis.

Introduction

The early stages of embryonic development are characterized by a remarkable degree of cellular heterogeneity, with common characteristics including asymmetric cell divisions, morphogen gradients, and localized signaling centers, contributing to the determination of distinct cell fates and the formation of complex tissues and organs(*1, 2*) For example, in the developing mouse embryo, cellular heterogeneity with different activation of Hippo signal pathway is crucial for the segregation of the inner cell mass and the trophectoderm, leading to the formation of the pluripotent epiblast and the extraembryonic tissues, respectively(*3*). Similarly, in zebrafish, the differential expression of Nodal signaling components drives the formation of mesoderm and endoderm, illustrating the functional significance of cellular heterogeneity in lineage specification (*4*). On the other hand, this heterogeneity emerges through a series of highly orchestrated processes that vary across different species(*5*). In mice and humans, early asymmetric divisions can be traced back to the 2-cell stage embryo, where differential distribution of cytoplasmic determinants and localized signaling events begin to establish distinct cell fates (*1, 6-8*). In model organisms such as zebrafish, the maternal-to-zygotic transition (MZT) marks a critical period where control of development shifts from maternally deposited RNAs and proteins to zygotic gene expression, further contributing to cellular diversity(*9*). Understanding the mechanisms that drive this heterogeneity is essential for elucidating the principles of developmental biology.

The emergence of cellular heterogeneity in embryonic development is orchestrated by a complex interplay of mechanisms in across various species(*1, 10*). Previous studies have mainly focused on regulatory processes of gene expression, including signal transduction pathways, transcriptional regulation, and epigenetic modifications(*3, 10-13*) For instance, in Caenorhabditis elegans, asymmetric cell divisions are pivotal in establishing cellular heterogeneity and the anterior-posterior axis, with PAR proteins crucially partitioning cell fate determinants during early embryonic divisions(*14, 15*). The identification of morphogen gradients in Drosophila melanogaster, such as the Bicoid and Hunchback gradients, provided foundational insights into how spatial and temporal regulation of signaling pathways can drive antero-posterior axis specification(*16-18*). In Xenopus laevis, the Spemann-Mangold organizer functions as a signaling center, emitting factors like Noggin and Chordin to pattern the embryonic axis(*19-21*). In vertebrates, regulatory mechanisms become increasingly intricate. In zebrafish, a widely used model organism for studying vertebrate development, several key mechanisms contribute to early developmental heterogeneity(*12, 22*). One of the pivotal events is zygotic genome activation (ZGA), marking the transition from maternal to zygotic control of development(*23*). The dramatic reprogramming of the transcriptome during this process is one of the main drivers of cellular heterogeneity(*24, 25*). The widespread application of high-throughput single-cell sequencing technology and spatial omics technology has clarified the differentiation fate of these heterogeneous cells and the changes in gene expression along the

developmental trajectory during zebrafish embryogenesis after ZGA(*26-28*). In addition to the cellular heterogeneity identified after ZGA, cell staining and microscopic imaging have shown that zebrafish embryos exhibit cellular heterogeneity and asynchronous cell division even before ZGA(*29*) However, the developmental fate and underlying mechanisms of these heterogeneous cells formed before the ZGA period remain unclear.

Transcript splicing and RNA processing play pivotal regulatory roles during embryonic development, orchestrating the precise temporal and spatial expression of transcripts necessary for cellular differentiation and tissue formation(*30*). For instance, the splicing factor Sxl (Sex-lethal) and Mbnl1 (Muscleblind-like 1) has been shown to regulate alternative splicing events crucial for sex determination, heart and skeletal muscle development(*9, 31*), demonstrating how splicing decisions can influence developmental outcomes. Furthermore, the transcript splicing processes significantly contribute to the generation of cellular heterogeneity by producing diverse transcript isoforms from a single gene, thereby expanding the functional repertoire of the genome(*32*). For example, the alternative splicing of the Dscam and Fgf8 gene produces isoforms with distinct functions in limb and nervous system development, highlighting the importance of splicing in tissue-specific transcript regulation(*33, 34*). Before the ZGA, maternal RNA processing and splicing also play a crucial role in setting the stage for subsequent developmental events. In Xenopus laevis and zebrafish, maternal mRNAs are selectively stabilized or degraded, and their splicing is tightly regulated to ensure proper embryonic development(*35, 36*). Additionally, the gene *YBX1*, associated with mRNA splicing and stability in mice, is crucial for embryonic development by regulating the alternative splicing of maternal mRNA and activating early zygotic genes(*37*). In vitro experiments with Pladienolide B (PlaB) demonstrated that inhibiting spliceosome function can re-induce pluripotent cells into totipotent stem cells, underscoring the necessity of normal alternative splicing for the transition from totipotency to pluripotency in early embryos(*38*). Therefore, systematically investigating alternative splicing events and transcript isoforms of maternal and zygotic genes using a model animal like zebrafish offers a novel perspective to understanding the mechanisms behind early embryonic cell heterogeneity, providing critical regulatory insights previously unattainable at the gene expression level.

The advent of long-read sequencing technologies has revolutionized our ability to study transcriptomic complexity. Unlike short-read sequencing, which often struggles with accurately reconstructing full-length transcripts, long-read sequencing provides the capability to sequence entire RNA molecules in a single read(*39*). This allows for the precise identification of transcript isoforms, including those generated by alternative splicing, and provides a more comprehensive view of the transcriptome. Advances in long-read sequencing technologies have enabled the comprehensive profiling of full-length transcript isoforms, revealing the complexity of splicing events and their regulatory roles in development(*40*). Recently, researchers have leveraged long-read sequencing technologies, such as PacBio's Single Molecule, Real-Time (SMRT) sequencing and Oxford Nanopore's platforms, to develop single-cell full-length transcriptomes(*41-43*). These advancements have extended beyond transcriptomics to include epigenomics, 3D genome organization, and other multi-omics analyses, offering deeper insights at single-cell resolution into the molecular mechanisms driving critical life processes such as nervous system function and early embryonic development(*44-46*) . These technologies have enabled the exploration of cellular heterogeneity,

revealing intricate details about cell-specific transcriptomes and the dynamic changes that occur during development.

Despite these advancements, the lack of comprehensive single-cell long-read transcriptomic studies on embryogenesis limits our ability to fully understand the complexity of transcript usage and alternative splicing, as well as its impact on cell fate determination. Here, we adopted the newly developed single-molecule nanopore sequencing platform (CycloneSEQ) (*47*)to establish a single-cell long-read transcriptome sequencing technology and dissect the global transcript heterogeneity landscapes based on comprehensive full-length transcript information during the critical window before and after ZGA of zebrafish embryo development. Our study provides important data resources and a robust and comprehensive full-length transcript data analysis methodology for researching the usage diversity of transcript isoforms and alternative splicing events during vertebrate embryogenesis.

Results

Full-length transcriptome diversity in zebrafish early embryonic development

Precise single-cell barcode demultiplexing is challenging and error-prone for long-read sequencing, due to the lower sequencing accuracy compared to the next-generation sequencing (NGS) platforms. To address this issue, we prepared a mix of cell suspension containing an equal amount of human (HEK293T) and mouse (NIH-3T3) cells for scRNA-seq library construction using the DNBelab C4 Droplet platform. Subsequently, we conducted the short-read sequencing by DNBSEQ and long-read sequencing by CycloneSEQ in parallel(*47*). By aligning to the valid cell barcode whitelist generated from short-read sequencing, we successfully demultiplexed single-cell barcodes for CycloneSEQ reads with high sequence consistency for analysis. (fig. S1A) (see Methods). To assess the accuracy of barcode demultiplexing, we conducted mixed-cell analyses on two datasets, revealing comparable mixing rates between the two replicates (fig. S1B). Additionally, we compared single-cell transcriptomes generated by CycloneSEQ with those from standard DNBSEQ and noted a high degree of concordance (fig. S1C), confirming the reliability of the barcoding detection approach. Consequently, we established the single-cell long-read sequencing technology by integrating short-read and long-read data to construct a more detailed information matrix that encompasses transcript expression and usage differences, thereby surpassing the limitations of traditional gene-level analytical approaches.

Zebrafish provide a vital model for embryogenesis research, enabling detailed investigation of molecular mechanisms from fertilization to lineage development due to their external and rapid early development(*48, 49*). We systematically investigated the dynamic changes in RNA processing and transcriptome diversity during zebrafish embryogenesis throughout MZT using long-read sequencing. Zebrafish embryos from various developmental stages including 128-cell (2.25 hours post-fertilization (hpf)), 512-cell (2.75 hpf), 1k-cell (3 hpf), Dome (4.3 hpf), and 50%-epiboly (5.25 hpf), were harvested for full-length transcriptome analysis (Fig.1A). We filtered for transcripts with an expression level greater than 5 (umi > 5) detected in at least three cells, identifying a total of 27,717 transcripts. Among these, 18,360 (66.24%) were full splicing isoforms (FSM), while 7,051 (25.44%) represented novel transcripts (Novel In Catalog (NIC) = 3.50%, Novel Not In Catalog (NNIC) = 21.92%) (fig. S1D). Additionally, the transcription start sites (TSS) of the representative novel transcripts

were supported by snATAC-seq peaks(*50*) (fig. S1E). Despite no discernible variation in length across transcript types, the novel transcripts exhibited significantly reduced expression levels compared to annotated transcripts, consistent with previous studies(*51*) (fig. S1, F and G). Both annotated and novel transcripts predominantly displayed protein-coding potential and maintained relative consistency across various developmental stages (fig. S1, H and I). The majority of genes corresponded to only one or two transcripts (fig. S1J), and a concordance was observed between genes and transcripts at each developmental stage (fig. S2, A and B).

Using short-read sequencing data, we integrated and clustered a total of 28,468 cells across five developmental time points, identifying 15 distinct cell clusters (Fig. 1B). Notably, prior to ZGA (at the 128-cell, 512-cell, and 1k-cell stages), there was considerable overlap among cell clusters (Fig. 1, B and C). These cell types were distinguished by specific marker genes: C0 exhibited high expression of *lsp1b*, C5 highly expressed *myof*, and C2 primarily expressed *grip2a* and *wnt8a* (fig. S2C). Unlike the well-documented *grip2a* and *wnt8a* functioning in dorsal-ventral axis formation in zebrafish(*12, 52*) ,the contributions of *lsp1b* and *myof* to early body axis patterning are yet to be defined. We therefore performed the *in situ* hybridization assay and observed a similar spatial expression pattern for *lsp1b* and *wnt8a* and a pronounced expression of *myof* in the ventral region (Fig. 1D). These findings hint at possible involvement of *lsp1b* and *myof* in dorsal-ventral axis development, while further research is needed to clarify their roles. However, prior to the 50%-epiboly stage, assigning definitive cellular identities remained challenging despite the presence of differential expression patterns. At 50%-epiboly, distinct cell types became discernible: C11 ectoderm was characterized by high levels of *sox2* and *sox3*, C10 mesoderm and endoderm by *msgn1* and *foxa3*, and the enveloping layer (EVL) by *krt4* and *cldne*. *Notch3*, crucial for the self-renewal and differentiation of cytotrophoblast progenitors in early human placental development(*53*), was prominent in C3, where *sox2* and *sox3* were expressed at lower levels, suggesting a presumptive ectoderm status. C4, lacking distinctive markers, was classified as deep cell layer (DEL) (Fig. 1B and fig. S2C).

Contrasting the distinct cell clustering result based on short-read data, the cell clusters with the long-read data becomes unclear due to the data sparsity (3 k reads/cell) (fig. S2, D and E). To resolve this issue, we adopted the SEACell strategy(*54*). We assigned the cell cluster information from short-read data to long-read data based on cell IDs and grouped cells into metacells within each cluster according to expression similarity (see Methods), yielding 3,851 metacells with mean gene expression exceeding 2000 (fig. S2F). Based on this matrix, we conducted cell clustering for long-read data and integrated the cluster information from short-read data (Fig.1B), revealing expression consistency between the two datasets (fig. S2, G and H). Our analysis of cell type-specific markers and their corresponding transcript information revealed that genes such as *wnt8a*, *krt4*, *apoc1* and *lsp1b* exhibit multiple transcript variants across cell types (Fig. 1E and fig. S2, I and J). In summary, we have developed a single-cell long-read sequencing technology integrated with the SEACell strategy, effectively addressing the sparsity challenges associated with long-read data and providing a comprehensive, transcriptome-wide resource throughout the MZT.

Global changes in transcriptome usage during zebrafish MZT

Transcript usage preference analysis provides a more nuanced understanding of gene functional changes throughout development than merely assessing gene expression levels(*55*). Although

MZT has been extensively studied at the level of transcription and epigenetic regulation(*56*), the dynamic alterations in transcript usage during this critical developmental period are merely explored. To identify specific transcript usage at each stage, we scrutinized isoform usage specific to each stage (see Methods). Our results revealed that the predominant transcripts at the 128-cell stage were primarily associated with genes involved in cell cycle, ribosome activity, and RNA metabolism, while transcripts predominantly used during the 512-cell and 1K-cell stages were linked to genes related to cell cycle transitions and gastrulation (Fig. 2A). Notably, the cell division-related gene *ccnd1* displayed different transcript variants with stage-specific usage patterns, suggesting that *ccnd1* has a dominant transcript at each developmental stages and potentially performs different functions (Fig. 2B).

To better understand transcript usage dynamics around ZGA, we categorized five developmental stages into three phases: pre-ZGA (128-cell stage), ZGA (512-cell and 1K-cell stages), and post-ZGA (Dome and 50% epiboly stages). Our findings indicated relatively few changes in transcript usage during ZGA, with more significant shifts occurring afterward (Fig. 2C). Initially, at ZGA, differential usage was prevalent in non-differentially expressed genes (non-DEGs), but this trend reversed post-ZGA, where over half of the changes involved differentially expressed genes (DEGs) (Fig. 2D). Moreover, transcript usage in non-DEGs during ZGA was characterized by a preference for more complete transcripts, featuring longer 3'-UTR (approximately 80%), completed open reading frame (ORF) gain (approximately 90%), and exon gain (approximately 87.5%), which are crucial for cell cycle regulation, transcription, and mRNA processing (Fig. 2E and fig. S3A). For example, *lrrc59*, key to mRNA translation and cell cycle regulation, underwent a significant isoform switch during ZGA. Specifically, Nonsense-Mediated Decay (NMD)-sensitive transcripts ENSDART00000137965 were prevalent before ZGA but diminished thereafter, coinciding with an upregulation of complete transcripts ENSDART00000137965 (Fig. 2F). Additionally, we observed a notable decrease in the expression of novel transcripts that were prevalent in the pre-ZGA phase, coinciding with a rise of annotated transcripts during ZGA (Fig. 2G and fig. S3B). This transition from novel- to annotated- transcripts indicated the significant shifts in functions related to cell cycling and RNA metabolism during ZGA activation.

