BMC Biology

RESEARCH ARTICLE

Open Access

A combination of transcriptomics and epigenomics identifies genes and regulatory elements involved in embryonic tail development in the mouse

Yong-Xuan Chen[1,2†], Xiu-Ping Zhang[1†], David N. Cooper[3], Dong-Dong Wu[1,4*] and Wan-Dong Bao[1*]

Abstract

Background The post-anal tail is a common physical feature of vertebrates including mammals. Although it exhibits rich phenotypic diversity, its development has been evolutionarily conserved as early as the embryonic period. Genes participating in embryonic tail morphogenesis have hitherto been widely explored on the basis of experimental discovery, whereas the associated *cis*-regulatory elements (CREs) have not yet been systematically investigated for vertebrate/mammalian tail development.

Results Here, utilizing high-throughput sequencing schemes pioneered in mice, we profiled the dynamic transcriptome and CREs marked by active histone modifications during embryonic tail morphogenesis. Temporal and spatial disparity analyses revealed the genes specific to tail development and their putative CREs, which facilitated the identification of novel molecular expression features and potential regulatory influence of non-coding loci including long non-coding RNA (lncRNA) genes and CREs. Moreover, these identified sets of multi-omics data supply genetic clues for understanding the regulatory effects of relevant signaling pathways (such as Fgf, Wnt) dominating embryonic tail morphogenesis.

Conclusions Our work brings new insights and provides exploitable fundamental datasets for the elucidation of the complex genetic mechanisms responsible for the formation of the vertebrate/mammalian tail.

Keywords Vertebrate/mammal/mouse, Embryonic tail morphogenesis, Transcriptome, Epigenomic histone modification, *Cis*-regulatory element

†Yong-Xuan Chen and Xiu-Ping Zhang contributed equally to this work.

*Correspondence:
Dong-Dong Wu
wudongdong@mail.kiz.ac.cn
Wan-Dong Bao
baowandong@mail.kiz.ac.cn
[1] State Key Laboratory of Genetic Resources and Evolution, Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650201, China
[2] Kunming College of Life Science, University of the Chinese Academy of Sciences, Kunming 650204, China
[3] Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, Wales, UK
[4] Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

Background

The post-anal tail is a prominent physiological structure in chordates and is associated with a rich diversity of traits in vertebrates, being involved in roles that are fundamental to locomotion and survival, including swimming [1], balance [2, 3], attack [4], grabbing [5, 6], body support [7], and autotomy for escape [8]. Although some adult vertebrates (such as human [9, 10] and anuran [11]) appear not to have an obvious tail, they still possess an embryonic or larval tail that is degraded during development. As early as the embryonic period, tail morphogenesis proceeds and is driven by a posterior extension of the embryo known as the tail bud [12, 13]. The axial progenitors in the tail bud, descended directly from a subset of those contributing to the trunk region, generate the structural components of the embryonic tail primarily including the paraxial mesoderm, neural tube, and notochord [14–17]. Not simply a developmental elongation of the anterior/posterior body axis, embryonic tail morphogenesis is regulated by specific molecular mechanisms that may be distinguished from those controlling trunk elongation [18–22].

To date, numerous genes have been identified as being important for tail morphogenesis, which is largely based on the discovery or construction of mutant animals with aberrant tail phenotypes. These genes generally act by participating in the fundamental biological processes dominating the posterior extension and specification. For instance, *Wnt3a*, a signal ligand gene essential for embryonic caudal development [23, 24], drives a conserved regulatory network for the cell fate choices of neuromesodermal progenitors [25]. In mouse, the functional disruption of *Wnt3a* or its responsive transcription factor (TF) genes, e.g., *Tbx6*, *Tcf1*, and *Lef1*, causes a severe deficit in the paraxial mesoderm and hence an aberration/absence of the tail structure [26–28]. The Fgf signaling pathway is essential for the maintenance of neuromesodermal progenitors [29] and cell migration in the presomitic mesoderm [30, 31] to ensure axis extension. Mouse embryos with a disrupted *Fgf3* or *Fgf4*/*Fgf8* compound deletion exhibit seriously abnormal elongation and shape of tail [32, 33]. Moreover, the genes crucial for somitogenesis, a basic function of the paraxial mesoderm in the tail bud, are needed to achieve normal morphogenesis of the embryonic tail [34–38].

High-throughput sequencing is widely used to reveal the gene expression dynamics associated with the embryological developmental processes [39–42], which could provide raw data for the mining of the marker or active genes involved but has still not been systematically designed for the exploration of embryonic tail morphogenesis in vertebrates. Furthermore, CREs represent vital structural means for the expression modulation of gene loci, and also a general target of variations leading to changes of gene expression and hence the evolution of traits [43, 44]. With respect to the study of the genetic mechanisms underlying the structural generation or even the morphological diversity of vertebrate tails, one fundamental task may be to interrogate the CRE atlases responsible for the modulation of genes involved in the development of the embryonic tail.

In the present study, taking advantage of the general character of tail bud-restricted expression of genes participating in tail morphogenesis, we implemented a temporal series of transcriptome sequencing and epigenome sequencing of active histone modifications focused on murine embryonic tail tip tissues. Multiple omics data are indicative of the gene expression and chromatin state dynamics that correlate with embryonic tail development, and help to reveal the novel molecular features and putative regulatory mechanisms underlying this basic developmental process that is strongly conserved in vertebrates.

Results

Transcriptome dynamics occurring in mouse embryonic tail tip tissue

To obtain the gene expression dynamics relevant to embryonic tail morphogenesis, we designed a systematic sampling scheme of mouse embryonic tail tip tissues for bulk transcriptome sequencing. Specifically, during the period of embryonic tail morphogenesis when the caudal somite count continuously increased [45], the tail terminal region (≈ 1 mm in anterior/posterior (AP) axis, termed T1) was sampled respectively at 6 consecutive stages E10.5 ~ E15.5, and so was its proximally contiguous region (also ≈ 1 mm, termed T2) as a control (Fig. 1a). Structurally, T1 contained tail bud and several newly formed somites, e.g., 2 ~ 3 somites at E10.5 based upon our observation, and would possess a higher somite count within itself as the developmental stage advances, when the tail bud area shrunk progressively, whereas T2 simply corresponded to the somite regions formed earlier. Principal component analysis (PCA) on the transcriptome profiles revealed that T1 and T2 samples exhibited similar and parallel distribution trajectories along the timeline (Fig. 1b), perhaps owing to their adjacent sampling positions from the same tissue origin of embryonic tail. Meanwhile, the samples of E10.5 ~ E12.5, especially those from T1, showed an obvious clustering trend by sampling position (Fig. 1b, c and Additional file 1: Fig. S1), indicating the high similarity of T1 transcriptome profiles and their collective distinction from those of T2 during this earlier phase. Yet the divergence between T1 and T2 transcriptome profiles appeared to become less significant during E13.5 ~ E15.5, given the

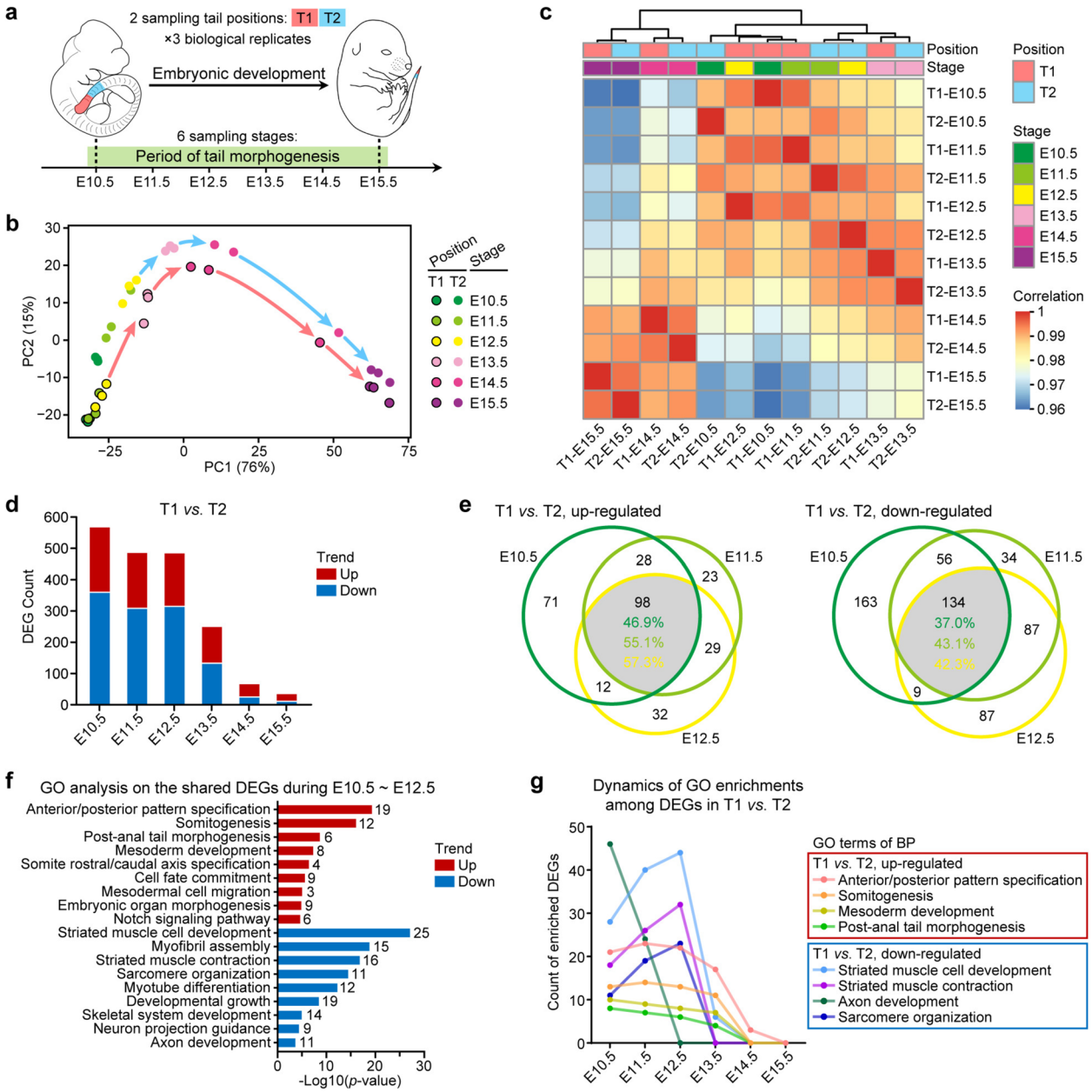

Fig. 1 Overview of the dynamic transcriptomic profiles of tail tip tissue during mouse embryonic development. **a** Summary of the sampling strategy underlying transcriptome sequencing. Two positions, viz. terminal and proximally adjacent regions (≈ 1 mm length for both) termed T1 and T2 respectively, of embryonic tail tip tissue were sampled at six consecutive stages (E10.5 ~ E15.5) during tail morphogenesis, with three biological replicates employed for each sampling. **b** PCA analysis of all the transcriptomic samples. The rlogTransformation function was used to normalize the count data for the PCA plot. The colored arrows indicate the trajectory direction of samples with time in the drawing, respectively for T1 and T2. **c** Correlation analysis on the transcriptomic profiles of all 12 sampling types (2 positions × 6 stages). The mean value of three biological replicates was calculated and the correlation heatmap was drawn with the value of log2(mean + 1). (d) Statistics of the DEG count in T1 *vs.* T2 for each stage. Log2(fold change) > 2, FDR < 0.05; zFPKM > −3 at least in one of T1 and T2 within each comparison. **e** Intersection analysis on the DEGs of E10.5 ~ E12.5 stages in T1 *vs.* T2, respectively for the up- and downregulated sets. The DEG counts in different parts of the Venn diagram are shown. For the E10.5 ~ E12.5-shared DEGs (grey part), their ratio in the set of each stage (ring) is indicated, in the accordant color with the set. **f** GO enrichment analysis on the E10.5 ~ E12.5-shared DEGs in T1 *vs.* T2. The DEG counts enriched in BP terms are indicated. $P < 0.001$. **g** Statistics of the dynamic DEG count enriched in the representative BP terms during embryonic tail morphogenesis after GO analysis on the DEGs in T1 *vs.* T2 for each stage. Four terms enriched in **f** are shown here, respectively for the up- and downregulated sets

uniform clustering by sampling stage for the samples from this later phase (Fig. 1c and Additional file 1: Fig. S1).

