scientific reports

OPEN

Accessible chromatin reveals regulatory mechanisms underlying cell fate decisions during early embryogenesis

Tongqiang Fan & Youjun Huang✉

This study was conducted to investigate epigenetic landscape across multiple species and identify transcription factors (TFs) and their roles in controlling cell fate decision events during early embryogenesis. We made a comprehensively joint-research of chromatin accessibility of five species during embryogenesis by integration of ATAC-seq and RNA-seq datasets. Regulatory roles of candidate early embryonic TFs were investigated. Widespread accessible chromatin in early embryos overlapped with putative *cis*-regulatory sequences. Sets of cell-fate-determining TFs were identified. *YOX1*, a key cell cycle regulator, were found to homologous to clusters of TFs that are involved in neuron and epidermal cell-fate determination. Our research provides an intriguing insight into evolution of cell-fate decision during early embryogenesis among organisms.

An outstanding challenge of developmental biology is to explain how multi-cell organisms originate from a single cell that mature through complex dynamic processes. In complex organs, the generation of a single lineage usually involves multiple steps of cell fate decisions. Comprehensively, understanding the pathways of cell lineage differentiation during in vivo development, especially transcriptional regulatory strategies at point of cell lineage segregation, and the common characteristics of multiple organisms, in critical for directing stem cell differentiation into desired cell types and the discovery of common ground of eukaryotes. Early embryo development of the metazoans and plants comprise a sequence of cell fate decisions in which cells are guided along a pathway of restricted potential and increasing specialisation.

During mammalian early embryo development, stem cell will undergo some cell fate determined phases, including ZGA (zygote genome activation), 1-, 2-, 4-, 8-, 16-, 32-cell, early blastocyst, and late blastocyst stages, in which 8–32-cell stages exhibit first cell fate decision, and early to late blastocyst stages occur second cell fate decision[1–3]. The embryonic cell lineage of *C. elegans* has been traced from zygote to newly hatched larva[4], in which 1-, 2-, 4-, 8-, and 16-cell stage embryos can represent early embryo cell lineage[5]. The control of the oocyte-to-embryo transition in Drosophila parallels that of other animals, however, in early *Drosophila* embryos, the master checkpoint laid on nuclear cycle from cycle 10 to 14[6].

Compared to mammals, plants are ancient organisms, which require many centuries to obtain new organs. Shoot, root, flowers, fruits are continuously augmented to build sophisticated post-embryonic tissues. This process requires coordination of cell divisions, cell fate determination, and cell communication. Starting from fertilization, embryogenesis produces the first stem cell, and by the end of embryogenesis the zygote has transformed into a mature embryo that comprises the basic tissue types identical to any post-embryonic plant[7]. The processes of plant early embryo development include multiple stages: two cell, eight cell, 16 cell, globular, heart, torpedo, and bent stages[8].

TFs (transcription factors) play an irreplaceable role in cell fate determination during embryogenesis[9]. The identities of cell types are also influenced by environmental signal transduction, in turn, results in the activation or inhibition of TFs[10]. Much progress has been made in understanding how core regulators such as *OCT4*, *NANOG*, and *SOX2* as well as transcriptional effectors of signaling pathways, such as *SMAD1/2/3/4*, and *TCF3*, control the regulatory circuity that control earliest stage of embryonic development[11]. During early embryogenesis, inner cells develop a stable regulation circuit, in which *OCT4*, *SOX2*, and *NANOG* are assigned to promote pluripotency to determine ICM (inner cell mass) cell fate[12]. In contrary, outer cells upregulate TFs such as *CDX2* and *EOMES* to promote the formation of trophectoderm-destined cells[13,14]. Regulation of trophectoderm targets

State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Lin'an, Hangzhou 311300, People's Republic of China. ✉email: hyj@zafu.edu.cn

by *OCT4*, *SOX2* and *NANOG* in the pluripotent lineage[15], accompanied by the autoregulatory properties of *OCT4* and *CDX2*[16], ensure the maintenance of lineage segregation. *SALL4*, which establishes and maintains ICM integrity by upregulating *Oct4* and Nanog expression[17]; *TEAD4*, which acts upstream regulator of *CDX2* during trophectoderm development[18]; and episomal expression of the early trophoblast TFs such as *TCFAP2C*, *GATA3*, *ELF5* or downregulation of the pluripotency factor *OCT4* can induce trophoblast cell fate determination in ES cells[19–21]; are all important supplements to the regulatory networks of the first cell fate decision.