Contrasting with the ZGA phase, we observed a notable rise in novel transcripts among those with differential usage for both non-DEGs and DEGs post-ZGA (fig. S3, C to F). This trend was associated with a preference for shorter ORF transcripts in the downregulated genes post-ZGA (fig. S3G), particularly in cell cycle and DNA damage response genes (fig. S3H). For instance, *cenpe*, essential for zebrafish zygotic cell division(*57*), preferred to be downregulated. Moreover, the transcripts that encoded shorter ORFs with fewer functional domains significantly increased in the post-ZGA phase (Fig. 2H), reflecting the gene regulation at both transcription and post-transcription level. Conversely, in the upregulated genes, most transcripts exhibited longer 3' UTRs in the post-ZGA phase (fig. S3G). The extension of 3' UTRs has been reported to regulate mRNA and protein translation efficiency, degradation, and subcellular localization through microRNAs (miRNAs) and RNA-binding proteins (RBPs) (*58*). The functions of these genes with longer 3'UTRs are mainly central to RNA splicing and processing (fig. S3H), indicating an increase in translational demand to support developmental progression post-ZGA. In summary, our single-cell long-read data analysis of zebrafish embryogenesis underscores a close association of differential transcript selection and gene

expression with developmental transitions before and after ZGA.

Global changes in alternative splicing during zebrafish MZT

Alternative splicing is an important post-transcriptional regulatory mechanism in eukaryotes that plays a critical role in cell differentiation, organ development, and tissue homeostasis maintenance(*59*). To elucidate the global changes in alternative splicing, as well as the potential interplay between gene expression and alternative splicing pre- and post-ZGA at the stage specific manner in zebrafish embryos, we employed Percent Spliced In (PSI) metric, gene expression level and exon inclusion level in our analysis (Fig. 3A) (see Methods). In contrast to the limited differential transcript usage between pre-ZGA and ZGA phases, we detected a comparable number of differential splicing events for the two comparisons (367 at ZGA vs Pre-ZGA and 436 at post-ZGA vs ZGA) (Fig. 3B). In ZGA compared to pre-ZGA, most splicing differences occurred in non-DEGs. Whereas in post-ZGA phase compared to ZGA, 66.50% of differential splicing events were associated with DEGs (Fig. 3C). To clarify the relationship between splicing events and DEGs, we selected genes with differential exon usage among the DEGs and categorized them into upregulated and downregulated groups (Fig. 3, D and E). Upregulated genes, such as *lin28a* and *zeb1a*, showed a higher incidence of exon inclusion events, whereas for genes encoding RNA-binding proteins, like *elval1a* and *hnrnpk,* a gradual transition from inclusion to exclusion was observed over time (Fig. 3D). Similarly, downregulated genes involved in cell cycle regulation, including *ccni, cdc20*, *anln* and *kif2c*, also displayed variations in exon usage (Fig. 3E).

RNA-binding proteins (RBPs) are key mediators of post-transcriptional RNA regulation(*60, 61*). To better understand the role of RBPs in alternative splicing throughout zebrafish embryogenesis, we performed the RBP motif prediction. First, we grouped these differential splicing events based on the occurrence of more inclusion or exclusion events during MZT. By employing rMAPS, we enriched for RBP motifs, thereby identifying 27 RBPs potentially involved in the regulation of alternative splicing (Fig. 3F) (see Methods). To find RBPs more closely associated with these splicing events, we selected 26 RBPs that showed gradually increasing expression over time (fig. S3I). Among these, we identified overlaps with candidate RBPs coding genes such as *snrpa*, *tial*, *rbm15b* and *srsf11* (Fig. 3F and fig. S3I). Notably, SNRPA binds to introns flanking target exons, facilitating the retention of those exons (Fig. 3G, left). In contrast, the RBP TIAL, which also increased over time, was primarily enriched in splicing events involving exclusion, suggesting a role in promoting exon skipping (Fig. 3G, right). To analyze which genes these RBPs might influence, we selected genes with differential splicing events across developmental stages. Among the downregulated genes, *ccni* exhibited increased usage of exon-included transcripts ENSDART00000129351 and transcript13732.21.nnic post-ZGA, while *kif2c* favored exon-skipped transcripts, showing a significant decrease in the expression of the exon-included transcript ENSDART00000109849 (Fig. 3, H and I). Taken together, our study provides insights into the nuanced dynamics of alternative splicing during zebrafish embryogenesis, identifying RNA-binding proteins (RBPs) such as SNRPA and TIAL in the regulation of exon inclusion and exclusion around the ZGA. Additionally, we observed stage-specific splicing variations in genes like *ccni* and *kif2c*.

Isoform diversity uncovering finer subpopulations

Previous research demonstrates that the transcript diversity analysis may uncover intermediate transitional cell types(*62*). To evaluate whether the transcript expression profiles infer more delicate state differences between cells than gene expression, we performed the entropy analysis to estimate the degree of expression heterogeneity for genes or transcripts within each cluster (see Methods). The results indicate that the gene expression profile of cells within a cluster is highly homogeneous for most cell clusters (rogue > 0.9), yet considerable intra-cluster heterogeneity at the transcript level (Fig. 4A). Therefore, we performed sub-clustering based on transcript diversity until the entropy values exceeded 0.9 for the majority of cells in a sub-cluster (fig. S4A) and identified a total of 24 sub-clusters. Among these clusters based on DEGs, further refinement was possible by leveraging transcript heterogeneity (Fig. 4B). The cluster C0 exhibited the highest heterogeneity and could be further divided into four major sub-clusters (Fig. 4C). Moreover, we conducted an in-depth analysis to explore isoforms and splicing diversity at the gene level across various cell clusters. The results revealed a modest increase in transcript diversity within groups C0 and C2, marked by a slightly higher proportion of genes with three or more isoforms (fig. S4B), while no significant variations were observed in the composition and proportion of splicing events (fig. S4C).

Next, we conducted a detailed analysis of cluster C0, identifying four sub-clusters that closely correspond to three developmental stages: 128-cell (mapped to C0_3), 512-cell (mapped to C0_1) and 1k-cell (mapped to C0_2 and C0_4). This suggested a temporal progression of cellular sub-clusters defined by transcriptomic diversity, despite minimal variations in overall gene expression (Fig. 4D). Our analysis of transcript usage preferences revealed a nuanced pattern of transcript selectivity among these sub-clusters. Notably, the sub-cluster C0_1 specifically favored transcripts ENSDART00000035235 (one transcript of *ubb*) and transcript19346.13.nnic (one transcript of *npm3*), which are implicated in proteasomal degradation(*63*) and chromatin remodeling(*64*), respectively (Fig. 4E). Sub-cluster C0_3 preferentially expressed transcripts, including transcript31038.10.nic from *npm1a*, a gene known for its roles in ribosome biogenesis, centrosome duplication, and hematopoiesis(*65*), as well as ENSDART00000149207 from *ccnb1*, which is involved in cell cycle regulation.

To identify sub-cluster-specific functions of differentially distributed transcripts, we performed structural prediction analysis. Notably, the predominant isoforms of *npm3* (transcript19346.13.nnic) and *npm1a* (transcript31038.10.nic) in C0_1 and C0_3 were tended to undergo NMD and predicted to be non-coding, respectively, which are in accordance with the downregulated levels at latter (Fig. 4F and fig. S4D). Conversely, *ubb (*ENSDART00000063357*)* and *ccnb1*(ENSDART00000035235) tend to utilize transcripts with more complete structural domains in C0_1, indicating their roles in cell cycle progression and proteasomal degradation (fig. S4, E and F). This underscores the importance of transcript usage in modulating gene activity in a stage-specific manner. In conclusion, focusing solely on gene expression levels can overlook the functional roles of specific transcripts. Our analysis underscores that transcript usage is pivotal for cellular function and development, offering a refined tool for deciphering cellular heterogeneity.

To further explore the relationship between alternative splicing and differential transcript usage among cell sub-clusters, we employed a splicing modalities analysis, which assesses changes in splicing levels at the single-cell level(*66*). From sub-cluster C0_3 (128-cell stage) to C0_1 (512-cell stage), 12% of splicing forms underwent modal transitions, primarily shifting

from bimodal to inclusion or exclusion patterns. However, the change in these splicing modalities from C0_1 to C0_2/C0_4 (1k-cell stage) showed a slight decrease, at only 9% (Fig. 4, G and H). In addition to transitions from multimodal to inclusion or exclusion patterns, there was also a notable increase in shift of inclusion towards bimodal splicing patterns from C0_1 to C0_2/C0_4 (Fig. 4G), which may reflect an upregulation of splicing activity associated with the ZGA.

Isoform subgrouping associated with distinct developmental trajectories

The majority of cell trajectory analyses are based on gene expression data, leaving the influence of transcript diversity on developmental trajectories an open question. To this end, we utilized fine-grained transcript sub-clusters to estimate the potential developmental trajectories using the KNN algorithm(*67*), leading to the construction of a transcript-based trajectory (see Method). We identified that the transcript sub-clusters C2_1 and C2_2 from the 512-cell stage C2 diverge along two distinct developmental trajectories, giving rise to cell types Presumptive ectoderm (C3_1: trajectory 2) and Ectoderm (C11_1: trajectory 1) at the 50%-epiboly stage (Fig. 5A and fig. S5A), each characterized by different expression profiles, such as *sox3* in the Ectoderm and *celsr1a*(*68*) in the Presumptive Ectoderm (Fig. 5B). To ascertain the role of alternative splicing in shaping distinct developmental trajectories, we analyzed the differential splicing events between trajectories 1 and 2. Our comparison revealed a higher incidence of splicing events in trajectory 1, predominantly characterized by skipped exons (SE) and alternative 3' splice site (A3) events (Fig. 5C). The genes undergoing differential splicing between the trajectories highlighted that trajectory 1 was predominantly associated with RNA processing, translation, and convergent extension (fig. S5B). In contrast, trajectory 2 was primarily linked to RNA metabolic processing and polycomb repressive complex (fig. S5B), with genes such as *bmi1a* and *rbbp4* involved in regulating gene expression and maintaining cellular pluripotency(*69, 70*).

We noted pronounced disparities in RBP expression between two trajectories: trajectory 1 displayed elevated RBP levels before ZGA (1k-cell stage), yet these differences waned post-ZGA, while the expression of the housekeeping gene actb1 remained constant (Fig. 5D and fig. S5C). To elucidate the RBPs driving these splicing variations, we conducted motif predictions for RBPs on differentially exon skipping events, revealing 28 motifs preferentially associated with trajectory 1 and 30 with trajectory 2 (Fig. 5E). The expression profiles of these RBPs mirrored this trend, showing marked divergence pre-ZGA (1k-cell), followed by a convergence (fig. S5D). Among these RBPs, SNRPA is mainly enriched in trajectory 1 and SFPQ in trajectory 2, participating in the regulation of alternative splicing via interactions with intronic regions of target genes (Fig. 5, E and F). Additionally, SFPQ has been reported to play a role in the development of the zebrafish nervous system(*71*). In addition, both splicing factors exhibited significant trajectory-specific expression compared to others (Fig. 5G). Upon closer examination of genes exhibiting differential exon usage between the trajectories, we observed distinct splicing patterns (Fig. 5H), even when the overall gene expression levels were comparable (fig. S5E). For example, *admp*, which is expressed in the anterior neuroectoderm and mesoderm during gastrulation(*72*), showed different transcript usage between the trajectories: trajectory 1 favored exon inclusion, whereas trajectory 2 had higher levels of exon-skipped transcripts (Fig. 5I and fig. S5F). In conclusion, transcript sub-clustering revealed two

developmental branches with distinct differentiation profiles. The early-stage RBP and splicing variations between these branches highlight the potential of alternative splicing to significantly influence cell fate determination at the onset of cellular commitment.

Discussion

Comprehensively employing the full-length transcript information through the single-cell long-read sequencing technology, our study elucidated, for the first time, the intricate isoform diversity, dynamic isoform usage, and alternative splicing in zebrafish embryos during MZT. Meanwhile, we developed a robust and comprehensive data analysis pipeline that encompasses long-read barcode demultiplexing, differential transcript usage, and alternative splicing analysis at single-cell resolution. This workflow is designed to reveal transcript heterogeneity and differential splicing among cell clusters, shedding light on the complexities of transcriptome diversity and the underlying regulatory mechanisms within the cellular. Our findings underscore the essential roles of isoform dynamics and alternative splicing in early development, complementing gene expression regulation. Simultaneously, the full-length transcript information enables us to delve deeper into the transcript usage switch over development, predict the biological functions of transcripts specific to various developmental stages, and infer the RBPs that regulate the corresponding splicing events. Moreover, at the onset of ZGA, we identified two distinct cell subpopulations with isoform diversities and divergent developmental pathways that converge on the ectoderm, each characterized by unique differentiation sequences. Notably, gene expression alone was insufficient to distinguish these diverse cell subclusters. Therefore, single-cell long-read sequencing technology has provided unprecedented insights into the complexity of transcriptome regulation during early embryonic development, underscoring the critical role of isoform diversity and alternative splicing in shaping cellular heterogeneity and developmental trajectories.

Around the critical developmental period of ZGA, we investigated developmental stage- and cell-specific isoforms. Yet, the functional significance of the numerous low-expression transcripts present remains unclear. In our analysis, we scrutinized the isoform usage frequency variations across the periods surrounding ZGA, including those isoforms with very low expression but notable changes in usage. Such as *admp*, is known for its role in early dorsoventral patterning and was found to utilize a novel transcript pre-ZGA. Though expression of this gene is documented to initiate during the blastula stage, it is yet to be determined if the protein functionality is active, given that the transcript variant appears to lack coding capacity(*72*). Additionally, our findings indicate that the extent of transcript variation during ZGA is comparatively minimal, with most isoforms exhibiting differential usage without corresponding changes in gene expression. This suggests the crucial role of transcript stability in safeguarding developmental processes leading up to ZGA. Conversely, post-ZGA, transcript changes are more pronounced, with over half of these changes leading to differential gene expression. This indicates that ZGA activates the differential usage of a substantial number of transcripts, thereby driving the differential expression of related genes to facilitate rapid embryonic development. This underscores the critical role of ZGA in orchestrating the complex regulatory networks that drive early embryonic development. Moreover, transcript structure predictions suggest that ZGA preferentially utilizes transcripts with more complete functions, further underscoring the pivotal role of ZGA in early embryonic development. Unlike transcript

usage, differential splicing events were already significantly occurring during ZGA. However, these differential splicing did not lead to a significant differential gene expression during ZGA, nor was there a necessary connection between changes in splicing events before and after ZGA and changes in gene expression, suggesting the distinct regulatory mechanisms governing gene expression and alternative splicing during early development.