According to the results of the differentially expressed gene (DEG) analysis in T1 *vs*. T2 (log2(fold change) > 2, FDR < 0.05; zFPKM > −3 at least in one of T1 and T2), the count of DEGs between T1 and T2 transcriptome profiles indeed decreased progressively throughout embryonic tail morphogenesis, with a relatively sharp decline after E12.5 (Fig. 1d). Thus, we then focused upon the DEGs during the earlier phase to reveal the gene expression characteristics respectively in T1 and T2 regions. Corresponding to the developmental continuity, a large proportion of DEGs were shared among the up/downregulated sets of E10.5 ~ E12.5 in T1 *vs*. T2 (Fig. 1e) such that numerous DEGs were identified to be continuously enriched in T1 (upregulated, $n = 98$) or T2 (downregulated, $n = 134$). Gene Ontology (GO) enrichment analysis revealed that these shared DEGs were significantly associated with biological processes (BPs) including AP pattern specification, somitogenesis, cell fate commitment, mesoderm development, and post-anal tail morphogenesis for T1, in contrast to the BPs relevant to musculature and nervous system development or function, e.g., striated muscle cell development, sarcomere organization, myotube differentiation, striated muscle contraction, neuron projection guidance, and axon development, for T2 ($p < 0.001$, Fig. 1f). This distinction between the transcriptome profiles of T1 and T2 was consistent with our expectation based on the different tissue components of the two sampling positions, i.e., T1 contained tail bud that drove embryonic tail morphogenesis, and overall had a lower degree of differentiation as compared with T2. Then, by reason of the processive tail bud diminution and tissue maturation, T1 gradually acquired more gene expression associated with the BPs responsible for tail tissue differentiation and functionalization, but less of that involved in the BPs dominating embryonic tail morphogenesis (Additional file 1: Fig. S2a, b). This could help to explain the decreasing count of DEGs relating to both BP groups, between T1 and T2 with the passage of developmental time (Fig. 1g). In brief, we obtained a systematic set of transcriptome profiles that could reflect the gene expression dynamics involved in mouse embryonic tail morphogenesis.

Weighted gene coexpression network analysis (WGCNA) to acquire an embryonic tail morphogenesis-specific gene module

Based on the general expression feature of genes responsible for embryonic tail morphogenesis, i.e., specific expression enriched in the tail bud area, it might be easy to infer that these genes should collectively possess a relatively stable low level of expression in T2, but a high initial expression level (IEL, herein referring to the expression level at the earliest stage E10.5) in T1, which would then continually decline over time and eventually approach the T2 expression level. Indeed, this particular expression pattern was exhibited by representative key genes modulating embryonic tail morphogenesis, e.g., *T* [46], *Wnt3a* [26], *Hes7* [37], and *Fgf3* [32] (Fig. 2a), with relatively high expressions at the earlier stages and usually an apparent downregulation from E12.5 to E14.5 in T1. Following this pattern, we next tried to holistically obtain the set of genes potentially playing roles in embryonic tail morphogenesis, by performing a WGCNA analysis [47] on the transcriptome profiles.

After coexpressed gene module calculation, 7 modules were finally identified (Fig. 2b and Additional file 2: Table S1), each of which represented genes that shared highly similar expression patterns in response to temporal changes during embryonic tail morphogenesis. The blue module showed the highest correlation with T1 ($r = 0.61$, $p = 9e-05$) and was revealed to be significantly associated with BPs such as somitogenesis, AP pattern specification, and post-anal tail morphogenesis (Fig. 2c). Importantly, the coexpressed genes dominating the blue module exhibited a fitted expression pattern that

(See figure on next page.)

Fig. 2 WGCNA analysis on the transcriptome profiles to yield an embryonic tail morphogenesis-specific gene module. **a** The expression profiles of the known essential genes responsible for mouse embryonic tail morphogenesis. These representative genes collectively possess relatively low and stable expression in T2 but a high IEL in T1 that progressively declines over time and eventually approaches the T2 expression level. $n = 3$ for each sampling type (Additional file 8: Table S41). Bars denote standard deviations. **b** Cluster dendrogram indicates the modules based on topological overlap of coexpressed genes in the WGCNA analysis on the DEG sets of T1 *vs*. T2 transcriptome profiles throughout embryonic tail morphogenesis. The colored bars below provide information on module membership. **c** Correlation analysis between the 7 coexpression modules and sampling positions or stages. Each square contains the corresponding correlation value and *p*-value (bracketed). The representative GO terms of BP enriched in the gene set are indicated on the right for each module. **d** The expression trajectory of the blue module for T1 and T2 throughout embryonic tail morphogenesis, shown by means of the module eigengene that represents the gene expression profile of this module. $n = 3$ for each sampling type (Additional file 8: Table S42). Fit curves are applied by locally weighted scatterplot smoothing, and shading represents the confidence interval (95% confidence). **e** Network of the top 30 hub genes in the blue module. The hub genes were obtained in rank order of their MCC scores calculated by cytoHubba in Cytoscape

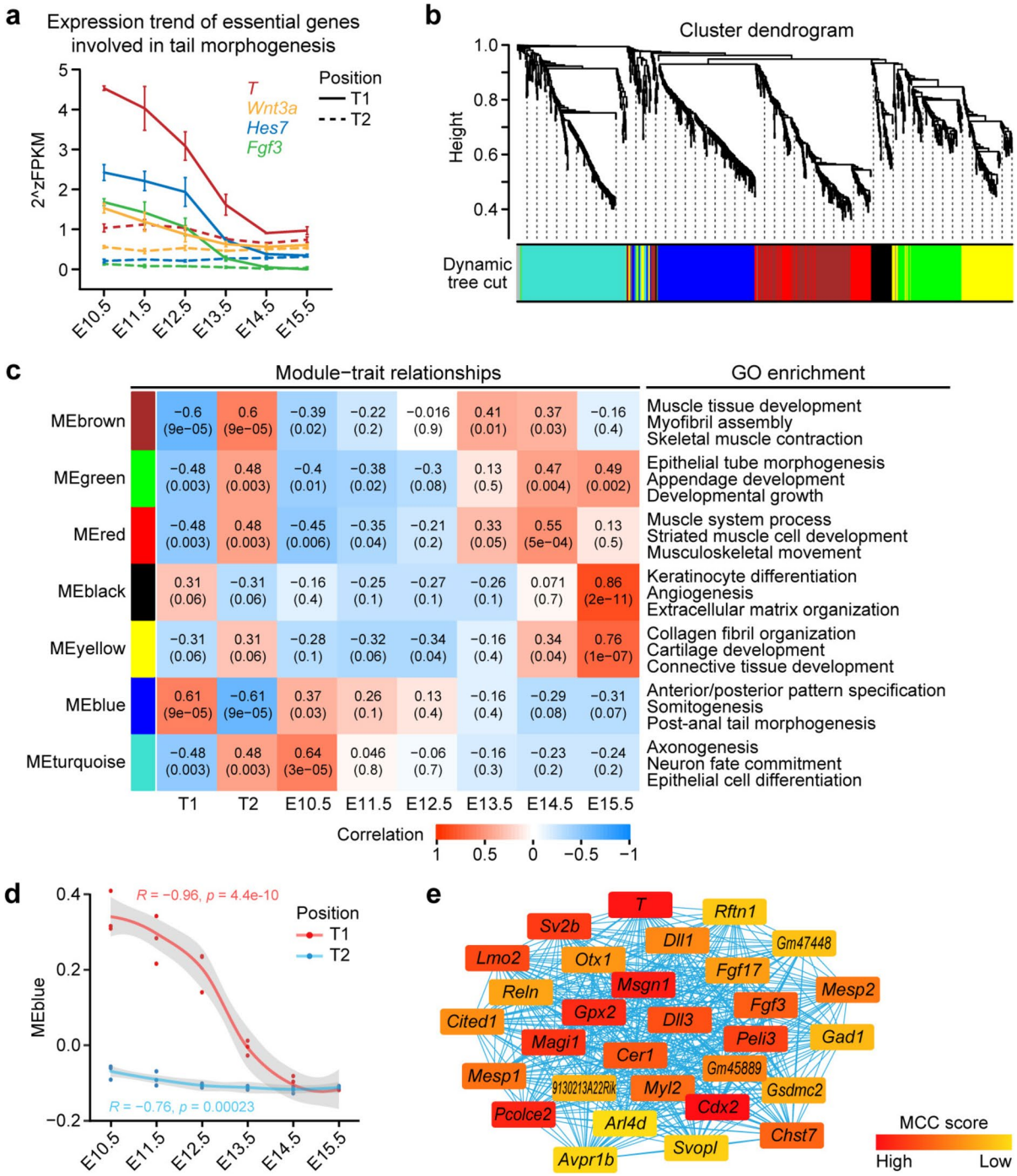

Fig. 2 (See legend on previous page.)

tallied perfectly with the type we had anticipated above (Fig. 2d). Moreover, numerous known important genes for embryonic tail morphogenesis were found among the top 30 hub genes of the blue module, including *T* [46], *Fgf3* [32, 48], *Msng1* [49], *Dll1* [50], *Dll3* [34], *Cdx2* [51, 52], and *Mesp2* [35] (Fig. 2e). Of them, *T* is known as a pivotal T-box TF gene that is autoregulated in a loop with the Wnt signaling pathway essential for embryonic caudal development [53–55], and the latter operates via the downstream TF effector genes such as *Tbx6*, *Msng1*, and certain *Cdx* genes, as reviewed in [25]. *Dll1* and *Dll3* encode ligands for activating Notch signaling pathway,

which is crucial for somite formation in embryonic tail and involves the functional participation of *Mesp2* gene [34, 56–58]. *Fgf3* is known as a key signal factor gene indispensable for normal AP axis extension including the tail region [32, 48]. All of these characteristics helped to identify the blue module as the one specific for embryonic tail morphogenesis.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis revealed that genes in the blue module were significantly correlated with regulation of the actin cytoskeleton (mmu04810, $p = 1.04e-03$) which closely relates to cell motility including migration that is abundant in the presomitic mesoderm (PSM) [30, 31, 59], and also with the signaling pathways respectively driven by Rap1 (mmu04015, $p = 6.16e-04$), Ras (mmu04014, $p = 2.16e-02$), calcium (mmu04020, $p = 3.25e-04$), and phosphoinositide 3-kinase (PI3K)-Akt (mmu04151, $p = 3.40e-03$). According to the corresponding KEGG pathway maps [60], these enriched signaling pathways are known to participate in the regulation of cell proliferation, adhesion, migration, gene activity, and cytoskeletal remodeling and could all be activated by Fgf signal transduction, suggesting a prominent role for the Fgf signaling pathway in the modulation of the developmental and cellular BPs related to embryonic tail morphogenesis. In accordance with this postulate, many Fgf signal factor genes were noted in the blue module, e.g., *Fgf3*, *Fgf4*, *Fgf15*, *Fgf17*, and *Fgf18*, each of which has been found to be expressed specifically in PSM [33, 48, 61, 62]; additionally, *Fgf3* and *Fgf4* have also been shown to be required for normal axial elongation of the embryonic tail [29, 33]. Besides, as is also indicated in those KEGG pathway maps [60], all of the enriched signaling pathways mentioned above could be regulated by the cyclic adenosine monophosphate (cAMP) signaling pathway (mmu04024), which was also significantly enriched in the blue module ($p = 1.77e-02$). Combined with the previous reports that mesodermal inhibition of the cAMP responsive element-binding proteins (CREBs), a TF family acting downstream of the cAMP pathway [63, 64] as their names describe, caused defective segmentation and somite polarity [65, 66], this signaling pathway could be a potentially important regulator involved in embryonic tail development.

Novel expression characteristics revealed by the blue module for embryonic tail morphogenesis

To discover novel genes marking embryonic tail morphogenesis, we then performed a detailed component analysis on the gene membership of the blue module. The vast majority of modular genes pertained to the positive correlation category (179/213), which we focused upon in the present study and referred to specifically as the blue module hereafter. Of the modular protein-coding and lncRNA genes, those within the top 100 gene set of positive correlation with high IEL (ranked by the *z*-normalized fragments per kilobase per million mapped reads (zFPKM) value at E10.5) in T1 were investigated preferentially (85/144 and 14/30, respectively, Fig. 3a). According to the relevant literature, more than half of these highly expressed protein-coding genes (51.8%, $n = 44$) are known to be expressed specifically in the tail bud area of mouse embryos (Additional file 2: Table S2), of which nearly half (45.5%, $n = 20$) have further been shown to play essential roles in the BPs governing tail morphogenesis by their mutant phenotypic data, primarily including the ligand and regulator/effector genes of Wnt (*Wnt3a* [23, 26], *Wnt5a* [67], *T* [46, 68], *Tbx6* [69], *Msgn1* [49], *Cdx1* [70], *Cdx2* [52], *Cdx4* [71], and *Sp5* [72]), Fgf (*Fgf3* [32, 48] and *Fgf4* [29, 33]), and Notch (*Dll1* [50], *Dll3* [34], *Hes7* [37], *Lfng* [38], *Ripply2* [36], and *Mesp2* [35]) signaling pathways (Fig. 3a). The enrichment of these known tail bud-expressed genes further illustrates the specificity of the blue module for embryonic tail morphogenesis and implies a high potential for other coexpressed genes within the module to participate in this conserved developmental process.