More intriguingly, this study offers a novel perspective on the fascinating question of cell developmental fate in developmental biology by examining transcript-isoform and alternative splicing dynamics. Our findings reveal that cells can be clustered into two distinct subclusters based on transcript isoform diversity and coding capacities, with a clear chronological relationship in their subsequent developmental differentiation processes. Furthermore, our analysis of alternative splicing events and RBP activity indicates that the developmental trajectory characterized by faster differentiation progress exhibits more active splicing. This heightened splicing activity suggests a pivotal role for alternative splicing in driving the rapid differentiation of specific cell lineages. Based on these results, we infer that cell-type specific activity of isoform switching and splicing is a critical determinant in lineage differentiation. The dynamic regulation of transcript isoforms and splicing events not only contributes to the diversification of cellular functions but also orchestrates the precise timing of developmental transitions. This underscores the importance of transcriptomic diversity in shaping the developmental fate of cells and highlights the potential of targeting splicing mechanisms to influence cell differentiation pathways.

On the other hand, several limitations in this study need to be addressed in the future. The current single-cell long-read technology based on single-molecule sequencing is constrained by limited sequencing throughput, resulting in sparse data acquisition. Although applying analytical methods such as SEACell strategy can mitigate the disadvantages of data sparsity, and strict thresholds set in the analysis also ensure the reliability of the results, it is inevitable that some crucial transcripts may not be captured. In addition, the sequencing accuracy of single-molecule sequencing remains suboptimal, which can affect the precision of cell data segmentation in single-cell sequencing applications. Moreover, we encountered challenges in accurately distinguishing between zygotic and maternal transcripts. Consequently, we were unable to definitively ascertain whether the major contributions to transcript changes during development before and after ZGA were derived from maternal or zygotic genes. This limitation underscores the need for more refined techniques to differentiate between these transcript sources. Future efforts should focus on enhancing the throughput and accuracy of single-cell long-read sequencing technologies. Improvements in sequencing chemistry and bioinformatics algorithms are essential to achieve more comprehensive and precise transcriptome profiling(*73*). Additionally, developing better methods to distinguish between maternal and zygotic transcripts will be crucial for understanding the dynamics of gene regulation during early embryonic development. Furthermore, integrating multi-omics approaches, such as combining full-length transcriptomics with proteomics and epigenomics, will provide a more holistic view of the regulatory networks that drive cellular heterogeneity and development.

Materials AND Methods

Cell line and zebrafish embryos samples

Human embryonic kidney (HEK293T) and mouse fibroblast (NIH-3T3) cell lines, sourced from American Type Culture Collection (ATCC), were employed in this study. Additionally, we utilized AB wild-type zebrafish embryos, with all experiments approved by the Animal Care and Use Committee of Huazhong Agriculture University (HZAUFI-2021-0001). Embryos were harvested at various developmental stages post-fertilization: 128-cell (2.25 hpf), 512-cell (2.75 hpf), 1k-cell (3 hpf), Dome (4.3 hpf), and 50%-epiboly (5.25 hpf).

Single cell suspension isolation and full-length cDNA construction

Single-cell suspensions from HEK293T and NIH-3T3 cell lines were prepared following the DNBelab C4 protocol (MGI, 940-001818-00). Zebrafish embryo single-cell suspensions were isolated using established methods(*26*). We processed 3,000 to 5,000 single cells using the DNBelab C-TaiM 4 (MGI, 900-000637-00) to generate barcoded single-cell cDNA libraries, adhering to the DNBelab C4 library construction protocol (MGI, 940-001818-00).

Generating benchmarking data via parallel NGS of fragmented cDNA

A quarter (25% volume) of the barcoded full-length cDNA was subjected to fragmentation to prepare for short-read sequencing, adhering to the DNBelab C4 library construction protocol (MGI, 940-001818-00) and subsequently sequenced using a DNBSEQ-T20 sequencer. The raw data were processed using DNBC4tools (https://pypi.org/project/DNBC4tools/) to generate gene expression matrix.

CycloneSEQ library preparation and sequencing

The remaining aliquot of the single-cell libraries underwent 15 cycles of amplification with cDNA Amp Primer-V3 and purified with 0.6x DNA Clean Beads from DNBelab C4 library construction kit (MGI, 900-000637-00). The integrity and size distribution of the full-length cDNA libraries were evaluated using an Agilent Bioanalyzer 2100 (Agilent Technologies, G2939A). Additionally, library purity was quantified by measuring the A260/A280 ratio with a NanoDrop spectrophotometer (Thermo Fisher Scientific, NanoDrop 2000). CycloneSEQ libraries were then constructed in accordance with the CycloneSEQ library kit instructions (CycloneSEQ, H940-000001-00) and subsequently sequenced using the CycloneSEQ WT02 platform.

In situ hybridization

Custom primers, incorporating the T7 promoter sequence, were designed for the genes *wnt8a*, *myof* and *lsp1b*. The target fragments, generated through PCR amplification, were transcribed in vitro using T7 RNA Polymerase (Thermo Scientific, EP0111) and DIG RNA Labeling Mix (Sigma, 11277073910) to produce probes, which were stored at -80°C. Whole-mount in situ hybridization on 128-cell and shield-stage zebrafish embryos was conducted as previously described(*74*). Briefly, embryos were fixed in 4% PFA at 4°C, dehydrated in methanol, and stored at -20°C. After rehydration, they were washed in 1x PBST (Sigma, PPB005) and pre-hybridized at 70°C. Hybridization was performed with 1 ug/ml probe and RNase inhibitor (Takara, 2313A) at 70°C overnight. Post-hybridization washes included 2x SSC (Thermo Fisher Scientific, AM9770) at 70°C, followed by 0.2x SSC for 1 hour. Embryos were blocked with 1x PBST containing 2% goat serum and 2 mg/ml BSA (Thermo Fisher Scientific, 37525),

then incubated with 1/5000 Anti-Digoxigenin-AP (Sigma, 11093274910) at 4°C. Final washes were in 1x PBST and Alkaline Tris buffer. Color development was achieved with NBT/BCIP Stock Solution (Sigma, 11681451001), followed by the application of a stop solution and storage or imaging at 4°C in the dark.

Preprocessing of the CycloneSEQ read

Raw FASTQ files were preprocessed using Seqkit(*75*) to exclude reads shorter than 300 bp. Quality control was performed with Fastp(*76*), applying parameters -q 10 and -u 40. The quality-controlled reads were screened for expected adapters and filtered using Cutadapt to remove reads without adapters(*77*). Barcode demultiplexing was conducted using BWA by aligning the reads to a valid barcode list from DNBSEQ, with an allowed edit distance of 6 for 26-base pair barcodes(*78*). Subsequently, barcode reads were aligned to the zebrafish genome (GRCz11) using Minimap2(*79*) with the parameters (-ax splice, --secondary=no, -k14) to generate BAM files. UMI correction was then applied using UMI-tools(*80*), followed by quantification with Isoquant(*81*) to generate the single transcript matrix. Finally, the identified transcripts were classified and annotated using Sqanti3(*82*).

Single cell short read data analysis

The single-cell gene expression matrix generated by DNBC4tools was processed for downstream analysis using the Seurat R package (v4.1.1)(*83*). Initially, we utilized DoubletFinder(*84*) to identify and remove potential doublets, and further filtered out cells with fewer than 500 genes expressed or with a mitochondrial percentage exceeding 5%. Following this, we conducted log normalization, proceeded with dimensionality reduction, and then performed clustering of cell types based on differential gene expression.

Single cell long read data analysis

Single-cell long-read matrices obtained from Isoquant(*81*) were processed to filter out cells with transcript expression below 5 UMI and presence in at least 3 cells. To address sparsity in the transcript matrix, we employed SEACells to generate metacells(*54*). We then assigned cell annotation information from short-read data to long-read data based on cell IDs. Subsequently, we grouped cells into metacells within each cell type from the same developmental stage, ensuring that each metacell contained approximately 4,000 transcripts, based on expression similarity. Finally, we conducted downstream analysis using Scanpy(*85*), which involved log normalization and dimensionality reduction.

Differential transcript usage (DTU) analysis

To assess differential transcript usage (DTU) between two consecutive developmental stages, we utilized IsoformSwitchAnalyzeR(*86*). The raw expression matrix for transcripts at each sta ge was randomly partitioned into two groups (each comprising 2 replicates), to generate bulk-level transcript and gene matrices. We then conducted DTU analysis between these stages usi ng IsoformSwitchAnalyzeR, with a statistical cutoff set at $p < 0.05$ and a minimum absolute d ifference (|dif|) of greater than 0.1.

To elucidate differential transcript usage (DTU) at various stage or cluster levels, we perform ed single-cell DTU analysis. Initially, we constructed a transcript usage matrix from the metac

ell data, excluding transcripts expressed in fewer than 10% of cells across any stage or cluster. We then conducted two steps of differential testing: local and individual. Local differential te sting identified transcripts with distinct usage patterns across at least two stages or clusters, us ing a significance threshold of $p < 0.05$ and $|dif| > 0.1$. Subsequently, individual differential te sting was performed on the transcripts that met the criteria from the local testing to identify st age- and cluster-specific DTU, applying the same significance threshold of $p < 0.05$ and $|dif| > 0.1$.

Differential percent spliced in (PSI) analysis

To evaluate differential PSI events across different stages or clusters, we merged the metacell transcript matrix with the splicing information generated by SUPPA2(*87*) to create a single-ce ll PSI matrix. For differential PSI analysis, each event was filtered to be expressed in at least 100 cells and in at least 10% of cells in any given stage or cluster. We conducted pairwise co mparisons, ensuring a minimum of 50 effective cells in each group, and performed 100 permu tation tests, following the methodology outlined in MARVEL(*88*). Significant differential spli cing events were identified based on a threshold of p-value < 0.05 and $|dpsi| > 0.05$. For each event, modal identification can be performed using Anchor(*66*).

RBP motif analysis

We utilized the rMAPS platform for the analysis of RNA-binding protein (RBP) motifs associ ated with alternative splicing events(*89*) (http://rmaps.cecsresearch.org). We initially downloa ded 453 RNA-binding protein (RBP) binding sites from the oRNAment database(*90*)(https://r nabiology.ircm.qc.ca/oRNAment/downloads) and integrated them with the existing RBP moti fs within the rMAPS platform. Subsequently, we input differential splicing events for predicti ve analysis, focusing on RBPs with homologs in zebrafish for further examination.

Transcript sub-clustering analysis

We utilized ROGUE(*91*) to evaluate cellular heterogeneity at both the transcript and gene lev els. We leveraged differentially expressed genes at the gene level, with a significance threshol d of corrected $p < 0.05$ and $|logFC| > 0.5$, to perform isoform sub-clustering at a resolution of 1.6.

Isoform trajectory analysis

We utilized the k-nearest neighbors (kNN) algorithm to perform trajectory analysis based on t he isoform sub-cluster matrix. Briefly, we selected the 10 closest neighbors for temporal anal ysis between adjacent time points, requiring that at least 5 neighbors originated from the same cluster. Distances were measured to be less than 5 times the median total distance at that time point, using median absolute deviations (MADs) as the unit of measurement.

REFERENCES AND NOTES

1. Q. Chen, J. Shi, Y. Tao, M. Zernicka-Goetz, Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo. *Nat Commun* **9**, 1819 (2018).
2. D. W. Houston, Vertebrate Axial Patterning: From Egg to Asymmetry. *Adv Exp Med Biol* **953**, 209-306 (2017).
3. M. Zhu, M. Zernicka-Goetz, Principles of Self-Organization of the Mammalian Embryo. *Cell* **183**, 1467-1478 (2020).
4. A. F. Schier, W. S. Talbot, Molecular genetics of axis formation in zebrafish. *Annu Rev Genet* **39**, 561-613 (2005).
5. M. S. Thomsen *et al.*, Heterogeneity within and among co-occurring foundation species increases biodiversity. *Nat Commun* **13**, 581 (2022).
6. M. Zernicka-Goetz, S. A. Morris, A. W. Bruce, Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**, 467-477 (2009).
7. J. Wang *et al.*, Asymmetric Expression of LincGET Biases Cell Fate in Two-Cell Mouse Embryos. *Cell* **175**, 1887-1901 e1818 (2018).
8. S. Junyent *et al.*, The first two blastomeres contribute unequally to the human embryo. *Cell* **187**, 2838-2854 e2817 (2024).
9. M. T. Lee *et al.*, Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360-364 (2013).
10. A. L. Wilkinson, I. Zorzan, P. J. Rugg-Gunn, Epigenetic regulation of early human embryo development. *Cell Stem Cell* **30**, 1569-1584 (2023).
11. Z. Xiao *et al.*, 3D reconstruction of a gastrulating human embryo. *Cell* **187**, 2855-2874 e2819 (2024).
12. R. Fuentes *et al.*, The maternal coordinate system: Molecular-genetics of embryonic axis formation and patterning in the zebrafish. *Curr Top Dev Biol* **140**, 341-389 (2020).
13. S. M. Morgani, A. K. Hadjantonakis, Signaling regulation during gastrulation: Insights from mouse embryos and in vitro systems. *Curr Top Dev Biol* **137**, 391-431 (2020).
14. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100**, 64-119 (1983).
15. K. J. Kemphues, J. R. Priess, D. G. Morton, N. S. Cheng, Identification of genes required for cytoplasmic localization in early C. elegans embryos. *Cell* **52**, 311-320 (1988).
16. L. Wolpert, Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25**, 1-47 (1969).
17. A. Huang, T. E. Saunders, A matter of time: Formation and interpretation of the Bicoid morphogen gradient. *Curr Top Dev Biol* **137**, 79-117 (2020).
18. T. Fukaya, Dynamic regulation of anterior-posterior patterning genes in living Drosophila embryos. *Curr Biol* **31**, 2227-2236 e2226 (2021).
19. R. Harland, Neural induction. *Curr Opin Genet Dev* **10**, 357-362 (2000).
20. Y. Azbazdar, E. M. De Robertis, Molecular analysis of a self-organizing signaling pathway for Xenopus axial patterning from egg to tailbud. *Proc Natl Acad Sci U S A* **121**, e2408346121 (2024).
21. H. Spemann, H. Mangold, Induction of embryonic primordia by implantation of organizers from a different species. 1923. *Int J Dev Biol* **45**, 13-38 (2001).
22. F. L. Marlow, Setting up for gastrulation in zebrafish. *Curr Top Dev Biol* **136**, 33-83 (2020).