Focusing on the unreported group of protein-coding genes for tail bud expression ($n = 41$, Fig. 3a and Additional file 2: Table S3), in situ hybridization experiments were performed to detect their expression distribution in E10.5 embryos, preferentially for those top ones possessing high IELs in T1 such as *Arl4d*, *Fabp7*, *Vwa2*, *Peli3*, *Magi1*, and *Rftn1* (IEL-T1 = 1.279, 0.960, 1.034, 0.713, 0.609, and 0.416 (zFPKM value), respectively). As expected, these genes displayed, without exception, specific expression in the tail bud area with varying distribution features (Fig. 3b). Specifically, *Arl4d*, encoding an adenosine diphosphate (ADP)-ribosylation factor (ARF)-like protein known to act so as to promote cell migration activity [73, 74], showed significant expression restricted to the posterior PSM, where abundant cell movements occur at the highest level along the AP axis in PSM [30, 31, 75]. Additional specific expression was also detected in the anterior-most region of PSM, implying the involvement of *Arl4d* in the formation of somites. The fatty acid-binding protein gene *Fabp7* exhibited mild but nevertheless identifiable expression that was restricted to the posterior PSM region in the vicinity of the axial progenitor zone, suggesting a potential expression label for the nascent mesoderm-fated progenies of this particular location. *Vwa2*, whose protein product is an extracellular matrix component able to mediate cell–matrix adhesion [76, 77], was detected specifically in PSM with fading expression in the anterior-most portion, which might emphasize the special extracellular matrix composition

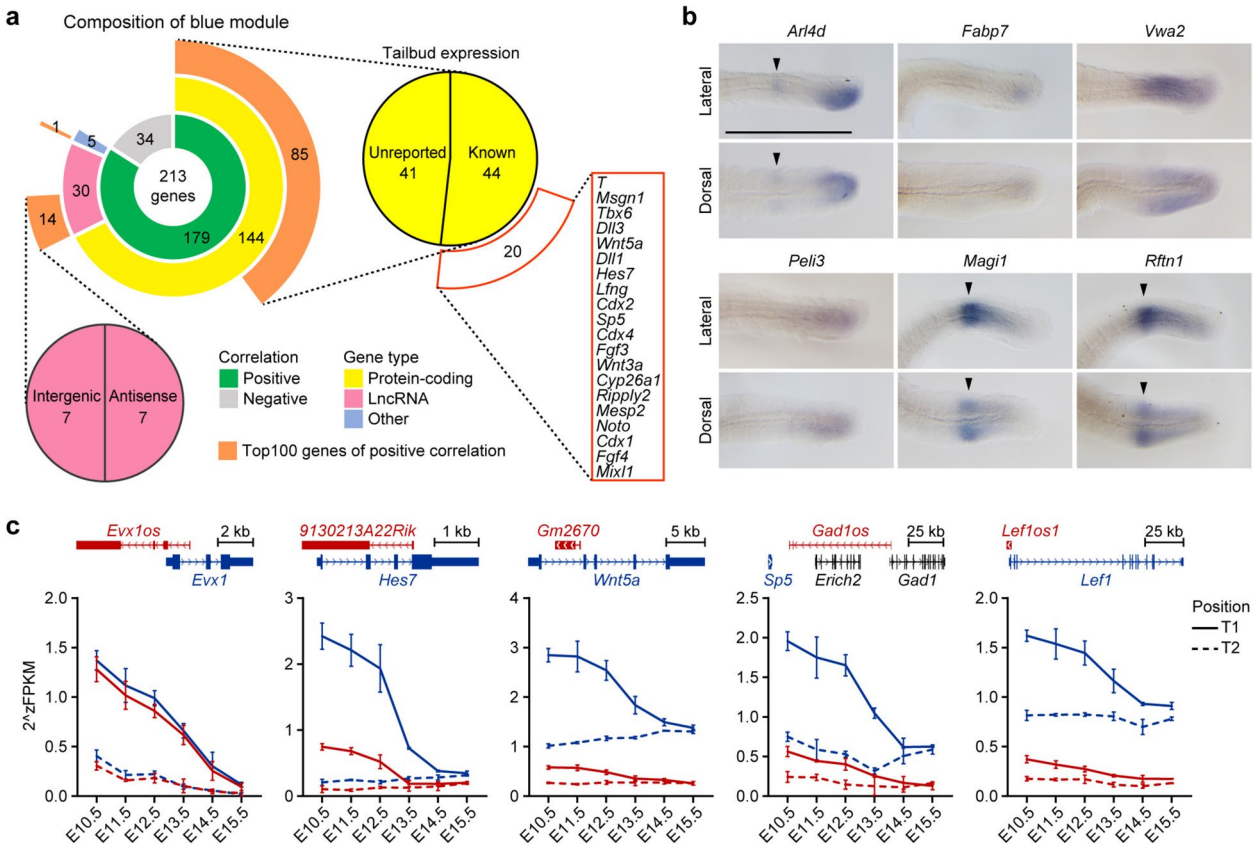

Fig. 3 Blue module reveals novel molecular expression characteristics involved in embryonic tail morphogenesis. **a** Overview of the gene components of the blue module. The gene members that were positively correlated with the module in terms of their expression pattern were focused upon here, of which the top 100 genes highly expressed (ranked by the zFPKM value at E10.5 in T1) were analyzed preferentially, primarily for the protein-coding genes ($n=85$) and lncRNA genes ($n=14$). The gene counts in different partitions are indicated. The red box represents the protein-coding genes with known tailbud expression that were further proved indispensable for the BPs participating in embryonic tail morphogenesis by their mutant data. **b** In situ hybridization experiments at E10.5 verify the specific tailbud expression of the protein-coding genes in the "unreported" group of **a**. Each gene was verified in at least four mouse embryos. The arrowheads indicate the location of the forming somite. Scale bar = 1 mm. **c** Coordinate expression between the lncRNA genes in the "antisense" group of **a** and their adjacent protein-coding genes in the sense strand during embryonic tail morphogenesis. The antisense lncRNA genes (red) that flank the known important protein-coding genes (blue) responsible for the fundamental BPs propelling embryonic tail morphogenesis are shown. $n=3$ for each sampling type (Additional file 8: Table S43). Bars denote standard deviations

within the PSM region prior to segmentation. *Peli3*, encoding an E3 ubiquitin protein ligase involved in the regulation of signal transduction [78–80], showed moderate expression that was primarily concentrated in the posterior PSM. The membrane-associated guanylate kinase gene, *Magi1*, encoding a scaffolding protein localized at cell junctions to help ensure signal transduction [81–83], exhibited robust expression at the sites of somite formation along with a relatively low level of expression in the more posterior region of PSM. *Rftn1*, a lipid raft linker gene whose protein product is necessary for the integrity of lipid raft in the plasma membrane and hence modulates signaling processes [84–86], exhibited an expression pattern resembling that of *Magi1* in the tail bud area, with relatively more visible detection in the PSM overall. The particular distributions of *Magi1* and *Rftn1* expression might be indicative of a special molecular environment respectively in the cortex and plasma membrane of PSM cells that guarantees signal transduction, perhaps specifically for somite formation. In any case, these detected genes represent novel markers of tail bud structures, thereby confirming the availability of the blue module for mining genes that participate in embryonic tail morphogenesis.

Meanwhile, the blue module suggests the involvement of lncRNA genes with potential regulatory roles in the process of tail morphogenesis. LncRNAs are a category of RNA transcripts with a length of more than

200 nucleotides but no capacity to encode proteins [87, 88], which have been shown to regulate the expression of neighboring genes in *cis* or remote genes in *trans* [89, 90]. With regard to *cis* regulation, it is inferred, based upon experimental validation, that protein-coding genes may often be fine-tuned by their juxtaposed lncRNA loci, especially the divergent ones that are transcribed from the antisense strand and positioned head-to-head relative to the neighboring protein-coding genes [90, 91]. Of these highly expressed lncRNA genes for preferential analyses in the blue module, the antisense class accounted for half of the total (7/14, Fig. 3a and Additional file 2: Table S4), of which most ($n=5$), including *Evx1os*, *9130213A22Rik*, *Gm2670*, *Gad1os*, and *Lef1os1*, were found to be located close to the known important signal ligand or TF genes responsible for the fundamental BPs relevant to tail morphogenesis, i.e., the protein-coding genes *Evx1* [92], *Hes7* [37], *Wnt5a* [67], *Sp5* [72], and *Lef1* [28], respectively (Fig. 3c). In line with the prevailing view that antisense lncRNAs tend to be tightly co-regulated with adjacent protein-coding genes on account of their apparently frequent role in facilitating the expression of the latter [91], each of these antisense/sense gene pairs, without exception, exhibited an overall highly coordinated expression trend during tail morphogenesis, with a relatively low expression level for the lncRNA genes in most cases (Fig. 3c). Moreover, as expected, nearly all the adjacent protein-coding genes were also allocated to the blue module (only missing *Lef1*), emphasizing the coexpression state of each antisense/sense gene pair.

Evx1os showed a high lncRNA transcript level measured by zFPKM value, which coincided strongly with that of *Evx1* (Fig. 3c), and represented a typical case of divergent lncRNAs to illustrate the biological functions of this subclass based on regulating adjacent protein-coding genes [91]. The other four antisense lncRNAs might all be classified as divergent, although they did not fully conform to the subclass as defined previously [91]. For instance, *Gm2670*, being located in the first intron of *Wnt5a* gene, could be regarded as a short divergent locus whose 3'-end did not extend beyond the body region of its neighboring protein-coding gene on the sense strand (Fig. 3c). Also, the exact divergent protein-coding gene of the *Gad1os* locus is *Gad1*, but intriguingly this 73.4-kb-long antisense locus also has two additional neighboring protein-coding genes on the sense strand, i.e., *Erich2* and *Sp5*, which are respectively located in an intron and downstream of *Gad1os* locus (Fig. 3c). Since *Gad1*, *Erich2*, and *Sp5* together exhibit prominent coexpression with *Gad1os* and were included in the blue module, *Gad1os* might appear to be a divergent lncRNA against the long genomic domain comprising the three protein-coding genes and thereby might be capable of promoting the gene expression of this entire domain. These cases of non-strictly defined divergent lncRNAs should concur with the proposition that lncRNA-mediated *cis*-regulation of adjacent protein-coding gene transcription might be a prevalent mechanism without limitation to the divergent lncRNA biotype [91]. Taken together, we may speculate that at least a subset of the lncRNA genes in the blue module, especially the antisense ones, might play potential regulatory roles in embryonic tail morphogenesis via *cis*-regulation of the expression of neighboring protein-coding genes responsible for relevant fundamental BPs.

Significant genome-wide decline of active histone modification levels in embryonic tail tip

For a more comprehensive understanding of the genetic modulation of embryonic tail morphogenesis in combination with the associated epigenomic dynamics indicative of CRE activities, we also performed a series of Cleavage Under Targets and Tagmentation-sequencing (CUT&Tag-seq) to map the dynamic genome-wide profiles of histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 acetylation (H3K27ac), which are known to serve as epigenetic marks of active promoters and enhancers respectively [93–96], in the T1 area of the embryonic tail tip with the passage of developmental time. For either modification, T1 was sampled selectively at the 3 stages of E10.5, E12.5, and E14.5 and was then sequenced (see Methods for more details).

After peak calling and biological replicate intersection, we obtained the genomic atlases of active epigenetic domains at each stage. Overall, the genome-wide levels of both H3K4me3 and H3K27ac dropped significantly in T1 during embryonic tail morphogenesis, with a very sharp decline (77.1 and 68.9% of domain count, 39.7 and 71.1% of genomic coverage, respectively) from E10.5 to E12.5 (Fig. 4a, b). Specifically, most of the H3K4me3 domains of E10.5 pertained to the non-transcriptional start site (TSS)-containing class, whose abrupt reduction was then principally responsible for the marked drop of H3K4me3 domain count after E10.5, rendering the TSS-containing ones as the main class at E12.5 and E14.5 (Fig. 4a). The H3K4me3 modification at E10.5 was not primarily confined to the proximal regions (≤ 3 kb from the TSS) as indicated at E12.5 and E14.5, but rather had more domains in the distal regions (>3 kb from the TSS) (Fig. 4a), pointing to a relatively special chromatin modification state with a wider range of H3K4me3-modified genetic locations at this early stage of embryonic tail morphogenesis. In contrast with H3K4me3, the reduction of H3K27ac domains did not exhibit any obvious regional preference, and the non-TSS-containing class (mainly the distal ones) were always in a majority

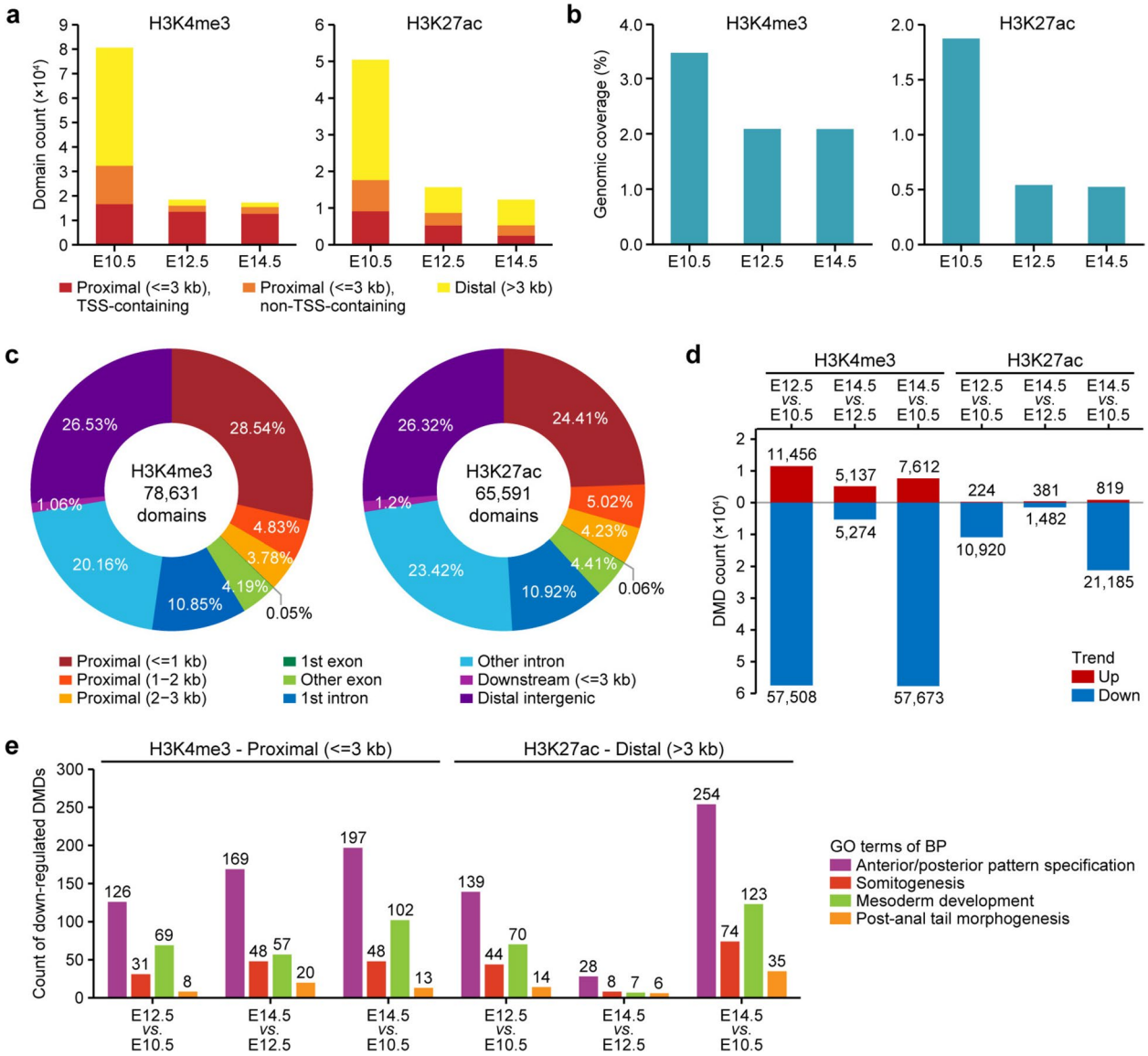

Fig. 4 Decline in genomic H3K4me3 and H3K27ac modification of T1 during embryonic tail morphogenesis. **a** Statistics of the count of H3K4me3 and H3K27ac domains identified at each sampling stage. **b** Statistics of the genome-wide coverage rate of H3K4me3 and H3K27ac domains identified at each sampling stage. **c** Total identified domains of H3K4me3 and H3K27ac after intertemporal merging and genomic annotation. The proportions of different genomic classes are indicated. **d** DMD analysis between every two stages respectively for H3K4me3 and H3K27ac. Fold change > 2, FDR < 0.05. The DMD counts of each comparison are indicated. **e** Statistics of the downregulated DMDs adjacent to the genes participating in the fundamental BPs relevant to embryonic tail morphogenesis, respectively in the proximal region for H3K4me3 and in the distal region for H3K27ac, based on the results of the analysis in **d**. The BP terms shown here are identical to those in Fig. 1g (upregulated in T1 *vs.* T2). Counts of downregulated adjacent DMDs are indicated for each BP term. Proximal and distal regions are delimited by the distance flanking TSS as presented in brackets for **a**, **c** and **e**, and so is the downstream region by the bracketed distance to the last exon for **c**

(Fig. 4a), which concurred with the characteristic of H3K27ac for marking active enhancers (typically distant from the TSSs) [95, 96]. Additionally, for both modifications, the TSS-containing class tended to have a larger width than the non-TSS-containing class irrespective of the developmental stage (Additional file 1: Fig. S3), perhaps suggesting a relatively richer distribution of *cis*-regulatory activities in the regions around TSSs (typically considered as promoters).

By integrating the biological replicates of all stages, a total of 78,631 and 65,591 domains were identified respectively for H3K4me3 and H3K27ac (Fig. 4c). To evaluate the modification level changes of the domains during embryonic tail morphogenesis, the analyses of

differentially modified domains (DMDs) were performed between every two stages (fold change > 2, FDR < 0.05). In line with the significant downward trend of genomic modifications, the downregulated DMDs invariably occupied a majority in any stage comparison for H3K4me3 and especially H3K27ac (Fig. 4d). Also, corresponding to the sharp decline of modifications that occurred preferentially between E10.5 and E12.5, both H3K4me3 and H3K27ac showed much more abundant modification changes with an overwhelming majority of downregulated DMDs (83.4 and 98.0%, $n = 57{,}508$ and 10,920, respectively) during the earlier period (E12.5 *vs*. E10.5) as compared with those during the later period (E14.5 *vs*. E12.5) (Fig. 4d). Here, we followed the downregulated DMDs with interest, given that the expression of genes promoting tail morphogenesis generally decreased over time in the T1 area, theoretically along with the activity and epigenetic mark of relevant promotive CREs including promoters and enhancers. Indeed, numerous downregulated DMDs were linked to genes involved in the essential BPs dominating embryonic tail development by annotation to the nearest TSS. Of these, the distal H3K27ac ones as putative enhancers were downregulated primarily during the earlier period as expected, whereas the downregulation of those H3K4me3-modified in the proximal regions appeared to display a relative delay with a preponderance of downregulated DMDs during the later period (Fig. 4e). This could suggest a tendency to be temporally out-of-sync between the active epigenetic de-modifications of promoters and enhancers associated with the fundamental BPs responsible for tail morphogenesis, when the BPs decline gradually in T1 over developmental time.

Trans- and self-regulation of the blue module

For exploring the potential mechanisms modulating embryonic tail morphogenesis based on the active epigenetic atlases identified above, we paid close attention to the domains annotated to genes within the blue module. Among the H3K4me3 and H3K27ac domains linked to the blue module, most (638/731 and 383/607, respectively) were significantly downregulated during tail morphogenesis (E14.5 *vs*. E10.5), which we then preferentially focused upon as stated earlier. Furthermore, of these, the downregulated ones of H3K4me3 located in proximal regions ($n = 185$) and of H3K27ac wholly ($n = 383$) were then selectively merged based on their genomic interval locations to obtain a total of 487 domains as the putative CRE profile of the blue module, given that H3K4me3 marks the active promoters and H3K27ac can be enriched in both promoters and distal CREs (i.e., enhancers) [93, 95, 97]. Since promoter and enhancer regions are generally rich in the specific binding sites of TFs that actuate the expression of corresponding genes in *trans* [98, 99], we performed a motif enrichment analysis on the 487 putative CREs and obtained the predicted TFs modulating blue module expression. GO analysis of the predicted TF list indicated significant associations with the BPs of tissue morphogenesis, stem cell population maintenance, cell fate commitment, regionalization (clustering with somitogenesis term in the GO enrichment network), muscle structure development, etc. (Fig. 5a and Additional file 1: Fig. S4), which actually occur in T1 for embryonic tail morphogenesis.

To help to explain the high coexpression state of blue module genes, we also attempted to parse out the key TF genes within the module itself, which could drive the expression of total modular genes to a large extent via their products acting in *trans*. It was noted that, of the motif-enriched TFs predicted to regulate blue module expression, 10 had their (mouse orthologous) coding genes included in the module. Further, 6 of the TF genes were expressed with a relatively high IEL in T1, viz. *Msgn1*, *Tbx6*, *Cdx2*, *Sp5*, *Cdx4*, and *Cdx1* (IEL-T1 = 1.980, 1.781, 0.962, 0.941, 0.755, and −0.435 (zFPKM value), respectively) (Fig. 5b), suggesting that they might be a priori putative internal drivers facilitating gene coexpression within the blue module. According to previous studies, all these 6 driver genes possess tail bud-specific expression and have been shown, by reference to their mutant data [49, 52, 69–72], to participate in the regulation of the fundamental BPs relevant to embryonic tail morphogenesis (Fig. 3a). Importantly, they also share another common point of significance, i.e., having been demonstrated/proposed as essential effectors at the downstream of Wnt(3a) signaling pathway [24, 25, 51, 72, 100–102] which dominates embryonic caudal development [23, 26, 101], implying an essential effect of the Wnt(3a) signaling pathway potentially on the expression modulation of the overall blue module.

Moreover, based on the information of putative binding interactions between TFs and true-positive sequences of CRE domains as indicated in the motif enrichment analysis, a presumptive *trans*-acting network of the 6 TF drivers could be constructed with regard to the self-regulation of the blue module. In the network, numerous modular gene loci ($n = 79$, for those highly expressed genes of the top 100 set ranked by the zFPKM value of IEL-T1) were speculated to be targeted via binding to their linked CRE domains in a proximal and/or distal *trans*-regulatory manner by the products of driver genes (Fig. 5c). Moreover, the targeted modular genes included many ligand genes of the Wnt, Fgf, and Notch signaling pathways (Fig. 5c), possibly suggesting the self-feedback regulation of the Wnt signaling pathway during embryonic tail morphogenesis and its regulatory effects on

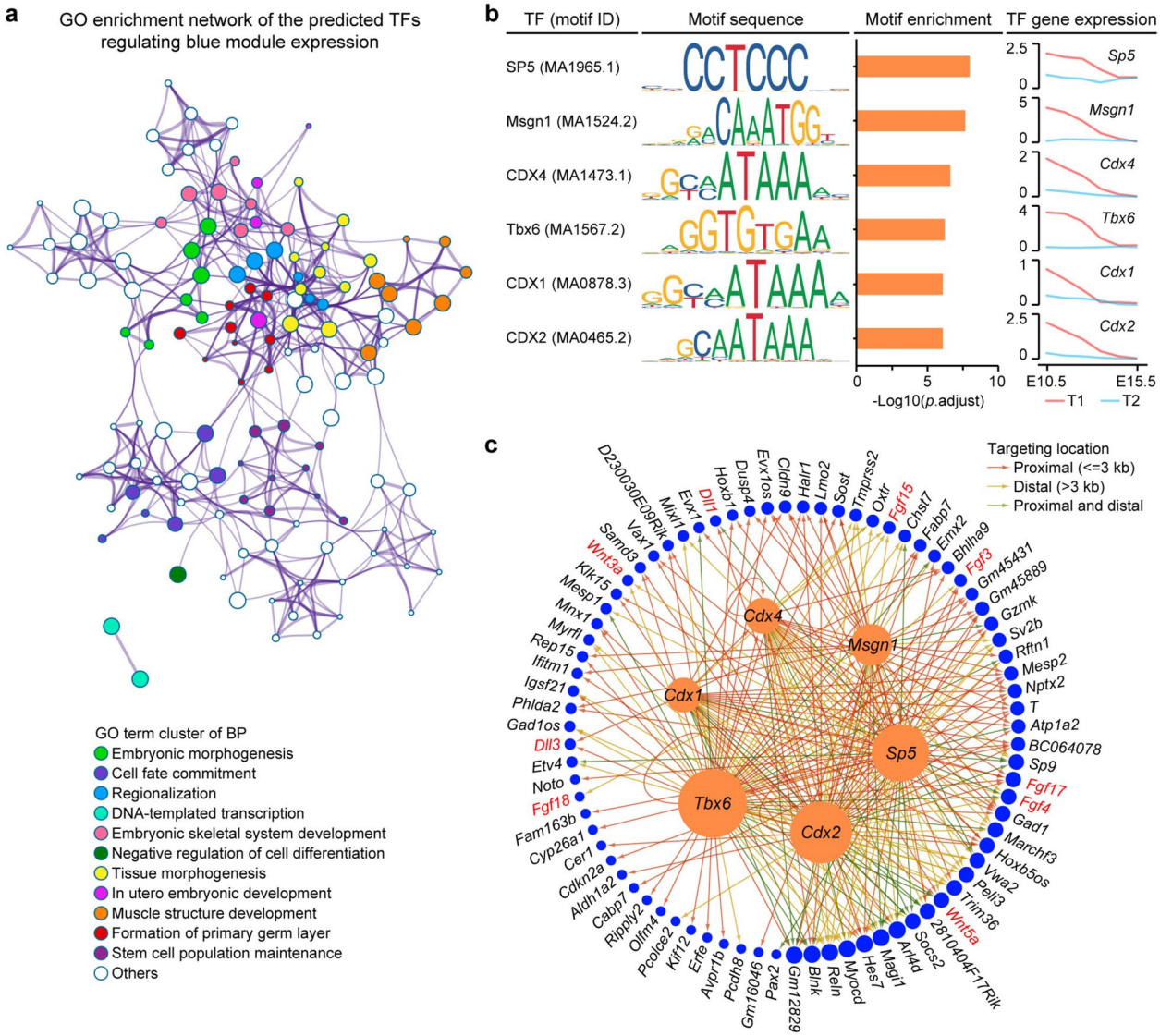

Fig. 5 *Trans*-regulation of blue module gene expression. **a** GO enrichment network of the TFs whose binding motifs are enriched within the putative CREs responsible for regulation of blue module gene expression. Nodes represent enriched BP terms, and their sizes positively correlate with the gene counts enriched in specific terms. Terms were grouped into clusters based on their membership similarities, and the most statistically significant term within a cluster was chosen to represent the cluster. Only the clusters relevant to the BPs involved in embryonic tail morphogenesis are shown here, with the apparently irrelevant ones labeled as 'Others'. **b** The 6 TFs with their binding motifs enriched within the putative CREs regulating blue module gene expression, which are also (homologously) encoded in the blue module, with a relatively high IEL (measured by the zFPKM value at E10.5) in T1 transcriptome data. For the curve diagrams of gene expression, the *y*-axis represents the 2^zFPKM values ($n = 3$ for each sampling type, Additional file 8: Table S44), whose mean is shown at each stage. **c** Self-regulation network construction for blue module with the 6 TF genes indicated in **b** as drivers, based on the presumed interactions between these TFs and their corresponding true-positive sequences of motif enrichment among the putative CREs regulating blue module gene expression. The blue module genes highly expressed within the top 100 set (ranked by the zFPKM value at E10.5 in T1) are selectively shown here. The ligand genes of Wnt, Fgf, and Notch signaling pathways are indicated (red italics). In the illustration of the types of targeting location (upper right corner), proximal and distal locations are delimited by the distance flanking TSS annotated in brackets. For either the TF genes (orange, inner) or other member genes (blue, outer) in the network diagram, node sizes are indicative of the counts of nodes' linked edges that represent the presumed regulatory relationships between genes

other vital signaling pathways responsible for normal tail development, via the downstream effector genes of Wnt such as these TF drivers. In general, the putative CRE profile of the blue module potentially provides valuable clues for studying the complicated regulatory mechanisms of the genes and relevant signaling pathways that participate in embryonic tail morphogenesis.