23. A. Sharma, G. J. Dsilva, G. Deshpande, S. Galande, Exploring the versatility of zygotic genome regulators: A comparative and functional analysis. *Cell Rep* **43**, 114680 (2024).
24. D. Jukam, S. A. M. Shariati, J. M. Skotheim, Zygotic Genome Activation in Vertebrates. *Dev Cell* **42**, 316-332 (2017).
25. J. Wragg, F. Muller, Transcriptional Regulation During Zygotic Genome Activation in Zebrafish and Other Anamniote Embryos. *Adv Genet* **95**, 161-194 (2016).
26. C. Liu *et al.*, Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev Cell* **57**, 1284-1298 e1285 (2022).
27. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987 (2018).
28. J. A. Farrell *et al.*, Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, (2018).
29. N. Olivier *et al.*, Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy. *Science* **329**, 967-971 (2010).
30. D. Torre *et al.*, Isoform-resolved transcriptome of the human preimplantation embryo. *Nat Commun* **14**, 6902 (2023).
31. K. Xiang *et al.*, Dinucleotide repeat polymorphism in the human SRC gene on chromosome 20. *Nucleic Acids Res* **19**, 6967 (1991).
32. T. Karakulak *et al.*, Heterogeneous and Novel Transcript Expression in Single Cells of Patient-Derived ccRCC Organoids. (2024).
33. G. Guo *et al.*, Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18**, 675-685 (2010).
34. D. Schmucker *et al.*, Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-684 (2000).
35. H. Aanes *et al.*, Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**, 1328-1338 (2011).
36. R. A. Blythe, K. Smith, A. D. Smith, Learning times for large lexicons through cross-situational learning. *Cogn Sci* **34**, 620-642 (2010).
37. M. Deng *et al.*, YBX1 mediates alternative splicing and maternal mRNA decay during pre-implantation development. *Cell Biosci* **12**, 12 (2022).
38. P. Y. Lin *et al.*, Comparative Analyses of Single-Cell Transcriptomic Profiles between In Vitro Totipotent Blastomere-like Cells and In Vivo Early Mouse Embryonic Cells. *Cells* **10**, (2021).
39. P. Kumari *et al.*, Advances in long-read single-cell transcriptomics. *Hum Genet*, (2024).
40. E. K. Oikonomou, M. Siddique, C. Antoniades, Artificial intelligence in medical imaging: A radiomic guide to precision phenotyping of cardiovascular disease. *Cardiovasc Res* **116**, 2040-2054 (2020).
41. T. Mantere, S. Kersten, A. Hoischen, Long-Read Sequencing Emerging in Medical Genetics. *Front Genet* **10**, 426 (2019).
42. M. MacKenzie, C. Argyropoulos, An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines (Basel)* **14**, (2023).
43. A. Payne, N. Holmes, V. Rakyan, M. Loose, BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* **35**, 2193-2198 (2019).
44. H. Yang *et al.*, A map of cis-regulatory elements and 3D genome structures in zebrafish.

Nature **588**, 337-343 (2020).
45. Y. Wang, Y. Zhao, A. Bollas, Y. Wang, K. F. Au, Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* **39**, 1348-1365 (2021).
46. A. Patowary *et al.*, Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders. *bioRxiv*, (2023).
47. J.-Y. Zhang *et al.*, A single-molecule nanopore sequencing platform. *bioRxiv*, (2024).
48. L. Ebarasi, A. Oddsson, K. Hultenby, C. Betsholtz, K. Tryggvason, Zebrafish: a model system for the study of vertebrate renal development, function, and pathophysiology. *Curr Opin Nephrol Hypertens* **20**, 416-424 (2011).
49. C. Shen, Z. Zuo, Zebrafish (Danio rerio) as an excellent vertebrate model for the development, reproductive, cardiovascular, and neural and ocular development toxicity study of hazardous chemicals. *Environ Sci Pollut Res Int* **27**, 43599-43614 (2020).
50. X. Lin *et al.*, Single-nucleus chromatin landscapes during zebrafish early embryogenesis. *Sci Data* **10**, 464 (2023).
51. H. Kiyose *et al.*, Comprehensive analysis of full-length transcripts reveals novel splicing abnormalities and oncogenic transcripts in liver cancer. *PLoS Genet* **18**, e1010342 (2022).
52. H. Hino *et al.*, Roles of maternal wnt8a transcripts in axis formation in zebrafish. *Dev Biol* **434**, 96-107 (2018).
53. B. Dietrich *et al.*, NOTCH3 signalling controls human trophoblast stem cell expansion and differentiation. *Development* **150**, (2023).
54. S. Persad *et al.*, SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data. *Nat Biotechnol* **41**, 1746-1757 (2023).
55. M. Werber, L. Wittler, B. Timmermann, P. Grote, B. G. Herrmann, The tissue-specific transcriptomic landscape of the mid-gestational mouse embryo. *Development* **141**, 2325-2330 (2014).
56. M. A. Eckersley-Maslin, C. Alda-Catalinas, W. Reik, Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol* **19**, 436-450 (2018).
57. K. W. Yu, Z. Y. She, Y. L. Wei, N. Zhong, Kinesin-7 CENP-E regulates cell division, gastrulation and organogenesis in development. *Eur J Cell Biol* **99**, 151107 (2020).
58. M. Chen *et al.*, 3' UTR lengthening as a novel mechanism in regulating cellular senescence. *Genome Res* **28**, 285-294 (2018).
59. F. E. Baralle, J. Giudice, Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* **18**, 437-451 (2017).
60. V. Despic, Identification of RNA-Binding Protein Landscapes Across Zebrafish Embryonic Transcriptome via iCLIP Approach. *Methods Mol Biol* **2218**, 331-345 (2021).
61. V. Despic *et al.*, Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Res* **27**, 1184-1194 (2017).
62. A. Patowary *et al.*, Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. *Science* **384**, eadh7688 (2024).
63. H. W. Ryu, C. W. Park, K. Y. Ryu, Restoration of cellular ubiquitin reverses impairments in neuronal development caused by disruption of the polyubiquitin gene Ubb. *Biochem Biophys Res Commun* **453**, 443-448 (2014).
64. N. Motoi *et al.*, Identification and characterization of nucleoplasmin 3 as a histone-binding protein in embryonic stem cells. *Dev Growth Differ* **50**, 307-320 (2008).

65. N. Bolli *et al.*, Expression of the cytoplasmic NPM1 mutant (NPMc+) causes the expansion of hematopoietic cells in zebrafish. *Blood* **115**, 3329-3340 (2010).
66. Y. Song *et al.*, Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol Cell* **67**, 148-161 e145 (2017).
67. P. Franti, O. Virmajoki, V. Hautamaki, Fast agglomerative clustering using a k-nearest neighbor graph. *IEEE Trans Pattern Anal Mach Intell* **28**, 1875-1881 (2006).
68. C. Li *et al.*, celsr1a is essential for tissue homeostasis and onset of aging phenotypes in the zebrafish. *Elife* **9**, (2020).
69. A. E. Herzog, R. Somayaji, J. E. Nor, Bmi-1: A master regulator of head and neck cancer stemness. *Front Oral Health* **4**, 1080255 (2023).
70. Y. Huang *et al.*, Rbbp4 Suppresses Premature Differentiation of Embryonic Stem Cells. *Stem Cell Reports* **16**, 566-581 (2021).
71. L. A. Lowery, J. Rubin, H. Sive, Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish. *Dev Dyn* **236**, 1347-1357 (2007).
72. Z. Lele, M. Nowak, M. Hammerschmidt, Zebrafish admp is required to restrict the size of the organizer and to promote posterior and ventral development. *Dev Dyn* **222**, 681-687 (2001).
73. S. L. Amarasinghe *et al.*, Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* **21**, 30 (2020).
74. C. Thisse, B. Thisse, High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* **3**, 59-69 (2008).
75. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962 (2016).
76. S. Chen, Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
77. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
78. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* **1303**, (2013).
79. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
80. T. Smith, A. Heger, I. Sudbery, UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491-499 (2017).
81. A. D. Prjibelski *et al.*, Accurate isoform discovery with IsoQuant using long reads. *Nat Biotechnol* **41**, 915-918 (2023).
82. F. J. Pardo-Palacios *et al.*, SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. *Nat Methods* **21**, 793-797 (2024).
83. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).
84. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
85. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).

86. K. Vitting-Seerup, A. Sandelin, IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**, 4469-4471 (2019).
87. J. L. Trincado *et al.*, SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**, 40 (2018).
88. W. X. Wen, A. J. Mead, S. Thongjuea, MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res* **51**, e29 (2023).
89. J. W. Park, S. Jung, E. C. Rouchka, Y. T. Tseng, Y. Xing, rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res* **44**, W333-338 (2016).
90. L. P. Benoit Bouvrette, S. Bovaird, M. Blanchette, E. Lecuyer, oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res* **48**, D166-D173 (2020).
91. B. Liu *et al.*, An entropy-based metric for assessing the purity of single cell populations. *Nat Commun* **11**, 3155 (2020).

ACKNOWLEDGMENTS

We thank all our teams' members, the China National GeneBank (CNGB) and Zhejiang Key Laboratory of Spatial Omics for their support. Project analysis was performed on the STOmics Cloud (https://cloud.stomics.tech).

Funding: This work was supported by the grant of Shenzhen Science and Technology Program (Grant No. RCJC20221008092804002).

Author Contributions

L.L., Z.D. and C.-Y. L. conceived the idea; C.-Y. L., C. L. and L.C. supervised the study; X.L., C. L. and C.-Y. L. designed the experiment; D.L. participated in the validation experiments with the help of M.H. and K.M.; K.Z., H.W., Y.W. and J.Z. collected the zebrafish embryo and generated single cell full-length cDNA libraries with the help of R.Z., G.H., Y.H. and J.P.; X.W. analyzed the data with the help of Z.Y., R.X., J.W., S.Y. and J.L.; T.Z., Y.D., J.C. and F.G. supported the single cell long read sequencing; X.L., C. L., L.C., X.W. and Z.Y. wrote the manuscript; L.L., Z.D., D.L., Y.G. and X.X. participated in the manuscript revised.

Competing Interests

The authors, affiliated with the BGI Group, have filed patent applications pertaining to the methodologies and outcomes detailed in this manuscript.

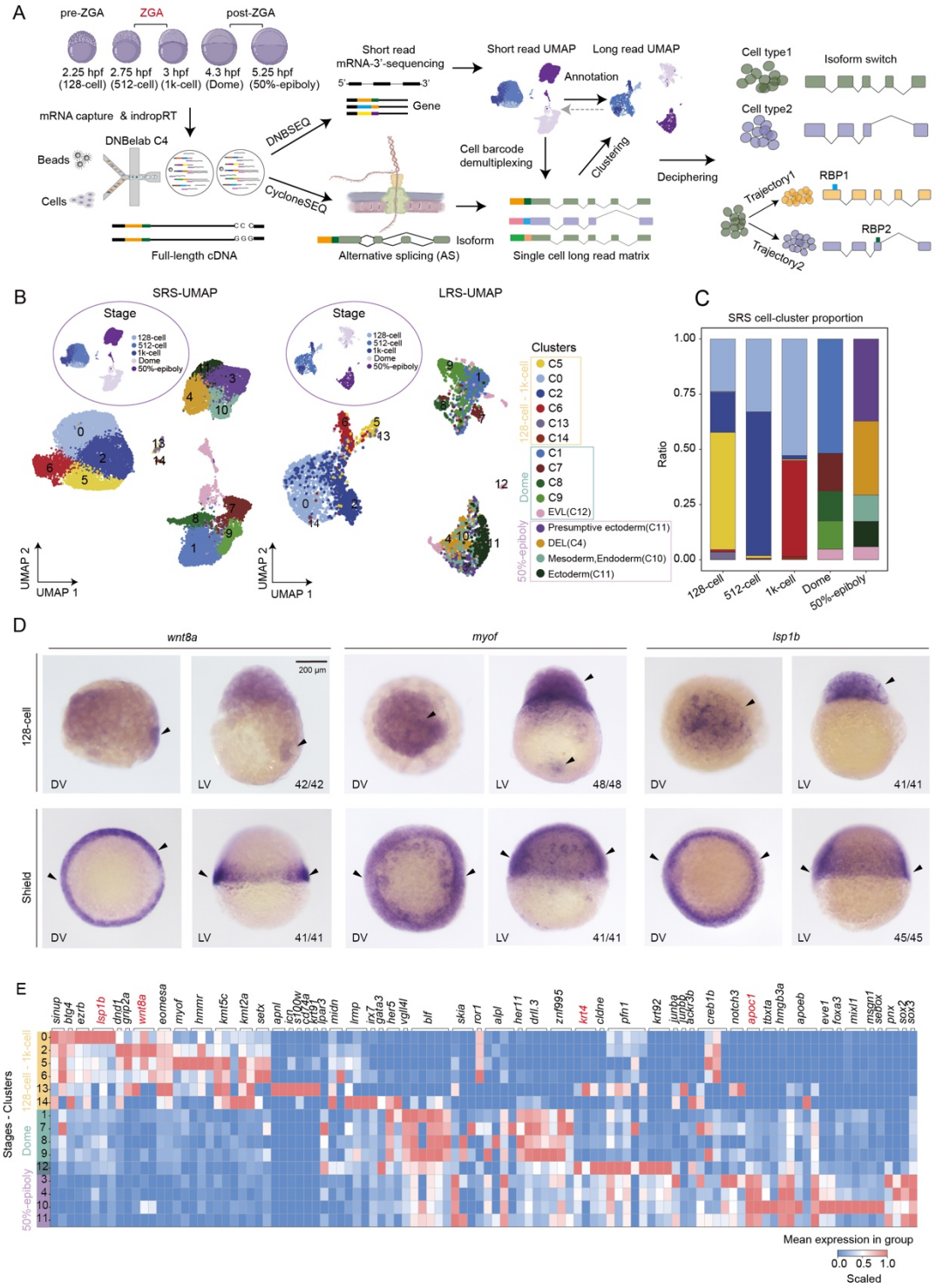
A
pre-ZGA
ZGA
post-ZGA
2.25 hpf (128-cell)
2.75 hpf (512-cell)
3 hpf (1k-cell)
4.3 hpf (Dome)
5.25 hpf (50%-epiboly)
mRNA capture & indropRT
DNBelab C4
Beads
Cells
Full-length cDNA
DNBSEQ
CycloneSEQ
Short read mRNA-3'-sequencing
Gene
Alternative splicing (AS)
Isoform
Short read UMAP
Long read UMAP
Annotation
Cell barcode demultiplexing
Clustering
Single cell long read matrix
Deciphering
Cell type1
Cell type2
Isoform switch
Trajectory1
Trajectory2
RBP1
RBP2
B
SRS-UMAP
LRS-UMAP
Stage
128-cell
512-cell
1k-cell
Dome
50%-epiboly
UMAP 1
UMAP 2
Clusters
128-cell - 1k-cell
C5
C0
C2
C6
C13
C14
Dome
C1
C7
C8
C9
EVL(C12)
50%-epiboly
Presumptive ectoderm(C11)
DEL(C4)
Mesoderm,Endoderm(C10)
Ectoderm(C11)
C
SRS cell-cluster proportion
Ratio
1.00
0.75
0.50
0.25
0.00
128-cell
512-cell
1k-cell
Dome
50%-epiboly
D
wnt8a
myof
lsp1b
128-cell
Shield
200 μm
DV
LV
42/42
48/48
41/41
45/45
E
Stages - Clusters
128-cell - 1k-cell
Dome
50%-epiboly
sinup
btg4
ezrb
lsp1b
dnd1
grip2a
wnt8a
eomesa
myof
hmmr
kmt5c
kmt2a
setx
apnl
icn
s100w
cd74a
krt91
lpar3
midn
lrmp
irx7
gata3
her5
vgll4l
blf
skia
ror1
alpl
her11
drll.3
znf995
krt4
cldne
pfn1
krt92
junba
junbb
ackr3b
creb1b
notch3
apoc1
tbxta
hmgb3a
apoeb
eve1
foxa3
mixl1
msgn1
sebox
pnx
sox2
sox3
Mean expression in group
0.0 0.5 1.0
Scaled

Fig. 1 Full-length transcriptome diversity in zebrafish early embryogenesis.