Discussion

Compared to other appendages of vertebrates (i.e., the forelimbs and hindlimbs), the tail has always received relatively lower scientific attention. Not simply a late-stage product of the posterior extension of body axis, the tail is an organ essential for the environmental adaptation and survival habits of animals. Until recent years, scientists in related fields have called for more research on the tail as a relatively independent object, considering its importance, trait diversity, and contribution in biomimetic applications [103, 104]. In this study, we designed a series of transcriptome sequencing and epigenome sequencing of active histone modifications focused on the process of tail morphogenesis, in order to systematically analyze this highly conserved developmental procedure with the widely applied high-throughput omics sequencing measures.

WGCNA analysis helped to reveal the specific gene set involved in tail morphogenesis, which included numerous known regulatory genes playing important roles in the related BPs, showing a significant consistency with the results in previous studies. Typically, the transcription factor gene *T*, also known as *Brachyury*/*Tbxt*, was indicated as one of the hub genes presenting high connectivity within the coexpression network of the gene set. According to existing literature, *T* gene function is vital for embryonic development including tail morphogenesis, during which *T* is indispensable for tail notochord development [46, 105] and can regulate the Wnt signaling pathway responsible for the formation of paraxial mesoderm [54, 106]. The mutations of *T* gene have proven to be the cause of abnormal tail (length) development in many mammals [107–109], and the generation of its novel type of spliced transcripts has been revealed to be the fundamental evolutionary mechanism for the tail loss in humans and apes [110]. This detail of our analysis results is well aligned with the hub status of *T* gene in modulating tail morphogenesis. Interestingly, based on the results of motif enrichment analysis on the associated CRE domains, the transcription factor T/Brachyury/Tbxt was not predicted to directly play roles in regulating the gene coexpression of the blue module, whereas the speculated modular driver TF genes point to the important regulatory role of Wnt signaling pathway in the gene coexpression of the entire module. This may suggest that *T* gene is at a relatively upstream position in the regulatory network controlling tail morphogenesis, and its profound effects tend to strongly rely on the downstream signaling pathways (such as Wnt) and related TF effector genes, as summarized in [25].

Our transcriptome data reveals numerous novel protein-coding genes with enriched expressions in tail bud, and the validated specific expression patterns suggest their involvement in the relevant BPs. Also, it reminds us to consider the role of lncRNA genes in regulating the tail development. Based on the current understanding of lncRNA loci in modulating the expression of neighboring protein-coding genes [91, 111], it should be inevitable that lncRNA loci participate in and play a certain regulatory effect in tail morphogenesis, at least for the ones neighboring the important protein-coding genes responsible for the fundamental BPs dominating tail formation. It may be an interesting research direction, but there is currently a lack of exploration.

The overall decrease in the level of genome-wide active histone modifications was indicated as an evident characteristic of the developing tail tip tissues of embryos. As the tissue differentiation and maturation proceed, the genome-wide H3K27ac modification level declines significantly, meaning that a huge number of putative enhancers are downregulated in activity and may further get resolved into heterochromatin from the poised state. Given the prevalence of H3K4me3 signal existing at highly active enhancers [112, 113] and its loss during the resolution of poising [114], the pronounced reduction of H3K4me3 signal away from genes may be correlated with a decrease in H3K4me3 signal at the enhancers. Based on the compositions of tail tip tissues of T1 samples at the different stages for CUT&Tag-seq, it could be theoretically inferred that the widespread downregulation of both active histone modifications should primarily occur in the tail bud region, which contains the precursor cell populations and abundant differentiating offspring cells for tail morphogenesis. Therefore, it seems interesting to analyze the epigenetic modification information of embryonic tail tip tissues, especially the tail bud, at the cell-type level in the subsequent work.

Overall, our multi-omics analyses yield new insights into the genetic modulation of vertebrate/mammalian tail development, as a basic data resource of exploitable values for relevant research. Yet, the present study may also have some limitations. For instance, the sampling strategies for sequencing just roughly utilized the differences in tissue composition of embryonic tail tip among the developmental stages, employing a simplification to sample the tail tip regions of the same physical length (≈ 1 mm) at each stage or position, without accurately quantifying the count of somites contained in each sampling type, which might vary depending on the stage or position. Given the general high degree of structural similarity among the determined somites, the overlook for this detail might not have actually caused interferences with the subject. Additionally, our multi-omics data was obtained at the bulk levels, so the genetic/genomic information indicated by the analyses could not be intuitively associated with specific cell types or tissue locations,

which however is able to be efficiently addressed utilizing single-cell and spatial transcriptome techniques. As an important breakthrough point for studying the genetic regulatory mechanisms underlying the tail morphogenesis procedure, the associated CREs identified by the CUT&Tag-seq may further require their corresponding three-dimensional genomic information, which could be acquired by the Hi-C sequencing techniques, to help achieve a more accurate genetic annotation for them. Therefore, the new path of utilizing high-throughput omics sequencing methods to explore the genetic mechanisms of the development of vertebrate/mammalian tail still has a lot of research work to be carried out, and the present study is only a foundational step at the beginning.

Conclusions

Utilizing systematic transcriptome and CUT&Tag sequencing approaches, the present study characterizes the gene expression and chromatin state dynamics that correlate with embryonic tail morphogenesis in the mouse, and identifies candidate genes and CREs that reveal novel molecular markers and potential regulatory mechanisms from non-coding chromatin regions for this evolutionarily conserved process. Our multi-omics analyses provide new insights into the genetic modulation of vertebrate/mammalian tail formation and serve as a fundamental data resource with exploitable value for future research.

Methods

Mouse model

All the mice used in present study were on the genetic background of wild-type C57BL/6 J and were purchased from Liaoning Changsheng Biotechnology Co. LTD (Benxi, China). Mouse care and all the experimental procedures were conducted in compliance with relevant guidelines (see our declaration related to ethics approval).

Sampling for high-throughput sequencing

For the transcriptome sequencing, 3 pregnant females were identified by abdominal morphology at each embryonic stage of E10.5, E11.5, E12.5, E13.5, E14.5, and E15.5, and were dissected to harvest embryos after sacrifice under deep anesthesia. Embryos were washed at least 3 times in precooled phosphate buffered saline (PBS) to remove maternal blood. For each embryo, the tail terminal region of ≈ 1 mm length, termed T1, was excised and transferred to a 1.5-ml Eppendorf (EP) tube filled with RNAlater stabilization solution (Invitrogen, AM7021). So was a tail region of ≈ 1 mm length proximally contiguous to T1, termed T2, using another 1.5-ml EP tube with an identical stabilization solution. The sampling procedure was performed strictly on ice, and in principle for either of T1 and T2, the tissue blocks from the litter embryos ($n = 6 \sim 8$) were integrated in the same tube as one biological replicate sample. Three replicates were obtained for T1 and T2 at each stage. Samples were then stabilized overnight at 4℃ and stored temporarily at − 30℃.

For the CUT&Tag-seq of H3K4me3 or H3K27ac histone modification, embryos were harvested at each stage of E10.5, E12.5, and E14.5 from pregnant females dissected after sacrifice under deep anesthesia, and were washed at least 3 times in precooled PBS. On ice, the T1 region was sampled and transferred into a 2-ml cryovial half-filled with PBS. Every 30 tissue blocks of the same stage were collected into a cryovial as one biological replicate, and 3 replicates were prepared for each stage. After the removal of PBS, the cryovials were immediately immersed in liquid nitrogen for flash freezing and short-term storage.

RNA extraction

All operations were performed on ice or at 4℃. Samples stabilized in RNAlater solution were transferred to 1 ml TRIzol reagent (Invitrogen, 15,596,026), homogenized by means of a SCIENTZ-48 tissue grinder with small steel balls, and then mixed with 200 μl chloroform with shaking, followed by standing for 3 min and centrifugation at 13,000 rpm for 15 min. The supernatant was carefully transferred into a new 1.5-ml EP tube, then mixed with isochoric 70% ethanol, and further transferred to a spin column of RNeasy Mini Kit (Qiagen, 74,104). Subsequently, total RNA was purified according to the protocol of Kit (HB-0570–003) and finally eluted with 30 μl RNase-free water. RNA integrality and contamination were monitored on 1% agarose gels, and RNA purity was checked using a Thermo Scientific NanoDrop-2000 spectrophotometer. RNA samples were then stored at − 80℃.

Transcriptomic library preparation and sequencing

RNA integrity was further assessed using the Agilent Bioanalyzer 2100 system. At least 1 μg total RNA per sample was used as input material for library preparation. The NEBNext UltraTM RNA Library Prep Kit for Illumina was used to generate sequencing libraries following the manufacturer's recommendations. Briefly, mRNAs were purified from total RNA using Oligo(dT)-attached magnetic beads, and then fragmented using divalent cations at elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×). First-strand cDNA was then synthesized with M-MuLV reverse transcriptase ($RNaseH^-$) and random hexamer primers. With RNaseH digestion, second-strand cDNA was subsequently synthesized using DNA Polymerase I. After the double-stranded cDNA was purified, remaining overhangs were converted into blunt ends by exonuclease/polymerase

activities, followed by adenylation of 3' ends of DNA fragments and ligation with sequencing adapter. The library fragments were then purified using the Beckman Coulter AMPure XP system to collect the cDNA fragments of preferentially 250 to 300 bp in length. The size-selected and adaptor-ligated cDNA was treated with NEB USER Enzyme at 37℃ for 15 min followed by 5 min at 95℃, and subsequently amplified with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) primer. Finally, amplified products were purified by Beckman Coulter AMPure XP system, and library quality was assessed with the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using Illumia TruSeq PE Cluster Kit v3-cBot-HS. The cDNA libraries were then sequenced on an Illumina Novaseq platform, and 150-bp paired-end reads were generated.

CUT&Tag library preparation and sequencing

CUT&Tag assays were performed as described previously [115]. Briefly, T1 samples were homogenized and nuclei were extracted. Approximately 50,000 nuclei were collected and incubated with Concanavalin A-coated magnetic beads for 15 min at room temperature. Bead-bound nuclei were then resuspended and incubated with H3K4me3 Rabbit pAb (1:100, Abcam, ab8580) or H3K27ac Rabbit pAb (1:100, Abcam, ab4729) overnight at 4℃. The magnetic stand was used to remove unbound antibodies, and the nuclei were incubated with Goat anti-Rabbit IgG H&L (1:100, ABclonal, AS014) for 30 min at room temperature. After several washes using the magnetic stand, the nuclei were incubated in pA-Tn5 adapter complex preparation for 1 h at room temperature, followed by removal of unbound pA-Tn5 proteins. The nuclei were resuspended and incubated in tagmentation buffer for 1 h at 37o; the mixture was then treated with Proteinase K to stop tagmentation. Afterwards, DNA was extracted using Beckman Coulter AMPure XP beads and amplified with NEBNext HiFi 2×PCR Master Mix, universal i5 primer, and uniquely barcoded i7 primer. After the purification of amplified products with AMPure XP beads, library quantity was determined by Life Invitrogen Qubit 3.0, whereas library quality was assessed by the Agilent Bioanalyzer 2100 system. Libraries were sequenced on an Illumina Novaseq 6000 platform using the PE150 method.

cDNA preparation and riboprobe synthesis

Using the BeyoRT First-Strand cDNA Synthesis Kit (Beyotime, D7166), cDNA was prepared with total RNA extracted from the T1 of E10.5, and was then used as a template to amplify the probe sequences designed for in situ hybridization with Ex Taq DNA Polymerase (TaKaRa, RR001A) and specific primers (SP6-prefixed for the reverse ones, Additional file 2: Table S5). Amplified fragments were purified by PCR Clean Up Kit (Beyotime, D0033), and used as templates for antisense riboprobe synthesis by DIG RNA Labeling Kit (Roche, 11,175,025,910, using the SP6 RNA polymerase). Riboprobes were purified with a RNA Probe Purification Kit (Omega, R6249), then quantitated by Thermo Scientific NanoDrop-2000 spectrophotometer, mixed well with isochoric prehybridization solution, and stored at − 30℃.

In situ *hybridization*

Whole-mount in situ hybridization was performed by integrating previous protocols [116, 117]. Embryos were harvested at E10.5, washed 3 times in cold PBS, and fixed overnight in 4% paraformaldehyde (PFA) in PBS at 4℃. After the removal of PFA by three cold PBTX (PBS + 0.1% Triton X-100) washes, embryos were dehydrated through a PBTX-methanol series and stored at − 30℃ in 100% methanol. For hybridization, embryos were rehydrated through a methanol-PBTX series, bleached in 6% hydrogen peroxide in PBTX for 1 h, and then washed twice in PBTX. After treatment with 10 μg/ml proteinase K in PBTX for 15 min at room temperature and two gentle PBTX washes, embryos were re-fixed with 4% PFA/0.2% glutaraldehyde in PBTX for 20 min and washed twice in PBTX. Embryos were transferred to prehybridization solution for 2 h at 68℃, followed by overnight incubation with the hybridization solution containing 500 ng/ml digoxigenin-labeled riboprobe at 68℃. The next day, after the recovery of riboprobe solution, embryos were washed 7 times (30 min each) with 2×saline sodium citrate (SSC)/50% formamide/0.1% Triton X-100 at 68℃, and then gradiently transferred to maleic acid buffer/0.1% Triton X-100 (MABTX) at room temperature, followed by blocking treatment with 10% goat serum (Solarbio, SL038)/2% blocking reagent (Roche, 11096176001) diluted in MABTX for 2 h at 4℃. Subsequently, embryos were incubated with an anti-digoxigenin primary antibody conjugated to alkaline phosphatase (Roche, 11093274910) at a concentration of 1:2000 in 2% blocking reagent in MABTX overnight at 4℃ with rocking. The following day, embryos were washed 6 times (1 h each) in MABTX at room temperature, and then washed overnight in MABTX at 4y antibody conjugated to alkaline phosphatase (Roche, 11093274910) at a concentration of 1:2000 in 2% blocking reagent in MABTX overnight at 4℃42074001), and then washed 3 times in PBTX. Lastly, embryos were post-fixed with 4% PFA, washed twice in PBTX, and transferred to 100% glycerol through a PBTX-glycerol series for storage and convenient imaging. Embryos were imaged using an Olympus SZX7 stereo microscope and a FluoCa BioHD-C20 digital camera.

A minimum of four embryos were processed for each riboprobe.

Transcriptome data processing and analysis

The FASTQ files of all samples were evaluated for quality control using FastQC [118] (v0.11.9). According to the results of FastQC, we further used TrimGalore (v0.6.4_dev) to trim FASTQ clean data with the parameters of "–quality 25 –stringency 4 –length 30 –clip_R1 10 –clip_R2 10 -e 0.1 –paired –phred33 –gzip." Sequence alignments were performed on all clean reads using HISAT2 (v2.1.0) with default parameters and the reference genome of GRCm38/mm10. Gene expression of RNA-seq was quantified using featureCounts (v2.0.1). DEG analyses were performed using edgeR (v4.0.16), with a multiple-factor analysis approach that included sampling stage and sampling position as factors in the design matrix. The functions of normLibSizes, estimateDisp, glmQLFit, glmQLFTest, and topTags in edgeR were used for normalization, dispersion estimation, hypothesis testing, and extraction of DEGs (log2(fold change) > 2, FDR < 0.05). Further, DEGs were filtered by zFPKM > −3 (the threshold for considering the genes actually expressed [119]) at least in one of the two compared samplings to obtain the final DEG set for each comparison.

WGCNA analysis and hub gene screening

Using the R package WGCNA (v1.70–3), WGCNA analysis was performed based on the DEG sets of T1 *vs*. T2 at all sampling stages. The values of 2^zFPKM of the DEGs were input for WGCNA analysis and were used for distance matrix computation to perform hierarchical cluster analysis, and no obvious outliers were indicated in the sample clustering (Additional file 1: Fig. S5). The correlation between modules and traits was calculated. For the interesting gene module (blue), the visual software Cytoscape [120] (v3.7.2) was used for network construction, and its cytoHubba plug-in was utilized to calculate gene node scores by the maximal clique centrality (MCC) algorithm. The higher score of MCC, the greater weight of node gene in the network. Based on the ranking of MCC scores, the top genes were recruited as the hub ones in the module (the top 30 here for the blue module).

CUT&Tag-seq data processing and analysis

Raw data of FASTQ files were filtered using TrimGalore software to obtain clean data with the parameters of "–quality 25 –stringency 4 –length 30 -e 0.1 –paired –phred33 –gzip," and FastQC was used for quality control before and after data filtering. Clean reads were mapped to GRCm38/mm10 using Bowtie2 (v2.4.1) with the parameters of "-I 10 -X 700 –phred33 –end-to-end –very-sensitive –no-unal –no-mixed –no-discordant." Briefly, we removed the reads from mitochondrial genome, included the reads with mapping quality (MAPQ) > 30, and removed duplicates by Picard (v2.23.8). Detailed statistical information of the reads of all samples is supplied in the supplementary materials (Additional file 2: Table S6). The length distribution of aligned fragments was assessed, and all the samples were confirmed to achieve a successful CUT&Tag experiment, with typical zigzag curves that peaked around the (multifold) length of nucleosomal DNA (Additional file 1: Fig. S6). We performed the peak calling analysis on CUT&Tag-seq data using MACS2 (v2.2.7.1) according to the code of "macs2 callpeak -p 1e-3 -g mm -f BAM -B –keep-dup all" (Additional file 2: Table S7), and annotated the peak sets by ChIPseeker (v1.26.2) and ChIPpeakAnno (v3.24.2). The correlation and PCA clustering analyses on samples were performed using affinity data for all domains, both of which indicated an obvious straying of H3K4me3 E10.5–2 sample from the other two replicates (Additional file 1: Fig. S7), and hence this sample was discarded in the subsequent analyses. All the samples utilized in the present study exhibited evident enrichment of H3K4me3/H3K27ac signal around TSSs (Additional file 1: Fig. S8 and Fig. S9), matching the genomic distribution characteristics of the both modifications. The DMDs (fold change > 2 and FDR < 0.05) between every two sampling stages were identified in a single test using stage as a factor with three levels (E10.5, E12.5, E14.5), by the R package DiffBind (v3.0.15).

GO and KEGG enrichment analyses

All the GO and KEGG enrichment analyses in the present study were performed online using Metascape [121] (https://metascape.org/gp/index.html#/main/step1). Associated significant results are shown in the supplementary materials for the GO enrichment analyses respectively on the DEGs consistently up/downregulated during E10.5 ~ E12.5 in T1 *vs*. T2 (Additional file 3: Table S8 and Table S9), the DEGs up/downregulated at each sampling stage in T1 *vs*. T2 (Additional file 4: Table S10-S21), the DEGs up/downregulated in the other sampling stages *vs*. E10.5 within the T1 region (Additional file 5: Table S22-S31), the genes in each module of WGCNA analysis (Additional file 6: Table S32-S38) and the predicted TFs regulating blue module expression (Additional file 7: Table S39), and for the KEGG pathway enrichment analysis on the genes (the positive-correlated ones, $n = 179$) in blue module (Additional file 7: Table S40). The gene lists of GO terms involved in the analysis of downregulated DMDs of histone modifications were downloaded from the GO database [122, 123] (http://geneontology.org/). The map information of the KEGG pathways enriched in the blue module was

acquired from the KEGG database [60] (https://www.kegg.jp/kegg/pathway.html).

Motif enrichment analysis and putative trans-regulatory network construction

The motif enrichment analysis on the 487 putative CREs linked to blue module genes was performed online using the Analysis of Motif Enrichment (AME, v5.5.7) tool in the MEME Suite [124] (https://meme-suite.org/meme/tools/ame), with the motif database of JASPAR CORE (2022) vertebrates, average odds scoring method of sequence, and Fisher's exact test. The motif logos used in the figure were acquired from the JASPAR database [125] (https://jaspar.elixir.no/). When using the AME tool online, all the primary sequences (input) are "labeled as positive", whereas all the control sequences (shuffled input sequences) are "labeled as negative". Then, sequences with Position Weight Matrix (PWM) score greater than or equal to the optimal threshold (PWM_min) are "classified as positive", thus defining the "true positive" (both labeled and classified as positive) for the primary sequences (input), but the "false positive" (labeled negative but classified as positive) for the control sequences—please refer to the sample output page of AME online (https://meme-suite.org/meme/doc/examples/ame_example_output_files/ame.html) for more information. For the construction of putative *trans*-regulatory networks within the blue module, the presumed binding interactions between the 6 driver TFs and their corresponding true-positive CRE sequences indicated in the motif enrichment analysis, were utilized. Combining with the genetic annotation result of the CRE sequences, the predicted *trans*-regulatory relationships between the TFs and the corresponding target genes were then obtained for the network construction. The targeting manner was classified (proximal or distal) according to the genetic location of CREs relative to the TSSs. The network diagram was plotted using Cytoscape [120] (v3.9.1).

Abbreviations

ADP	Adenosine diphosphate
ARF	ADP-ribosylation factor
AME	Analysis of Motif Enrichment software
AP	Anterior/posterior
BP	Biological process
CREB	CAMP responsive element-binding protein
CRE	*cis*-Regulatory element
CUT&Tag-seq	Cleavage Under Targets and Tagmentation-sequencing
cAMP	Cyclic adenosine monophosphate
DEG	Differentially expressed gene
DMD	Differentially modified domain
EP	Eppendorf
GO	Gene Ontology
H3K27ac	Histone H3 lysine 27 acetylation
H3K4me3	Histone H3 lysine 4 trimethylation
IEL	Initial expression level
KEGG	Kyoto Encyclopedia of Genes and Genomes
lncRNA	Long non-coding RNA
MCC	Maximal clique centrality
PFA	Paraformaldehyde
PBS	Phosphate buffered saline
PI3K	Phosphoinositide 3-kinase
PWM	Position weight matrix
PSM	Presomitic mesoderm
PCA	Principal component analysis
SSC	Saline sodium citrate
TF	Transcription factor
TSS	Transcriptional start site
WGCNA	Weighted gene coexpression network analysis
zFPKM	Z-normalized fragments per kilobase per million mapped reads

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02192-0.

Additional file 1: Figures S1–S9. Fig. S1. Correlation analysis on all the 36 samples of transcriptome sequencing. Fig. S2. Dynamics of T1 transcriptome profiles with the developmental time. Fig. S3. Statistics of the width of H3K4me3 and H3K27ac domains identified at each sampling stage. Fig. S4. *P*-value annotations for the network of enriched GO terms in Fig. 5a. Fig. S5. Sample clustering to detect outliers before WGCNA analysis. Fig. S6. Assessment on the length distribution of aligned fragments for the CUT&Tag-seq samples of T1. Fig. S7. Clustering analyzes on the CUT&Tag-seq samples of T1. Fig. S8. Analysis of H3K4me3 signal near TSSs in CUT&Tag-seq. Fig. S9. Analysis of H3K27ac signal near TSSs in CUT&Tag-seq.

Additional file 2: Tables S1–S7. Table S1. Module memberships of WGCNA analysis. Table S2. Protein-coding genes with known tail bud-specific expression within the blue module ($n = 44$). Table S3. Protein-coding genes without known tail bud-specific expression within the blue module ($n = 41$). Table S4. LncRNA genes contained in the top100 gene set of positive correlation with high IEL in T1 within the blue module ($n = 14$). Table S5. Primers used to amplify the probe sequences designed for *in situ* hybridization from cDNA. Table S6. Statistical information of reads of CUT&Tag-seq samples. Table S7. Peak number and fraction of reads in peak (FRiP) value of CUT&Tag-seq samples.

Additional file 3: Tables S8, S9. Table S8. GO analysis result of the DEGs consistently up-regulated during E10.5 ~ E12.5 in T1 *vs.* T2. Table S9. GO analysis result of the DEGs consistently down-regulated during E10.5 ~ E12.5 in T1 *vs.* T2.