(A) Schematic illustrating transcriptomic diversity and alternative splicing in zebrafish embryogenesis, resolved with single-cell DNBSEQ and CycloneSEQ technologies. (B) UMAP plots illustrating cell clusters and stages: 28,468 cells detected by short read sequencing (left) and 3,851 metacells detected by long read sequencing (right). Colored by clusters and the developmental stages. (C) A bar chart shows cluster proportions across five developmental stages in short-read sequencing. The color legend of is the same as that in Fig.1B. (D) In situ imaging of *wnt8a*, *myof* and *lsp1b* at the 128-cell and shield stages. Scale bar, 0.2 mm. (E) The heatmap displays the expression of cluster specific genes with various isoforms by long-read sequencing, though isoform details are omitted from the heatmap.

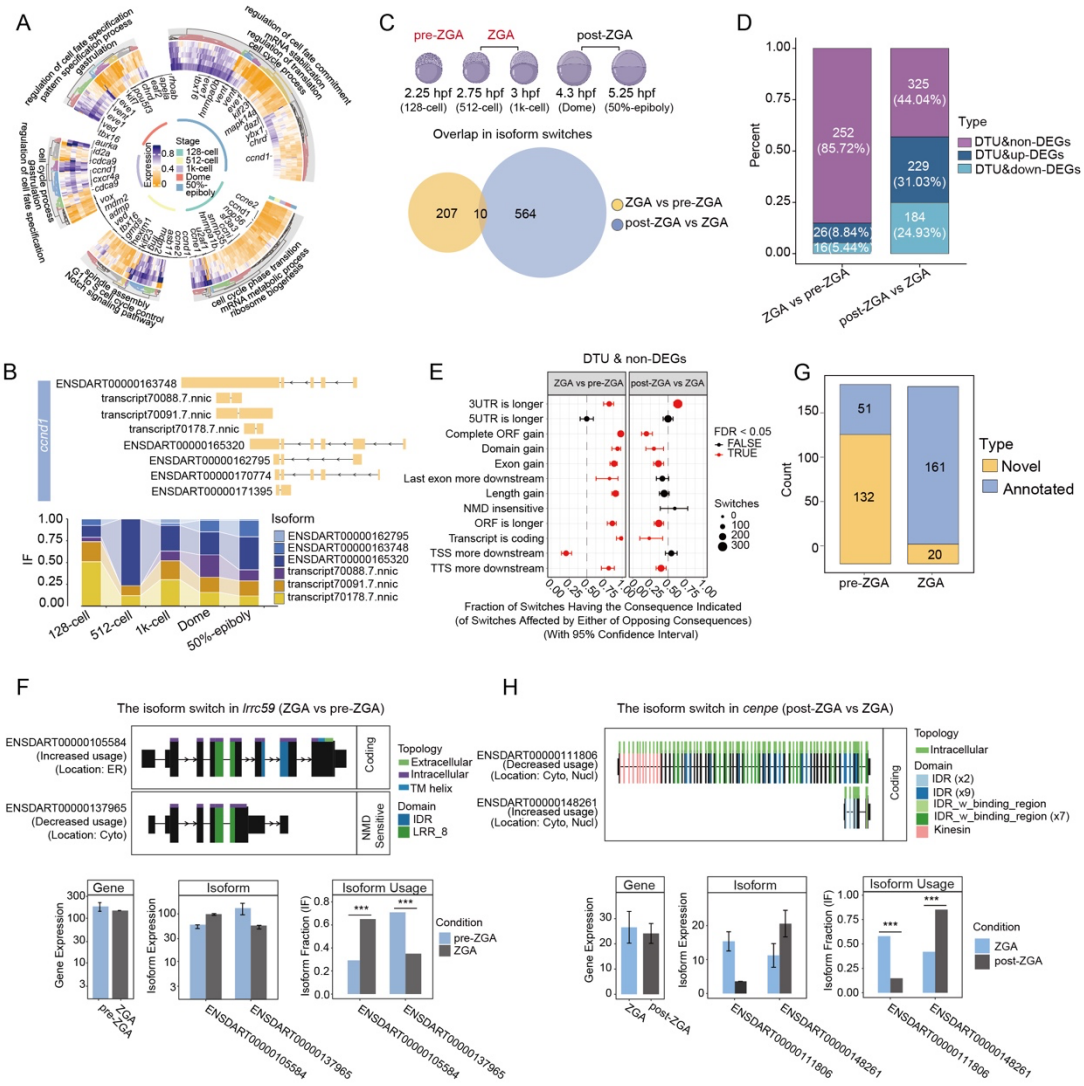
A
B
C
D
E
F
G
H
pre-ZGA
ZGA
post-ZGA
2.25 hpf (128-cell)
2.75 hpf (512-cell)
3 hpf (1k-cell)
4.3 hpf (Dome)
5.25 hpf (50%-epiboly)
Overlap in isoform switches
207
10
564
ZGA vs pre-ZGA
post-ZGA vs ZGA
Percent
252 (85.72%)
26(8.84%)
16(5.44%)
325 (44.04%)
229 (31.03%)
184 (24.93%)
Type
DTU&non-DEGs
DTU&up-DEGs
DTU&down-DEGs
ccnd1
ENSDART00000163748
transcript70088.7.nnic
transcript70091.7.nnic
transcript70178.7.nnic
ENSDART00000165320
ENSDART00000162795
ENSDART00000170774
ENSDART00000171395
IF
Isoform
128-cell
512-cell
1k-cell
Dome
50%-epiboly
DTU & non-DEGs
3UTR is longer
5UTR is longer
Complete ORF gain
Domain gain
Exon gain
Last exon more downstream
Length gain
NMD insensitive
ORF is longer
Transcript is coding
TSS more downstream
TTS more downstream
FDR < 0.05
FALSE
TRUE
Switches
Fraction of Switches Having the Consequence Indicated
(of Switches Affected by Either of Opposing Consequences)
(With 95% Confidence Interval)
Count
51
132
161
20
Novel
Annotated
The isoform switch in lrrc59 (ZGA vs pre-ZGA)
ENSDART00000105584
(Increased usage)
(Location: ER)
ENSDART00000137965
(Decreased usage)
(Location: Cyto)
Coding
NMD Sensitive
Topology
Extracellular
Intracellular
TM helix
Domain
IDR
LRR_8
Gene
Isoform
Isoform Usage
Gene Expression
Isoform Expression
Isoform Fraction (IF)
Condition
pre-ZGA
ZGA
The isoform switch in cenpe (post-ZGA vs ZGA)
ENSDART00000111806
(Decreased usage)
(Location: Cyto, Nucl)
ENSDART00000148261
(Increased usage)
(Location: Cyto, Nucl)
Intracellular
IDR (x2)
IDR (x9)
IDR_w_binding_region
IDR_w_binding_region (x7)
Kinesin
post-ZGA

Fig. 2 Global changes in transcriptome usage during zebrafish MZT.

(A) A circle heatmap illustrates differential transcript usage (DTU) patterns across developmental stages in long-read sequencing data. dIF > 0.1 and adjusted p-value < 0.05. (B) A schematic shows *ccnd1* transcript structure with a stacked bar chart detailing isoform fraction (IF) values at various stages. (C) The Venn diagram illustrates isoform switch events, comparing ZGA to pre-ZGA and post-ZGA to ZGA. (D) The stacked bar chart shows DTU and DGE between pre-ZGA vs ZGA, and ZGA vs post-ZGA. DTU is defined by dIF > 0.1 and p < 0.05, while DGE is categorized as upregulated (log2FC > 0.5, p < 0.05), downregulated (log2FC < -0.5, p < 0.05), or stable otherwise. (E) The illustration depicts functional consequences of isoform switching in non-DGE categories, comparing ZGA to pre-ZGA and post-ZGA to ZGA. (F) Isoform switching of the *lrrc59* gene was observed between pre-ZGA and ZGA stages. A bar chart illustrates the gene expression, isoform expression, and isoform usage. (G)The stacked bar chart displays the counts of annotated and novel transcripts, exhibiting isoform switching in non-DEG categories between pre-ZGA and ZGA stages. (H) Isoform switching of the *cenpe* gene observed between ZGA and post-ZGA stages. A bar chart shows the gene expression, isoform expression, and isoform usage.

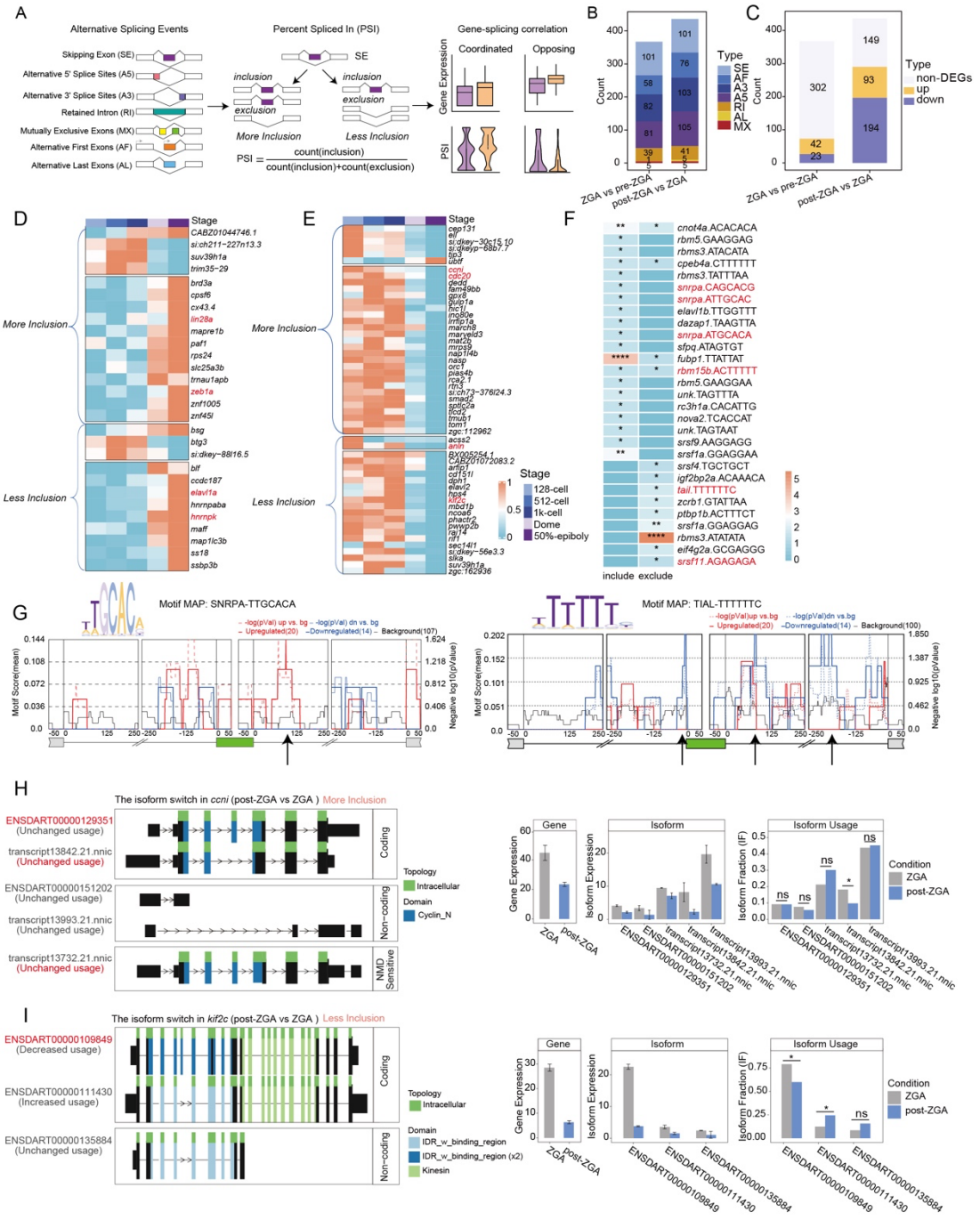
A
Alternative Splicing Events
Skipping Exon (SE)
Alternative 5' Splice Sites (A5)
Alternative 3' Splice Sites (A3)
Retained Intron (RI)
Mutually Exclusive Exons (MX)
Alternative First Exons (AF)
Alternative Last Exons (AL)
Percent Spliced In (PSI)
SE
inclusion
exclusion
More Inclusion
Less Inclusion
PSI = count(inclusion) / count(inclusion)+count(exclusion)
Gene-splicing correlation
Coordinated
Opposing
Gene Expression
PSI
B
Count
Type
SE
AF
A3
A5
RI
AL
MX
ZGA vs pre-ZGA
post-ZGA vs ZGA
C
Type
non-DEGs
up
down
D
Stage
More Inclusion
Less Inclusion
E
Stage
128-cell
512-cell
1k-cell
Dome
50%-epiboly
F
include
exclude
G
Motif MAP: SNRPA-TTGCACA
Motif MAP: TIAL-TTTTTTC
Motif Score(mean)
Negative log10(pValue)
Upregulated(20)
Downregulated(14)
Background(107)
Background(100)
H
The isoform switch in ccni (post-ZGA vs ZGA) More Inclusion
ENSDART00000129351 (Unchanged usage)
transcript13842.21.nnic (Unchanged usage)
ENSDART00000151202 (Unchanged usage)
transcript13993.21.nnic (Unchanged usage)
transcript13732.21.nnic (Unchanged usage)
Coding
Non-coding
NMD Sensitive
Topology
Intracellular
Domain
Cyclin_N
Gene
Isoform
Isoform Usage
Gene Expression
Isoform Expression
Isoform Fraction (IF)
Condition
ZGA
post-ZGA
I
The isoform switch in kif2c (post-ZGA vs ZGA) Less Inclusion
ENSDART00000109849 (Decreased usage)
ENSDART00000111430 (Increased usage)
ENSDART00000135884 (Unchanged usage)
Topology
Intracellular
Domain
IDR_w_binding_region
IDR_w_binding_region (x2)
Kinesin

Fig. 3 Global changes in alternative splicing during zebrafish MZT.