Additional file 4: Tables S10–S21. Table S10. GO analysis result of the DEGs up-regulated at E10.5 in T1 *vs.* T2. Table S11. GO analysis result of the DEGs up-regulated at E11.5 in T1 *vs.* T2. Table S12. GO analysis result of the DEGs up-regulated at E12.5 in T1 *vs.* T2. Table S13. GO analysis result of the DEGs up-regulated at E13.5 in T1 *vs.* T2. Table S14. GO analysis result of the DEGs up-regulated at E14.5 in T1 *vs.* T2. Table S15. GO analysis result of the DEGs up-regulated at E15.5 in T1 *vs.* T2. Table S16. GO analysis result of the DEGs down-regulated at E10.5 in T1 *vs.* T2. Table S17. GO analysis result of the DEGs down-regulated at E11.5 in T1 *vs.* T2. Table S18. GO analysis result of the DEGs down-regulated at E12.5 in T1 *vs.* T2. Table S19. GO analysis result of the DEGs down-regulated at E13.5 in T1 *vs.* T2. Table S20. GO analysis result of the DEGs down-regulated at E14.5 in T1 *vs.* T2. Table S21. GO analysis result of the DEGs down-regulated at E15.5 in T1 *vs.* T2.

Additional file 5: Tables S22–S31. Table S22. GO analysis result of the DEGs up-regulated in E11.5 *vs.* E10.5 within the T1 region. Table S23. GO analysis result of the DEGs up-regulated in E12.5 *vs.* E10.5 within the T1 region. Table S24. GO analysis result of the DEGs up-regulated in E13.5 *vs.* E10.5 within the T1 region. Table S25. GO analysis result of the DEGs up-regulated in E14.5 *vs.* E10.5 within the T1 region. Table S26. GO analysis result of the DEGs up-regulated in E15.5 *vs.* E10.5 within the T1 region. Table S27. GO analysis result of the DEGs down-regulated in E11.5 *vs.* E10.5 within the T1 region. Table S28. GO analysis result of the DEGs down-regulated in E12.5 *vs.* E10.5 within the T1 region. Table S29. GO analysis result of the DEGs down-regulated in E13.5 *vs.* E10.5 within the T1 region. Table S30. GO

analysis result of the DEGs down-regulated in E14.5 *vs*. E10.5 within the T1 region. Table S31. GO analysis result of the DEGs down-regulated in E15.5 *vs*. E10.5 within the T1 region.

Additional file 6: Tables S32–S38. Table S32. GO analysis result of the genes in the blue module of WGCNA analysis. Table S33. GO analysis result of the genes in the turquoise module of WGCNA analysis. Table S34. GO analysis result of the genes in the green module of WGCNA analysis. Table S35. GO analysis result of the genes in the red module of WGCNA analysis. Table S36. GO analysis result of the genes in the brown module of WGCNA analysis. Table S37. GO analysis result of the genes in the black module of WGCNA analysis. Table S38. GO analysis result of the genes in the yellow module of WGCNA analysis.

Additional file 7: Tables S39, S40. Table S39. GO (biological process) analysis result of the predicted TFs modulating the gene expression of blue module, performed with Metascape. Table S40. Result of KEGG pathway enrichment analysis on the genes (the positive-correlated ones, $n = 179$) in blue module, performed with Metascape.

Additional file 8: Tables S41–S44. Table S41. Gene expression values (2^zFPKM, $n = 3$) used for plotting Fig. 2a. Table S42. MEblue values ($n = 3$ for each sampling) used for plotting Fig. 2d. Table S43. Gene expression values (2^zFPKM, $n = 3$) used for plotting Fig. 3c. Table S44. Gene expression values (2^zFPKM, $n = 3$) used for plotting Fig. 5b.

Acknowledgements
We sincerely thank colleagues Yong Shao, Sheng Wang, and Yun-Mei Wang from the State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, for suggestions on the project.

Authors' contributions
DDW and WDB designed and leaded the project. WDB sampled the sequencing materials, performed the experiments, and drafted the paper. YXC analyzed the transcriptome and CUT&Tag-seq data. WDB, YXC, and XPZ drew the diagrams. XPZ, DNC, DDW, and YXC revised the manuscript. All authors discussed the results and implications, and commented on the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the CAS Light of West China Program (xbzg-zdsys-202213) and the Yunnan Applied Basic Research Projects (202401AT070303, 202305AH340006, 2019FI010, 2019FA008, and 2019FJ003).

Data availability
The authors declare that all data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. The individual data values of replicates used for plotting in the figures are provided in Additional file 8: Table S41–S44. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive [126] in National Genomics Data Center [127], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA013613) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Declarations

Ethics approval and consent to participate
Mouse care and all the experimental procedures in the present study were conducted in compliance with the guidelines of the Institutional Animal Care and Use Committee (IACUC) of the Kunming Institute of Zoology, Chinese Academy of Sciences. The approval number for all the contents of this research is IACUC-RE-2023–01-002.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 4 January 2024 Accepted: 12 March 2025
Published online: 26 March 2025

References

1. Webb PW. Locomotor patterns in the evolution of actinopterygian fishes. Am Zool. 1982;22:329–42.
2. Shield S, Jericevich R, Patel A, Jusufi A. Tails, flails, and sails: how appendages improve terrestrial maneuverability by improving stability. Integr Comp Biol. 2021;61:506–20.
3. Young JW, Chadwell BA, Dunham NT, McNamara A, Phelps T, Hieronymus T, Shapiro LJ. The stabilizing function of the tail during arboreal quadrupedalism. Integr Comp Biol. 2021;61:491–505.
4. Arbour VM. Estimating impact forces of tail club strikes by ankylosaurid dinosaurs. PLoS One. 2009;4:e6738.
5. Hale ME. Functional morphology of ventral tail bending and prehensile abilities of the seahorse. Hippocampus kuda J Morphol. 1996;227:51–65.
6. Nelson EL, Kendall GA. Goal-directed tail use in Colombian spider monkeys (Ateles fusciceps rufiventris) is highly lateralized. J Comp Psychol. 2018;132:40–7.
7. O'Connor SM, Dawson TJ, Kram R, Donelan JM. The kangaroo's tail propels and powers pentapedal locomotion. Biol Lett. 2014;10:20140381.
8. Higham TE, Russell AP, Zani PA. Integrative biology of tail autotomy in lizards. Physiol Biochem Zool. 2013;86:603–10.
9. Fallon JF, Simandl BK. Evidence of a role for cell death in the disappearance of the embryonic human tail. Am J Anat. 1978;152:111–29.
10. Tubbs RS, Malefant J, Loukas M, Jerry Oakes W, Oskouian RJ, Fries FN. Enigmatic human tails: a review of their history, embryology, classification, and clinical manifestations. Clin Anat. 2016;29:430–8.
11. Yaoita Y. Tail resorption during metamorphosis in xenopus tadpoles. Front Endocrinol (Lausanne). 2019;10:143.
12. Holmdahl DE. Experimentelle untersuchungen uber die lage der grenze primarer und sekundarer korperentwicklung beim huhn. Anat Anz. 1925;59:393–6.
13. Griffith CM, Wiley MJ, Sanders EJ. The vertebrate tail bud: three germ layers from one tissue. Anat Embryol (Berl). 1992;185:101–13.
14. Cambray N, Wilson V. Axial progenitors with extensive potency are localised to the mouse chordoneural hinge. Development. 2002;129:4855–66.
15. Cambray N, Wilson V. Two distinct sources for a population of maturing axial progenitors. Development. 2007;134:2829–40.
16. Jurberg AD, Aires R, Varela-Lasheras I, Novoa A, Mallo M. Switching axial progenitors from producing trunk to tail tissues in vertebrate embryos. Dev Cell. 2013;25:451–62.
17. McGrew MJ, Sherman A, Lillico SG, Ellard FM, Radcliffe PA, Gilhooley HJ, Mitrophanous KA, Cambray N, Wilson V, Sang H. Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. Development. 2008;135:2289–99.
18. Shifley ET, Vanhorn KM, Perez-Balaguer A, Franklin JD, Weinstein M, Cole SE. Oscillatory lunatic fringe activity is crucial for segmentation of the anterior but not posterior skeleton. Development. 2008;135:899–908.
19. Casaca A, Novoa A, Mallo M. Hoxb6 can interfere with somitogenesis in the posterior embryo through a mechanism independent of its rib-promoting activity. Development. 2016;143:437–48.
20. DeVeale B, Brokhman I, Mohseni P, Babak T, Yoon C, Lin A, Onishi K, Tomilin A, Pevny L, Zandstra PW, et al. Oct4 is required ~E7.5 for proliferation in the primitive streak. PLoS Genet. 2013;9:e1003957.
21. Aires R, de Lemos L, Novoa A, Jurberg AD, Mascrez B, Duboule D, Mallo M. Tail bud progenitor activity relies on a network comprising Gdf11, Lin28, and Hox13 genes. Dev Cell. 2019;48:383-95 e8.
22. Robinton DA, Chal J, da Lummertz Rocha E, Han A, Yermalovich AV, Oginuma M, Schlaeger TM, Sousa P, Rodriguez A, Urbach A, et al. The Lin28/let-7 pathway regulates the mammalian caudal body axis elongation program. Dev Cell. 2019;48:396-405 e3.
23. Takada S, Stark KL, Shea MJ, Vassileva G, McMahon JA, McMahon AP. Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev. 1994;8:174–89.
24. Ikeya M, Takada S. Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of cdx-1 expression. Mech Dev. 2001;103:27–33.

25. Kimelman D. Tales of tails (and trunks): forming the posterior body in vertebrate embryos. Curr Top Dev Biol. 2016;116:517–36.
26. Greco TL, Takada S, Newhouse MM, McMahon JA, McMahon AP, Camper SA. Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. Genes Dev. 1996;10:313–24.
27. White PH, Farkas DR, McFadden EE, Chapman DL. Defective somite patterning in mouse embryos with reduced levels of Tbx6. Development. 2003;130:1681–90.
28. Galceran J, Farinas I, Depew MJ, Clevers H, Grosschedl R. Wnt3a-/–like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev. 1999;13:709–17.
29. Naiche LA, Holder N, Lewandoski M. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. Proc Natl Acad Sci U S A. 2011;108:4018–23.
30. Benazeraf B, Francois P, Baker RE, Denans N, Little CD, Pourquie O. A random cell motility gradient downstream of FGF controls elongation of an amniote embryo. Nature. 2010;466:248–52.
31. Regev I, Guevorkian K, Gupta A, Pourquie O, Mahadevan L. Rectified random cell motility as a mechanism for embryo elongation. Development. 2022;149:dev199423.
32. Mansour SL, Goddard JM, Capecchi MR. Mice homozygous for a targeted disruption of the proto-oncogene int-2 have developmental defects in the tail and inner ear. Development. 1993;117:13–28.
33. Boulet AM, Capecchi MR. Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. Dev Biol. 2012;371:235–45.
34. Dunwoodie SL, Clements M, Sparrow DB, Sa X, Conlon RA, Beddington RS. Axial skeletal defects caused by mutation in the spondylocostal dysplasia/pudgy gene Dll3 are associated with disruption of the segmentation clock within the presomitic mesoderm. Development. 2002;129:1795–806.
35. Saga Y, Hata N, Koseki H, Taketo MM. Mesp2: a novel mouse gene expressed in the presegmented mesoderm and essential for segmentation initiation. Genes Dev. 1997;11:1827–39.
36. Chan T, Kondow A, Hosoya A, Hitachi K, Yukita A, Okabayashi K, Nakamura H, Ozawa H, Kiyonari H, Michiue T, et al. Ripply2 is essential for precise somite formation during mouse early development. FEBS Lett. 2007;581:2691–6.
37. Bessho Y, Sakata R, Komatsu S, Shiota K, Yamada S, Kageyama R. Dynamic expression and essential functions of Hes7 in somite segmentation. Genes Dev. 2001;15:2642–7.
38. Zhang N, Gridley T. Defects in somite formation in lunatic fringe-deficient mice. Nature. 1998;394:374–7.
39. Shangguan Y, Li C, Lin H, Ou M, Tang D, Dai Y, Yan Q. Application of single-cell RNA sequencing in embryonic development. Genomics. 2020;112:4547–51.
40. Aanes H, Collas P, Alestrom P. Transcriptome dynamics and diversity in the early zebrafish embryo. Brief Funct Genomics. 2014;13:95–105.
41. Singh AJ, Ramsey SA, Filtz TM, Kioussi C. Differential gene regulatory networks in development and disease. Cell Mol Life Sci. 2018;75:1013–25.
42. Karlsson G, Sommarin MNE, Boiers C. Defining the emerging blood system during development at single-cell resolution. Front Cell Dev Biol. 2021;9:660350.
43. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell. 2016;167:1170–87.
44. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2011;13:59–69.
45. Hill MA. Mouse timeline detailed. Embryology. 2023. https://embryology.med.unsw.edu.au/embryology/index.php/Mouse_Timeline_Detailed.
46. Schifferl D, Scholze-Wittler M, Wittler L, Veenvliet JV, Koch F, Herrmann BG. A 37 kb region upstream of brachyury comprising a notochord enhancer is essential for notochord and tail development. Development. 2021;148:dev200059.
47. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat. 2010;20:281–300.
48. Anderson MJ, Schimmang T, Lewandoski M. An FGF3-BMP signaling axis regulates caudal neural tube closure, neural crest specification and anterior-posterior axis extension. PLoS Genet. 2016;12:e1006018.
49. Chalamalasetty RB, Dunty WC Jr, Biris KK, Ajima R, Iacovino M, Beisaw A, Feigenbaum L, Chapman DL, Yoon JK, Kyba M, Yamaguchi TP. The Wnt3a/beta-catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. Nat Commun. 2011;2:390.
50. de Hrabe Angelis M, McIntyre J 2nd, Gossler A. Maintenance of somite borders in mice requires the Delta homologue Dll1. Nature. 1997;386:717–21.
51. Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F. Cdx2 is essential for axial elongation in mouse development. Proc Natl Acad Sci U S A. 2004;101:7641–5.
52. Chawengsaksophak K, James R, Hammond VE, Kontgen F, Beck F. Homeosis and intestinal tumours in Cdx2 mutant mice. Nature. 1997;386:84–7.
53. Arnold SJ, Stappert J, Bauer A, Kispert A, Herrmann BG, Kemler R. Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. Mech Dev. 2000;91:249–58.
54. Martin BL, Kimelman D. Regulation of canonical Wnt signaling by Brachyury is essential for posterior mesoderm formation. Dev Cell. 2008;15:121–33.
55. Yamaguchi TP, Takada S, Yoshikawa Y, Wu N, McMahon AP. T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes Dev. 1999;13:3185–90.
56. Wahi K, Bochter MS, Cole SE. The many roles of Notch signaling during vertebrate somitogenesis. Semin Cell Dev Biol. 2016;49:68–75.
57. Sasaki N, Kiso M, Kitagawa M, Saga Y. The repression of Notch signaling occurs via the destabilization of mastermind-like 1 by Mesp2 and is essential for somitogenesis. Development. 2011;138:55–64.
58. Hofmann M, Schuster-Gossler K, Watabe-Rudolph M, Aulehla A, Herrmann BG, Gossler A. WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. Genes Dev. 2004;18:2712–7.
59. Seetharaman S, Etienne-Manneville S. Cytoskeletal Crosstalk in Cell Migration. Trends Cell Biol. 2020;30:720–35.
60. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587–92.
61. Maruoka Y, Ohbayashi N, Hoshikawa M, Itoh N, Hogan BL, Furuta Y. Comparison of the expression of three highly related genes, Fgf8, Fgf17 and Fgf18, in the mouse embryo. Mech Dev. 1998;74:175–7.
62. Wright TJ, Ladher R, McWhirter J, Murre C, Schoenwolf GC, Mansour SL. Mouse FGF15 is the ortholog of human and chick FGF19, but is not uniquely required for otic induction. Dev Biol. 2004;269(1):264–75.
63. Mahmut A, Boulanger MC, Bouchareb R, Hadji F, Mathieu P. Adenosine derived from ecto-nucleotidases in calcific aortic valve disease promotes mineralization through A2a adenosine receptor. Cardiovasc Res. 2015;106:109–20.
64. Xie Z, Liu D, Liu S, Calderon L, Zhao G, Turk J, Guo Z. Identification of a cAMP-response element in the regulator of G-protein signaling-2 (RGS2) promoter as a key cis-regulatory element for RGS2 transcriptional regulation by angiotensin II in cultured vascular smooth muscles. J Biol Chem. 2011;286:44646–58.
65. Lopez TP, Fan CM. Dynamic CREB family activity drives segmentation and posterior polarity specification in mammalian somitogenesis. Proc Natl Acad Sci U S A. 2013;110:E2019–27.
66. Lutz B, Schmid W, Niehrs C, Schutz G. Essential role of CREB family proteins during Xenopus embryogenesis. Mech Dev. 1999;88:55–66.
67. Yamaguchi TP, Bradley A, McMahon AP, Jones S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. Development. 1999;126:1211–23.
68. Gluecksohn-Schoenheimer S. The development of normal and homozygous brachy (T/T) mouse embryos in the extraembryonic coelom of the chick. Proc Natl Acad Sci U S A. 1944;30:134–40.
69. Chapman DL, Papaioannou VE. Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. Nature. 1998;391:695–7.
70. van den Akker E, Forlani S, Chawengsaksophak K, de Graaff W, Beck F, Meyer BI, Deschamps J. Cdx1 and Cdx2 have overlapping functions

in anteroposterior patterning and posterior axis elongation. Development. 2002;129:2181–93.
71. van Nes J, de Graaff W, Lebrin F, Gerhard M, Beck F, Deschamps J. The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. Development. 2006;133:419–28.
72. Dunty WC Jr, Kennedy MW, Chalamalasetty RB, Campbell K, Yamaguchi TP. Transcriptional profiling of Wnt3a mutants identifies Sp transcription factors as essential effectors of the Wnt/beta-catenin pathway in neuromesodermal stem cells. PLoS One. 2014;9:e87018.
73. Hofmann I, Thompson A, Sanderson CM, Munro S. The Arl4 family of small G proteins can recruit the cytohesin Arf6 exchange factors to the plasma membrane. Curr Biol. 2007;17:711–6.
74. Lin MC, Yu CJ, Lee FS. Phosphorylation of Arl4A/D promotes their binding by the HYPK chaperone for their stable recruitment to the plasma membrane. Proc Natl Acad Sci U S A. 2022;119:e2207414119.
75. Lawton AK, Nandi A, Stulberg MJ, Dray N, Sneddon MW, Pontius W, Emonet T, Holley SA. Regulated tissue fluidity steers zebrafish body elongation. Development. 2013;140:573–82.
76. Gebauer JM, Keene DR, Olsen BR, Sorokin LM, Paulsson M, Wagener R. Mouse AMACO, a kidney and skin basement membrane associated molecule that mediates RGD-dependent cell attachment. Matrix Biol. 2009;28:456–62.
77. Hiroyasu S, Jones JCR. A new component of the Fraser complex. J Invest Dermatol. 2014;134:1192–3.
78. Jensen LE, Whitehead AS. Pellino3, a novel member of the Pellino protein family, promotes activation of c-Jun and Elk-1 and may act as a scaffolding protein. J Immunol. 2003;171:1500–6.
79. Siednienko J, Jackson R, Mellett M, Delagic N, Yang S, Wang B, Tang LS, Callanan JJ, Mahon BP, Moynagh PN. Pellino3 targets the IRF7 pathway and facilitates autoregulation of TLR3- and viral-induced expression of type I interferons. Nat Immunol. 2012;13:1055–62.
80. Butler MP, Hanly JA, Moynagh PN. Pellino3 is a novel upstream regulator of p38 MAPK and activates CREB in a p38-dependent manner. J Biol Chem. 2005;280:27759–68.
81. Ni J, Bao S, Johnson RI, Zhu B, Li J, Vadaparampil J, Smith CM, Campbell KN, Grahammer F, Huber TB, et al. MAGI-1 interacts with nephrin to maintain slit diaphragm structure through enhanced Rap1 activation in podocytes. J Biol Chem. 2016;291:24406–17.
82. Wright GJ, Leslie JD, Ariza-McNaughton L, Lewis J. Delta proteins and MAGI proteins: an interaction of Notch ligands with intracellular scaffolding molecules and its significance for zebrafish development. Development. 2004;131:5659–69.
83. Mizuhara E, Nakatani T, Minaki Y, Sakamoto Y, Ono Y, Takai Y. MAGI1 recruits Dll1 to cadherin-based adherens junctions and stabilizes it on the cell surface. J Biol Chem. 2005;280:26499–507.
84. Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. EMBO J. 2003;22:3015–26.
85. Saeki K, Fukuyama S, Ayada T, Nakaya M, Aki D, Takaesu G, Hanada T, Matsumura Y, Kobayashi T, Nakagawa R, Yoshimura A. A major lipid raft protein raftlin modulates T cell receptor signaling and enhances th17-mediated autoimmune responses. J Immunol. 2009;182:5929–37.
86. Watanabe A, Tatematsu M, Saeki K, Shibata S, Shime H, Yoshimura A, Obuse C, Seya T, Matsumoto M. Raftlin is involved in the nucleocapture complex to induce poly(I:C)-mediated TLR3 activation. J Biol Chem. 2011;286:10702–11.
87. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458:223–7.
88. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316:1484–8.
89. Kopp F, Mendell JT. functional classification and experimental dissection of long noncoding RNAs. Cell. 2018;172:393–407.
90. Yan P, Luo S, Lu JY, Shen X. Cis- and trans-acting lncRNAs in pluripotency and reprogramming. Curr Opin Genet Dev. 2017;46:170–8.
91. Luo S, Lu JY, Liu L, Yin Y, Chen C, Han X, Wu B, Xu R, Liu W, Yan P, et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. Cell Stem Cell. 2016;18:637–52.
92. Bell CC, Amaral PP, Kalsbeek A, Magor GW, Gillinder KR, Tangermann P, di Lisio L, Cheetham SW, Gruhl F, Frith J, et al. The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation. Sci Rep. 2016;6:26657.
93. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell. 2007;25:15–30.
94. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21:381–95.
95. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107:21931–6.
96. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell. 2013;155:1521–31.
97. Li X, Zhang J, Li D, He C, He K, Xue T, Wan L, Zhang C, Liu Q. Astrocytic ApoE reprograms neuronal cholesterol metabolism and histone-acetylation-mediated memory. Neuron. 2021;109:957-70 e8.
98. Zabidi MA, Stark A. Regulatory enhancer-core-promoter communication via transcription factors and cofactors. Trends Genet. 2016;32:801–14.
99. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. Genome Res. 2010;20:565–77.
100. Wittler L, Shin EH, Grote P, Kispert A, Beckers A, Gossler A, Werber M, Herrmann BG. Expression of Msgn1 in the presomitic mesoderm is controlled by synergism of WNT signalling and Tbx6. EMBO Rep. 2007;8:784–9.
101. Dunty WC Jr, Biris KK, Chalamalasetty RB, Taketo MM, Lewandoski M, Yamaguchi TP. Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. Development. 2008;135:85–94.
102. Pilon N, Oh K, Sylvestre JR, Bouchard N, Savory J, Lohnes D. Cdx4 is a direct target of the canonical Wnt pathway. Dev Biol. 2006;289:55–63.
103. Schwaner MJ, Hsieh ST, Braasch I, Bradley S, Campos CB, Collins CE, Donatelli CM, Fish FE, Fitch OE, Flammang BE, et al. Future tail tales: a forward-looking, integrative perspective on tail research. Integr Comp Biol. 2021;61:521–37.
104. Schwaner MJ, Hsieh ST, Swalla BJ, McGowan CP. An introduction to an evolutionary tail: EvoDevo, structure, and function of post-anal appendages. Integr Comp Biol. 2021;61:352–7.
105. Kemmler CL, Smolikova J, Moran HR, Mannion BJ, Knapp D, Lim F, Czarkwiani A, Hermosilla Aguayo V, Rapp V, Fitch OE, et al. Conserved enhancers control notochord expression of vertebrate Brachyury. Nat Commun. 2023;14:6594.
106. Mariani L, Guo X, Menezes NA, Drozd AM, Cakal SD, Wang Q, Ferretti E. A TALE/HOX code unlocks WNT signalling response towards paraxial mesoderm. Nat Commun. 2021;12:5136.
107. Wu B, Shao Y, Chen B, Liu C, Xue Z, Wu P, Li H. Identification of a novel mouse brachyury (T) allele causing a short tail mutation in mice. Cell Biochem Biophys. 2010;58:129–35.
108. Buckingham KJ, McMillin MJ, Brassil MM, Shively KM, Magnaye KM, Cortes A, Weinmann AS, Lyons LA, Bamshad MJ. Multiple mutant T alleles cause haploinsufficiency of Brachyury and short tails in Manx cats. Mamm Genome. 2013;24:400–8.
109. Han J, Yang M, Guo T, Niu C, Liu J, Yue Y, Yuan C, Yang B. Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. Anim Genet. 2019;50:772–7.
110. Xia B, Zhang W, Zhao G, Zhang X, Bai J, Brosh R, Wudzinska A, Huang E, Ashe H, Ellis G, et al. On the genetic basis of tail-loss evolution in humans and apes. Nature. 2024;626:1042–8.
111. Villegas VE, Zaphiropoulos PG. Neighboring gene regulation by antisense long non-coding RNAs. Int J Mol Sci. 2015;16:3251–66.

112. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet. 2014;46:1311–20.
113. Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC, Adelman K. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. 2018;32:26–41.
114. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125:315–26.
115. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019;10:1930.
116. Beverdam A. Whole mount in situ hybridization of mid-gestation mouse embryos. Curr Protoc Mouse Biol. 2020;10:e77.
117. Long HK, Osterwalder M, Welsh IC, Hansen K, Davies JOJ, Liu YE, Koska M, Adams AT, Aho R, Arora N, et al. Loss of extreme long-range enhancers in human neural crest drives a craniofacial disorder. Cell Stem Cell. 2020;27:765-83 e14.
118. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
119. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. Finding the active genes in deep RNA-seq gene expression studies. BMC Genomics. 2013;14:778.
120. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
121. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10:1523.
122. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
123. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, et al. The gene ontology knowledgebase in 2023. Genetics. 2023;224:iyad031.
124. McLeay RC, Bailey TL. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 2010;11:165.
125. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Perez N, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022;50:D165–73.
126. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, et al. The genome sequence archive family: toward explosive data growth and diverse data types. Genomics Proteomics Bioinformatics. 2021;19:578–83.
127. Database Resources of the National Genomics Data Center. China National Center for Bioinformation in 2022. Nucleic Acids Res. 2022;50:D27-d38.

Publisher's Note