(A) The diagram illustrates splicing event patterns and PSI calculation. (B) The stacked bar chart displays the proportion of differential splicing events (DPSI) between ZGA vs pre-ZGA and post-ZGA vs ZGA. |DPSI| > 0.05 and p-value < 0.05. (C) The stacked bar chart visualizes the relationship between the differential splicing events and DEGs, upregulation gene: adjusted p-value < 0.05 and log fold change > 0.5; downregulation gene: adjusted p-value < 0.05 and log fold change < -0.5. (D) The heatmap visualizes upregulated genes across developmental stages that correspond to DPSI in exon skipping events. DPSI > 0.05 and adjusted p-value < 0.05. (E) The heatmap visualizes downregulated genes across developmental stages that correspond to DPSI in exon skipping events. DPSI < 0.05 and p-value < 0.05. (F) The heatmap displays the enrichment results of RNA-binding protein (RBP) motifs corresponding to DPSI in exon skipping event across time points. |DPSI| > 0.05 and p-value < 0.05. (G) The enrichment results for the SNRPA and TIAL1 RNA-binding proteins (RBPs) are depicted, with arrows indicating the binding positions of these RBPs. (H-I) The switch plot presents results for the *cnni* (H) and *kif2c* (I) between ZGA and post-ZGA. A bar chart illustrates gene expression, isoform expression and isoform usage.

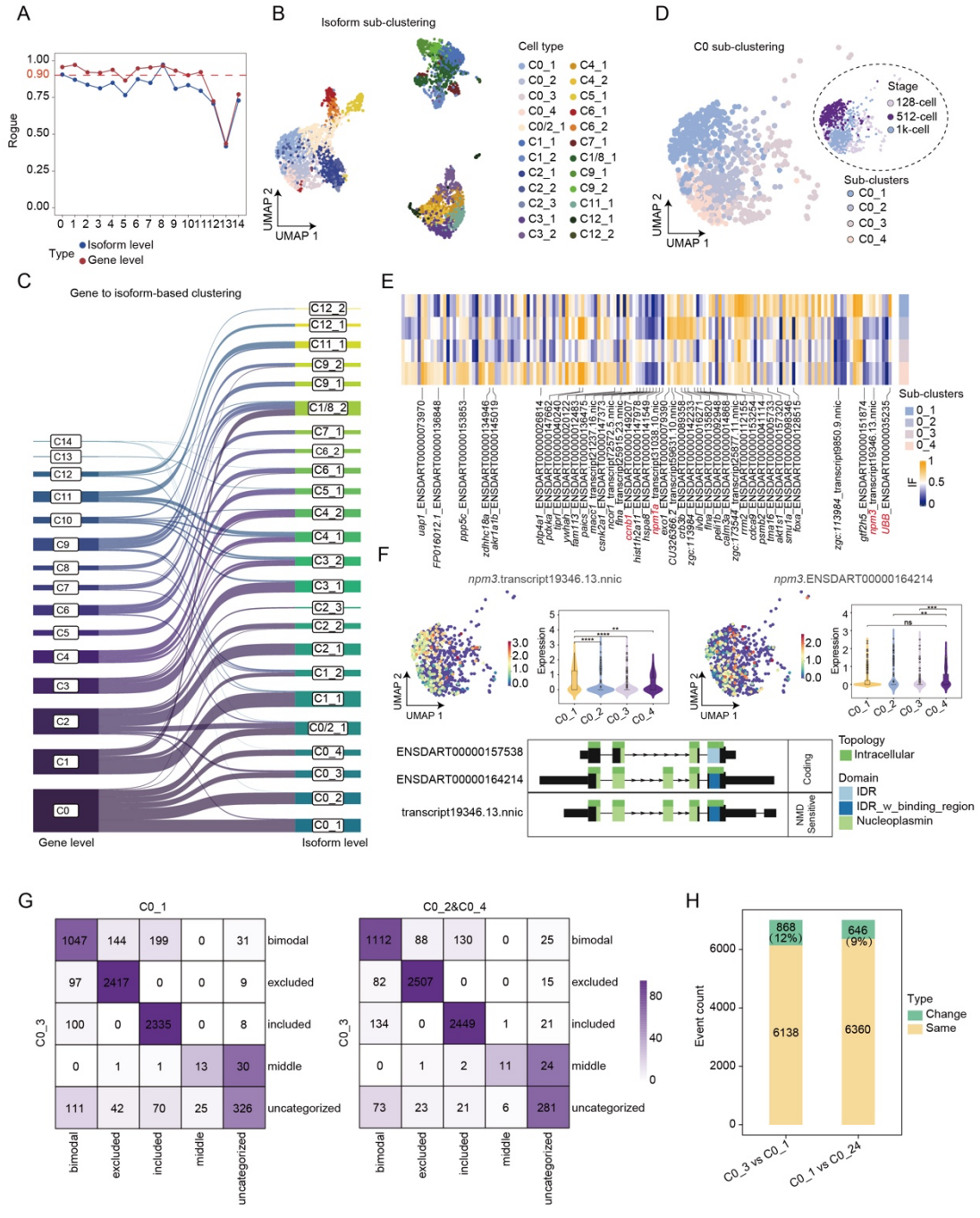
A
Rogue
Type
Isoform level
Gene level
B
Isoform sub-clustering
UMAP 1
UMAP 2
Cell type
C0_1
C0_2
C0_3
C0_4
C0/2_1
C1_1
C1_2
C2_1
C2_2
C2_3
C3_1
C3_2
C4_1
C4_2
C5_1
C6_1
C6_2
C7_1
C1/8_1
C9_1
C9_2
C11_1
C12_1
C12_2
D
C0 sub-clustering
Stage
128-cell
512-cell
1k-cell
Sub-clusters
C0_1
C0_2
C0_3
C0_4
C
Gene to isoform-based clustering
Gene level
Isoform level
E
Sub-clusters
IF
F
npm3.transcript19346.13.nnic
npm3.ENSDART00000164214
Expression
ENSDART00000157538
ENSDART00000164214
transcript19346.13.nnic
Coding
NMD Sensitive
Topology
Intracellular
Domain
IDR
IDR_w_binding_region
Nucleoplasmin
G
C0_1
C0_2&C0_4
C0_3
bimodal
excluded
included
middle
uncategorized
H
Event count
868 (12%)
646 (9%)
6138
6360
Type
Change
Same
C0_3 vs C0_1
C0_1 vs C0_24

Fig. 4 Isoform diversity uncovering finer subpopulations.

(A) A line chart displays cellular heterogeneity at gene and transcript levels in long read data, as assessed by ROGUE. (B) A UMAP plot illustrates detailed sub-clustering at the transcript level from long read data. (C) River plot showing mapping of cells from gene-based (left) to isoform-based (right) clustering. Each line represents a single cell. (D) A UMAP plot displays the sub-cluster C0 at the transcript level from long read data. (E) Heatmap showing the isoform faction with DTUs between sub-cluster in C0. dIF> 0.1 and adjusted p-value < 0.05. (F) Violin and UMAP plots (top) display the expression of *npm3* transcripts within the C0 sub-cluster. The switch plot (bottom) presents the results for *npm3* transcript isoform switching. (G) Stacked bar chart illustrating modal switching events between C0_3 and C0_1, and between C0_3 and C0_2&C0_4. (H) Stacked bar chart illustrating modal switching events between C0_3 to C0_1 and C0_3 to C0_2 & C0_4.

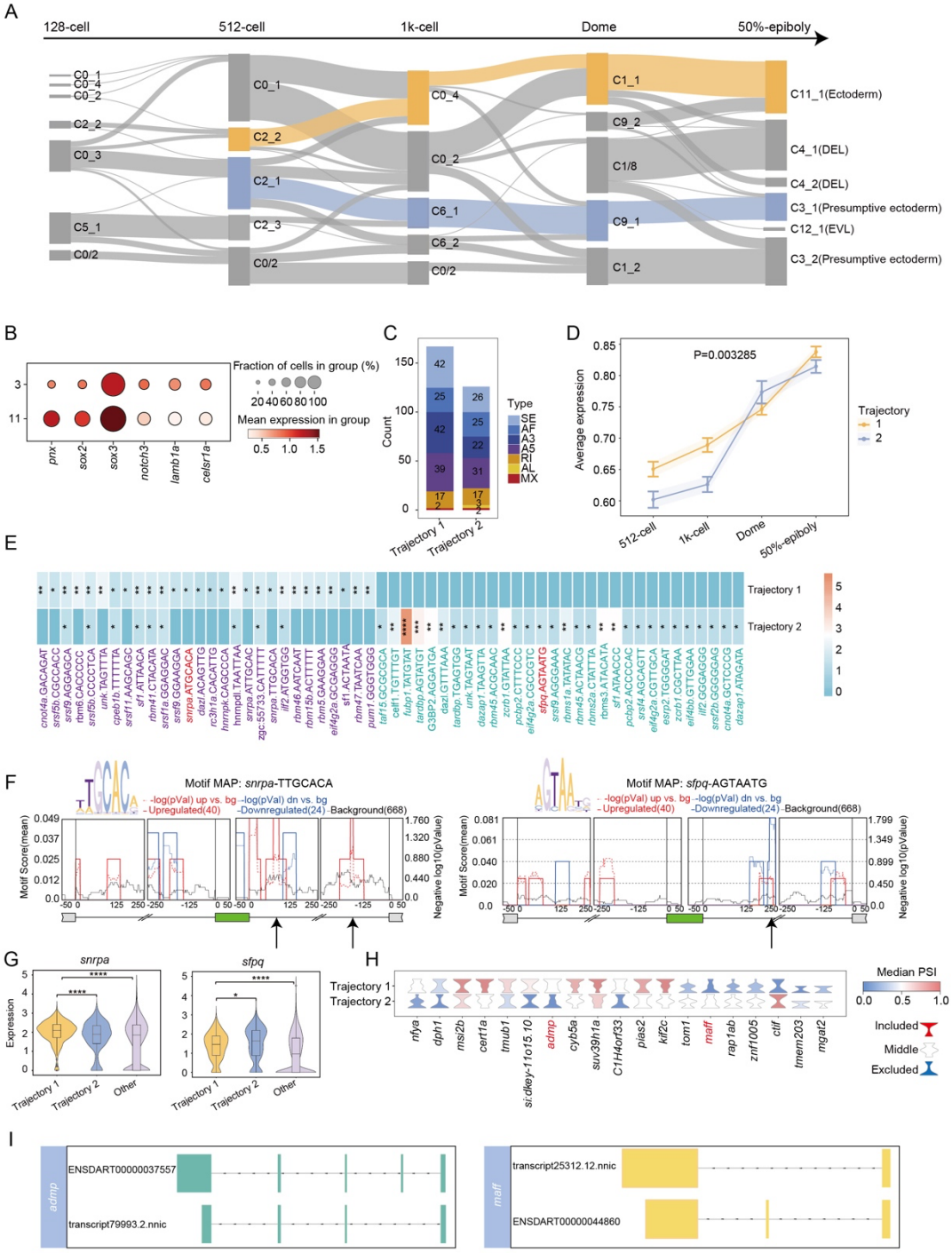
A
128-cell
512-cell
1k-cell
Dome
50%-epiboly
C11_1(Ectoderm)
C4_1(DEL)
C4_2(DEL)
C3_1(Presumptive ectoderm)
C12_1(EVL)
C3_2(Presumptive ectoderm)
B
Fraction of cells in group (%)
Mean expression in group
pnx
sox2
sox3
notch3
lamb1a
celsr1a
C
Count
Type
Trajectory 1
Trajectory 2
D
P=0.003285
Average expression
Trajectory
512-cell
1k-cell
Dome
50%-epiboly
E
Trajectory 1
Trajectory 2
F
Motif MAP: snrpa-TTGCACA
Motif MAP: sfpq-AGTAATG
Motif Score(mean)
Negative log10(pValue)
G
snrpa
sfpq
Expression
Trajectory 1
Trajectory 2
Other
H
Median PSI
Included
Middle
Excluded
I
admp
maff
ENSDART00000037557
transcript79993.2.nnic
transcript25312.12.nnic
ENSDART00000044860

Fig. 5 Isoform subgrouping associated with distinct developmental trajectories.
(A) A Sankey diagram visualizes the developmental trajectory of zebrafish embryos, incorporating isoform sub-clusters identified in long read data. Two developmental branches, leading to the ectoderm and presumptive ectoderm, are highlighted with distinct colors. (B) A dotted heatmap illustrating the marker gene expression profiles for presumptive ectoderm (C3) and ectoderm (C11). (C) The stacked bar chart displays the proportion of differential splicing events between two trajectories. $\Delta PSI > 0.1$ and $p < 0.05$. (D) Line chart of RBP expression profiles from long read sequencing across time points in two trajectories. (E) The heatmap displays the enrichment results of RNA-binding protein (RBP) motifs corresponding to differential splicing events (DPSI) associated with exon skipping event between two trajectories. $P < 0.05$ and $\Delta PSI > 0.1$. (F) The enrichment results for the SNRPA and SFPQ RNA-binding proteins (RBPs) are depicted, with arrows indicating the binding positions of these RBPs. (G) Violin plot depicting the expression patterns of *snrpa* and *sfpq* across two specified trajectories in comparison to other trajectories. (H) Violin plot showcasing genes with differential exon usage between two trajectories. $\Delta PSI > 0.3$ and $p < 0.05$. (I) Transcript structure diagram of *admp* and *maff*.

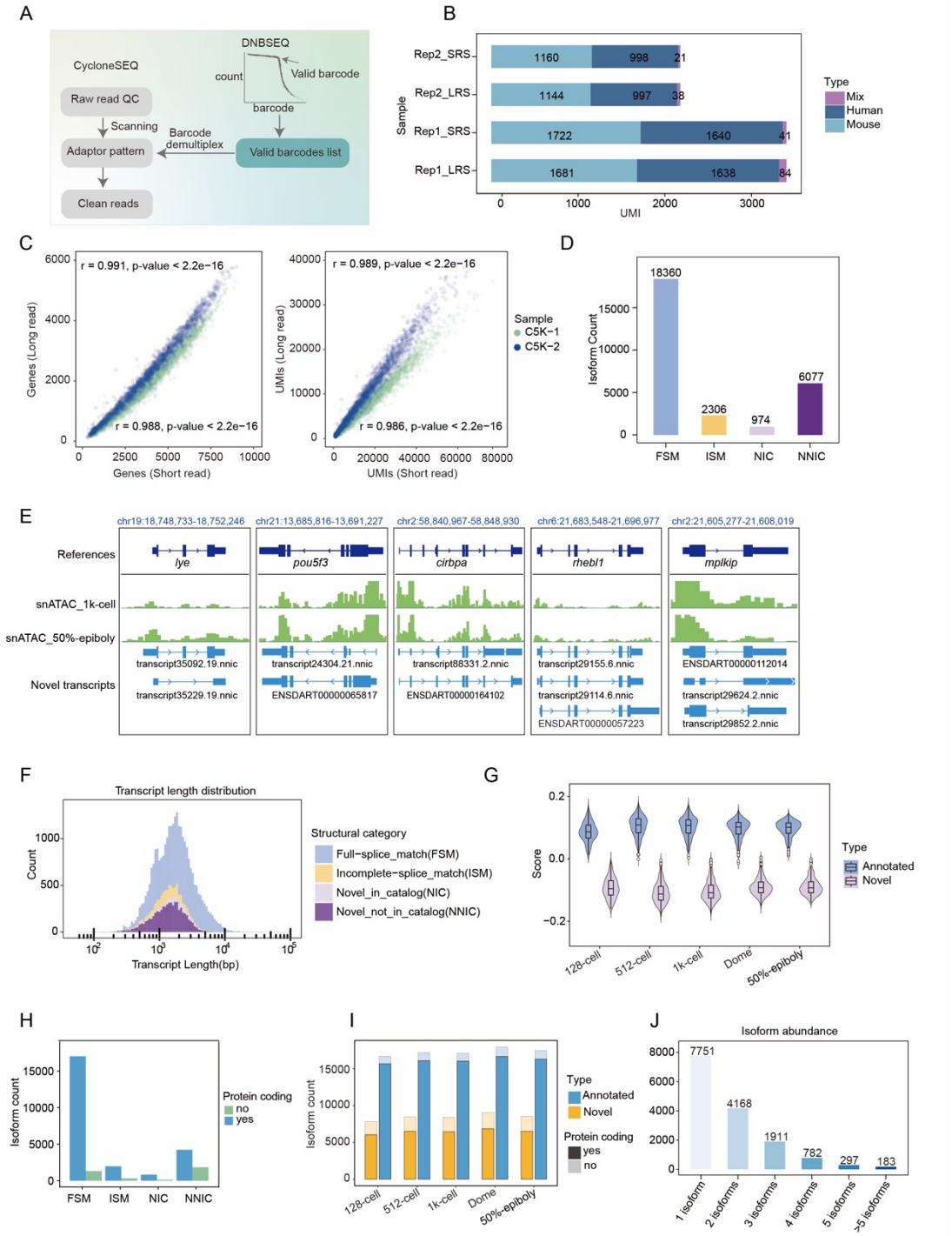
A
CycloneSEQ
Raw read QC
Scanning
Adaptor pattern
Clean reads
Barcode demultiplex
DNBSEQ
count
barcode
Valid barcode
Valid barcodes list
B
Rep2_SRS
Rep2_LRS
Rep1_SRS
Rep1_LRS
Sample
1160
998
21
1144
997
38
1722
1640
41
1681
1638
84
UMI
Type
Mix
Human
Mouse
C
r = 0.991, p-value < 2.2e−16
r = 0.988, p-value < 2.2e−16
Genes (Long read)
Genes (Short read)
r = 0.989, p-value < 2.2e−16
r = 0.986, p-value < 2.2e−16
UMIs (Long read)
UMIs (Short read)
Sample
C5K−1
C5K−2
D
Isoform Count
18360
2306
974
6077
FSM
ISM
NIC
NNIC
E
chr19:18,748,733-18,752,246
chr21:13,685,816-13,691,227
chr2:58,840,967-58,848,930
chr6:21,683,548-21,696,977
chr2:21,605,277-21,608,019
References
lye
pou5f3
cirbpa
rhebl1
mplkip
snATAC_1k-cell
snATAC_50%-epiboly
Novel transcripts
transcript35092.19.nnic
transcript35229.19.nnic
transcript24304.21.nnic
ENSDART00000065817
transcript88331.2.nnic
ENSDART00000164102
transcript29155.6.nnic
transcript29114.6.nnic
ENSDART00000057223
ENSDART00000112014
transcript29624.2.nnic
transcript29852.2.nnic
F
Transcript length distribution
Count
Transcript Length(bp)
Structural category
Full−splice_match(FSM)
Incomplete−splice_match(ISM)
Novel_in_catalog(NIC)
Novel_not_in_catalog(NNIC)
G
Score
Type
Annotated
Novel
128-cell
512-cell
1k-cell
Dome
50%-epiboly
H
Isoform count
Protein coding
no
yes
FSM
ISM
NIC
NNIC
I
Isoform count
Type
Annotated
Novel
Protein coding
yes
no
128-cell
512-cell
1k-cell
Dome
50%-epiboly
J
Isoform abundance
7751
4168
1911
782
297
183
1 isoform
2 isoforms
3 isoforms
4 isoforms
5 isoforms
>5 isoforms

fig. S1 Full-length transcriptome diversity in zebrafish early embryogenesis.
(A) Schematic diagram illustrating the long read single cell barcode demultiplexing. (B) The stacked histogram depicts UMI expression levels in mouse (NIH-3T) and human (HEK293T) cells across long read and short read data. (C) A scatter plot shows the correlation of gene and UMI counts per cell between short-read and long-read sequencing in C5K samples. (D) Bar chart showing the distribution of counts across various transcript structural categories. (E) Novel transcripts of the *lye*, *pou53*, *cirbpa*, *rhebl1* and *mplkip* genes were validated using snATAC-seq datasets, showing peak evidence at the transcription start site (TSS), as visualized in IGV. (F) A histogram displays the length distribution and counts of transcripts across different transcript structural categories. (G) A violin plot depicts the overall transcript expression scores for both annotated and novel transcripts across five developmental stages. (H) A bar chart illustrates the coding potential of transcripts, as predicted by CPAT, across different transcript structural categories. (I) A stacked bar chart depicts the coding potential of both annotated and novel transcripts across different developmental stages. (J) A bar chart shows the count of different isoform abundance per gene.

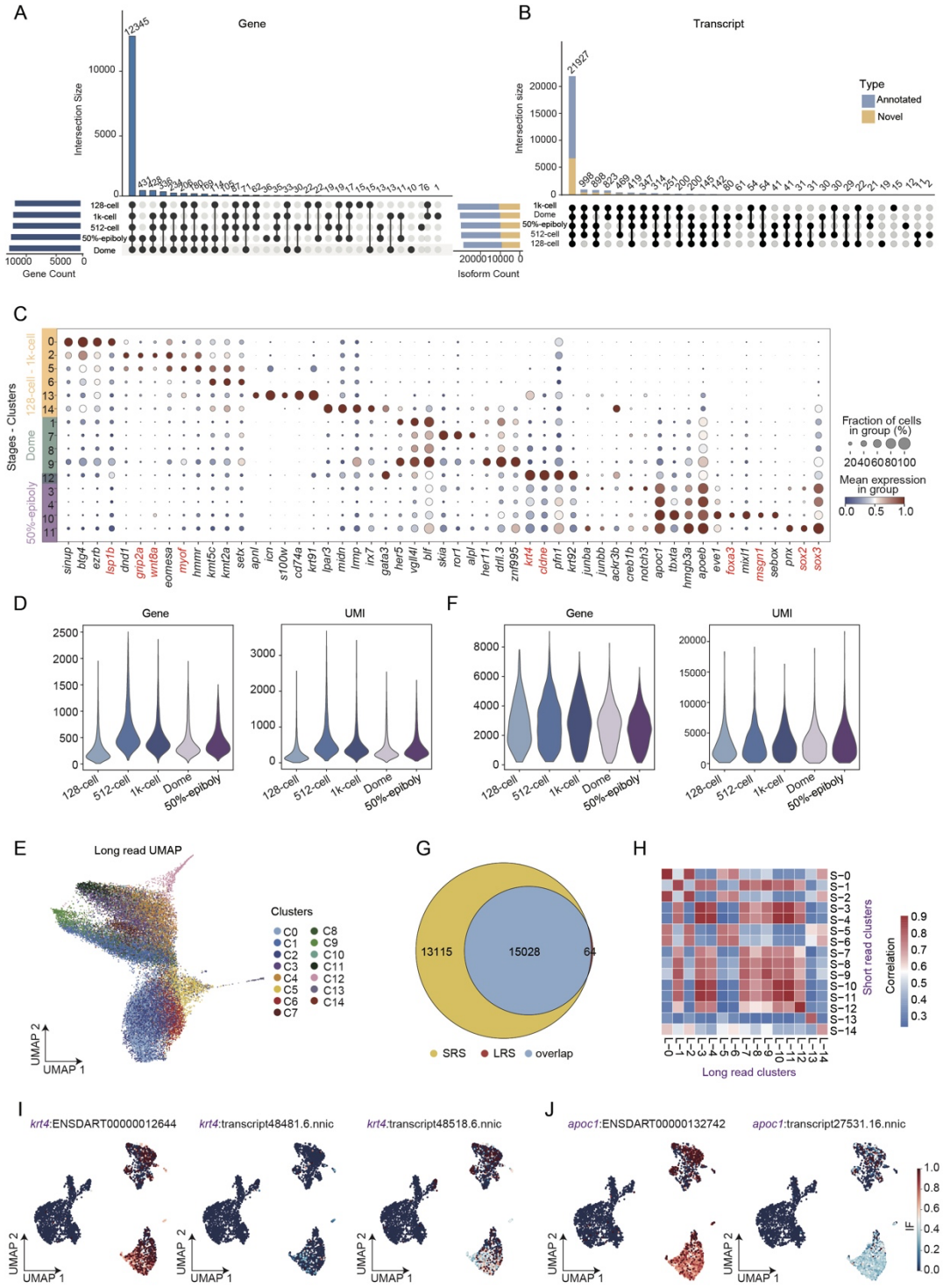
A
Gene
Intersection Size
Gene Count
128-cell
1k-cell
512-cell
50%-epiboly
Dome
B
Transcript
Intersection size
Isoform Count
Type
Annotated
Novel
C
Stages - Clusters
128-cell - 1k-cell
Dome
50%-epiboly
Fraction of cells in group (%)
Mean expression in group
D
Gene
UMI
F
Gene
UMI
E
Long read UMAP
Clusters
UMAP 1
UMAP 2
G
13115
15028
64
SRS
LRS
overlap
H
Short read clusters
Long read clusters
Correlation
I
krt4:ENSDART00000012644
krt4:transcript48481.6.nnic
krt4:transcript48518.6.nnic
J
apoc1:ENSDART00000132742
apoc1:transcript27531.16.nnic
IF

fig. S2 Full-length transcriptome diversity in zebrafish early embryogenesis.
(A) An upset plot illustrates the counts of genes detected in long read data across five developmental stages. (B) An upset plot displays the counts of transcripts detected in long read data across five developmental stages. (C) A dotted heatmap illustrates the expression of cluster specific gene across cell clusters identified from short-read data. (D) A violin plot displays the distribution of genes and UMIs across different stages in long read data prior to the application of SEACells. (E) UMAP plot of 28,468 cells from long read sequencing data prior to the application of SEACells. (F) A violin plot displays the distribution of genes and UMIs across different stages in long read data following SEACells application. (G) Venn diagram showing the overlap of genes detected by long read and short read sequencing. (H) Heatmap showing the correlation of gene expression across cell clusters between long read and short read data. (I-J) A UMAP plot illustrates the distribution of isoform fractions (IF) among the transcripts of *krt4* (I) and *apoc1* (J) in long read data.

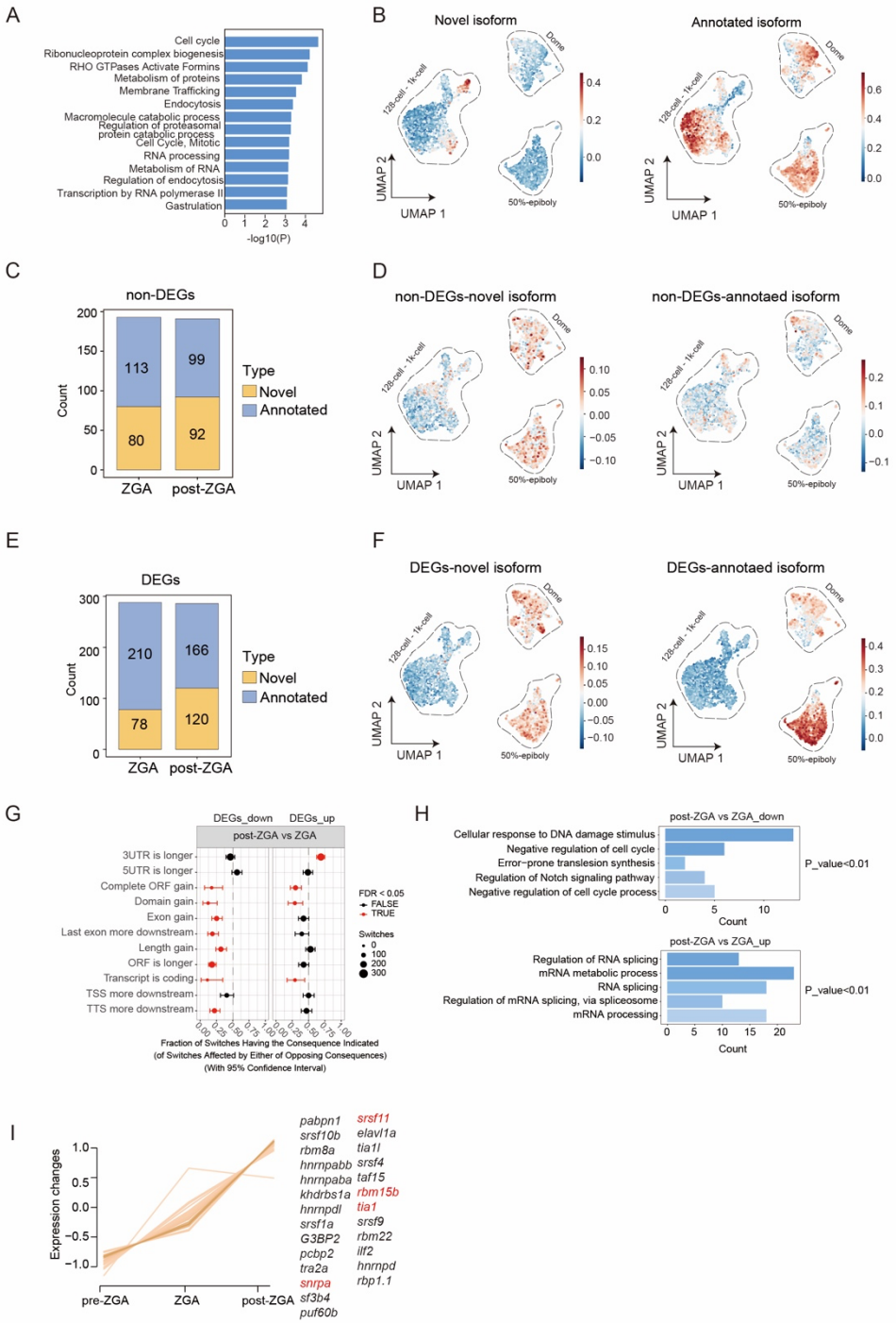
A
Cell cycle
Ribonucleoprotein complex biogenesis
RHO GTPases Activate Formins
Metabolism of proteins
Membrane Trafficking
Endocytosis
Macromolecule catabolic process
Regulation of proteasomal protein catabolic process
Cell Cycle, Mitotic
RNA processing
Metabolism of RNA
Regulation of endocytosis
Transcription by RNA polymerase II
Gastrulation
-log10(P)
B
Novel isoform
Annotated isoform
128-cell - 1k-cell
Dome
50%-epiboly
UMAP 1
UMAP 2
C
non-DEGs
Count
ZGA
post-ZGA
Type
Novel
Annotated
D
non-DEGs-novel isoform
non-DEGs-annotaed isoform
E
DEGs
F
DEGs-novel isoform
DEGs-annotaed isoform
G
DEGs_down
DEGs_up
post-ZGA vs ZGA
3UTR is longer
5UTR is longer
Complete ORF gain
Domain gain
Exon gain
Last exon more downstream
Length gain
ORF is longer
Transcript is coding
TSS more downstream
TTS more downstream
FDR < 0.05
FALSE
TRUE
Switches
Fraction of Switches Having the Consequence Indicated
(of Switches Affected by Either of Opposing Consequences)
(With 95% Confidence Interval)
H
post-ZGA vs ZGA_down
Cellular response to DNA damage stimulus
Negative regulation of cell cycle
Error−prone translesion synthesis
Regulation of Notch signaling pathway
Negative regulation of cell cycle process
P_value<0.01
Count
post-ZGA vs ZGA_up
Regulation of RNA splicing
mRNA metabolic process
RNA splicing
Regulation of mRNA splicing, via spliceosome
mRNA processing
I
Expression changes
pre-ZGA
ZGA
post-ZGA
pabpn1
srsf10b
rbm8a
hnrnpabb
hnrnpaba
khdrbs1a
hnrnpdl
srsf1a
G3BP2
pcbp2
tra2a
snrpa
sf3b4
puf60b
srsf11
elavl1a
tia1l
srsf4
taf15
rbm15b
tia1
srsf9
rbm22
ilf2
hnrnpd
rbp1.1

fig. S3 Global changes in transcriptome usage during zebrafish MZT.

(A) A bar plot illustrates the GO enrichment analysis for genes associated with isoform switching in the non-DGE categories, comparing ZGA to pre-ZGA. (B) A UMAP plot displays the expression distribution of annotated and novel transcripts involved in isoform switching within the non-DGE categories, comparing ZGA to pre-ZGA. (C) The stacked bar chart displays the counts of annotated and novel transcripts, exhibiting isoform switching in non-DEG categories between ZGA and post-ZGA stages. (D) A UMAP plot displays the expression distribution of annotated and novel isoforms involved in isoform switching within the non-DGE categories, comparing post-ZGA to ZGA. (E) The stacked bar chart displays the counts of annotated and novel transcripts, exhibiting isoform switching in DEG categories between ZGA and post-ZGA stages. (F) A UMAP plot displays the expression distribution of annotated and novel isoforms involved in isoform switching within DGE categories, comparing post-ZGA to ZGA. (G) The illustration shows the functional consequences of isoform switching in DEGs categories (downregulated and upregulated), comparing post-ZGA to ZGA. (H) The bar graphs illustrate the GO enrichment analysis for genes with isoform switching, categorized by downregulation gene (top) and upregulation gene (bottom), comparing post-ZGA to ZGA. (I) The line graph showcases the expression levels of RNA-binding proteins (RBPs) that exhibit an increasing trend with stage progression.

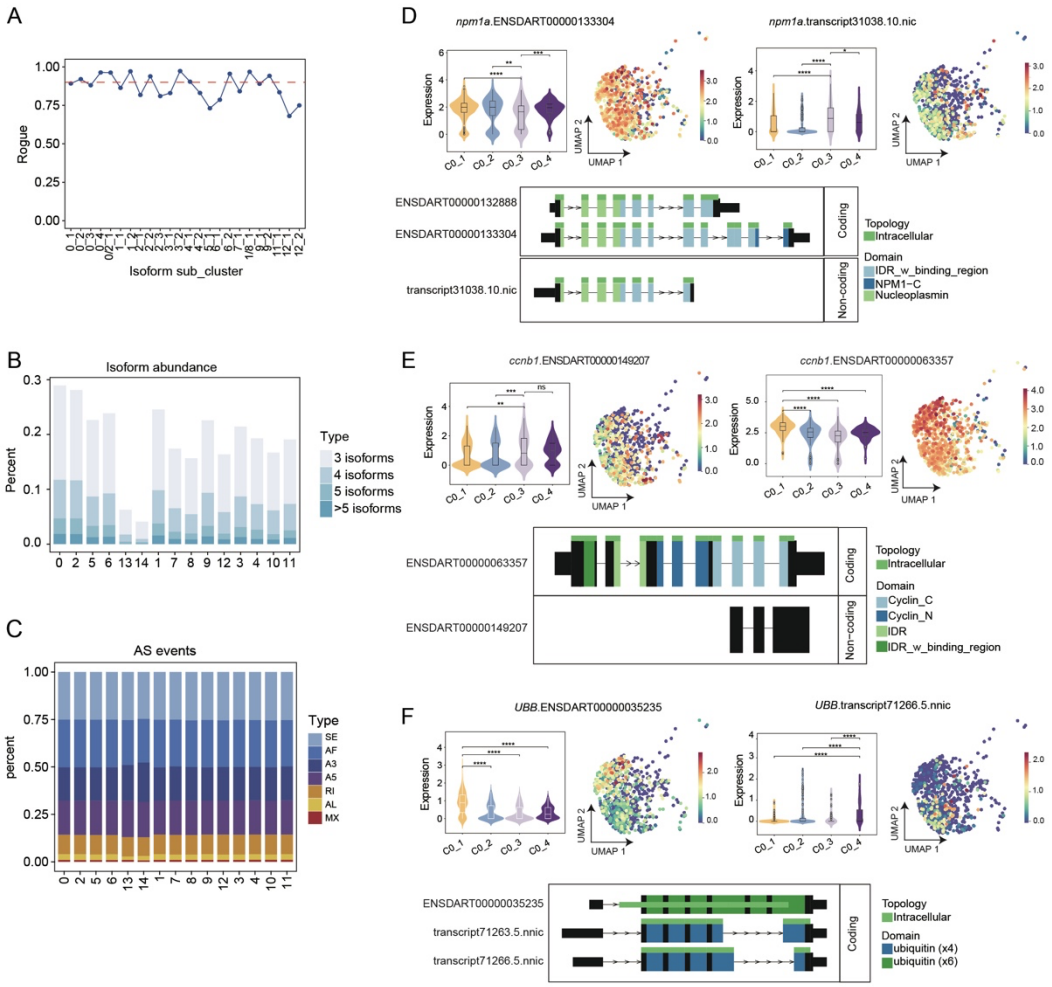
A
Rogue
Isoform sub_cluster
B
Isoform abundance
Percent
Type
3 isoforms
4 isoforms
5 isoforms
>5 isoforms
C
AS events
percent
Type
SE
AF
A3
A5
RI
AL
MX
D
npm1a.ENSDART00000133304
npm1a.transcript31038.10.nic
Expression
UMAP 1
UMAP 2
ENSDART00000132888
ENSDART00000133304
transcript31038.10.nic
Coding
Non-coding
Topology
Intracellular
Domain
IDR_w_binding_region
NPM1−C
Nucleoplasmin
E
ccnb1.ENSDART00000149207
ccnb1.ENSDART00000063357
ENSDART00000063357
ENSDART00000149207
Cyclin_C
Cyclin_N
IDR
IDR_w_binding_region
F
UBB.ENSDART00000035235
UBB.transcript71266.5.nnic
ENSDART00000035235
transcript71263.5.nnic
transcript71266.5.nnic
ubiquitin (x4)
ubiquitin (x6)

fig. S4 Isoform diversity uncovering finer subpopulations.

(A) A line graph illustrates the assessment of cellular heterogeneity within the sub-clusters of long read transcript levels using ROGUE. (B) A stacked bar chart illustrates the transcript abundance per gene form long read data. (C) A stacked bar chart illustrates the detected splicing events in long read data, with clusters delineated at the gene level. (D-F) Violin and UMAP plots (top) display the expression distribution of different transcripts of *npm1a* (D), *ccb1* (E), and *ubb* (F) within the C0 sub-clusters. The switch plot (bottom) presents the results of transcript isoform switching related to the aforementioned genes.

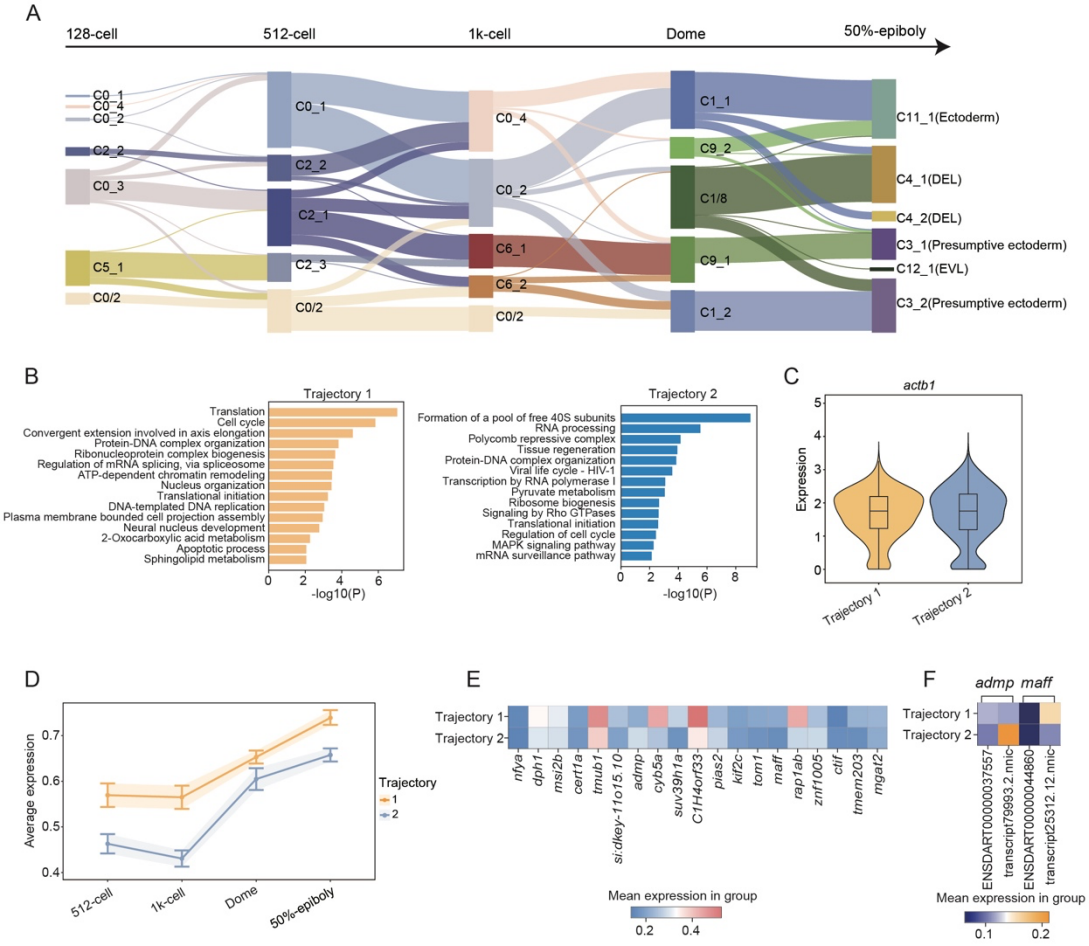
A
128-cell
512-cell
1k-cell
Dome
50%-epiboly
C0_1
C0_4
C0_2
C2_2
C0_3
C5_1
C0/2
C0_1
C2_2
C2_1
C2_3
C0/2
C0_4
C0_2
C6_1
C6_2
C0/2
C1_1
C9_2
C1/8
C9_1
C1_2
C11_1(Ectoderm)
C4_1(DEL)
C4_2(DEL)
C3_1(Presumptive ectoderm)
C12_1(EVL)
C3_2(Presumptive ectoderm)
B
Trajectory 1
Translation
Cell cycle
Convergent extension involved in axis elongation
Protein-DNA complex organization
Ribonucleoprotein complex biogenesis
Regulation of mRNA splicing, via spliceosome
ATP-dependent chromatin remodeling
Nucleus organization
Translational initiation
DNA-templated DNA replication
Plasma membrane bounded cell projection assembly
Neural nucleus development
2-Oxocarboxylic acid metabolism
Apoptotic process
Sphingolipid metabolism
-log10(P)
Trajectory 2
Formation of a pool of free 40S subunits
RNA processing
Polycomb repressive complex
Tissue regeneration
Protein-DNA complex organization
Viral life cycle - HIV-1
Transcription by RNA polymerase I
Pyruvate metabolism
Ribosome biogenesis
Signaling by Rho GTPases
Translational initiation
Regulation of cell cycle
MAPK signaling pathway
mRNA surveillance pathway
-log10(P)
C
actb1
Expression
Trajectory 1
Trajectory 2
D
Average expression
512-cell
1k-cell
Dome
50%-epiboly
Trajectory
1
2
E
Trajectory 1
Trajectory 2
nfya
dph1
msi2b
cert1a
tmub1
si:dkey-11o15.10
admp
cyb5a
suv39h1a
C1H4orf33
pias2
kif2c
tom1
maff
rap1ab
znf1005
ctif
tmem203
mgat2
Mean expression in group
F
admp
maff
Trajectory 1
Trajectory 2
ENSDART00000037557
transcript79993.2.nnic
ENSDART00000044860
transcript25312.12.nnic
Mean expression in group

fig. S5 Isoform subgrouping associated with distinct developmental trajectories.

(A) A Sankey diagram visualizes the developmental trajectory of zebrafish embryos, incorporating isoform sub-clusters identified in long read data. (B) A bar plot illustrates the GO enrichment of genes with differential exon usage between two trajectories with $p < 0.05$. (C) Violin plot showing the expression profile of the housekeeping gene actb1 across two developmental trajectories. (D) Expression dynamics of RBP genes identified via RNA-binding protein (RBP) motif prediction across two trajectories at different time points. (E) Heatmap of gene expression with significant differential exon usage between two trajectories. $P < 0.05$, $\Delta PSI > 0.3$. (F) Heatmap displaying the expression patterns of distinct transcripts for *admp* and *maff* across two lineages.