Generally, TFs bind to open chromatin of *cis*-regulatory regions, such as promoters, and enhancers, however, a subset called 'pioneer factor' are dominant in their ability to engage silent, unmarked chromatin and initiate the recruitment of other factors, thereby creating a permissive state for gene activation[22,23].

The development of a multicellular organism with its organs and tissues is a reproducible event that shows high coordination between the increase of cell mass and the diversification of cell population. These reproducibility programmes underly species cell-fate decision events[24,25], which are implemented by gene regulatory networks (GRNs), which are basic units of molecular activity that establish and maintain functional tissues by promoting sequential, irreversible gene expression that link genes to cell lineages[26,27]. The regulation mechanisms of cell-fate decisions lack of systematic research during early embryonic development at the chromatin level, especially with regard to cross-species joint research. Here, we focused on the functions of TFs and GRNs associated with cell-fate decision events during early embryogenesis. We used publicly available ATAC-seq[28] datasets of six species, human, mouse, *Drosophila*, *C. elegans*, *Arabidopsis*, and yeast (we termed cell cycle stage as early embryogenesis, as lifespan of yeast is short and cell-fate decision events may occur as early as cell cycle) to question whether there are common traits of TFs during early embryonic development and to investigate underlying transcriptional regulatory mechanism of cell fate control. We found many TFs match their functions of various cell fate determination by investigation of the dynamic chromatin changes and gene expression patterns of various species during different development stages. Furthermore, we quantified TFs abundance from different cells during early embryonic development. Finally, we investigated the evolutionary mechanisms underlying cell-fate determination. After data mining, we focused on *YOX1*, a key cell-cycle regulator in yeast, combined with homology, transcriptomes, and regulatory networks, the conserved roles of the homologs were found to be crucial for cell fate determination.

Results and discussion

Dynamic chromatin changes over developmental stages.
Epigenome mapping is a powerful method for cataloging functional elements throughout the genome[29], and it can provide insights into the regulatory mechanisms that underlie changes of cell fate[30]. To investigate the mechanisms underlying cell fate determination, we applied ATAC-seq datasets and standard data analysis pipeline (Figure S1) of six species (*A. thaliana*, *C. elegans*, *H. sapiens*, *M. musculus*, *D. melanogaster*, *S. cerevisiae*) as they have emerged as most appreciated models for system biological research. The detailed information about the ATAC-seq samples that we used were listed in Table S1 and S2. Firstly, we checked quality of all the raw materials, and the results showed that the insert size distribution of each ATAC-seq library displays a stereotypical 150 bp periodicity that consistent with the expected nucleosome occupancy of chromatin. However, the nucleosome occupancy of *Arabidopsis* was not so obvious, as plants have mitochondrial and chloroplast genomes, which are completely accessible to Tn5, and likely depletes Tn5 activity from the nuclear genome[31] (Figure S2). Then, we checked the number of reads mapped to each chromosome (Fig. 1D, Fig. S5). The result showed highly similar reads distribution pattern, indicating of high sample quality. We designed a stringent computational framework to integrate all the samples from different species with unified parameters, resulting in the identification of 25,000–65,000 high-confidence, accessible peaks for *Arabidopsis*, 30,000–54,000 for *Drosophila*, 28,000–1,250,000 for human, 8000–650,000 for mouse, 2000–3500 for yeast, and 15,000–28,000 for nematode (Fig. 1A, Fig. S3). Examination of peak signals versus uniquely mapped reads revealed that the signal enrichments consistently plateau at greater sequencing depths (Fig. 1B, Fig. S4).

To investigate dynamics chromatin changes over different developmental stages of each species, we used deepTools2 software[32]. Visualization of all the ATAC-seq datasets revealed that with the developmental stages proceeding, most peaks were in promoter-TSS (transcription start site) region (Fig. 1C,E, Figs. S6, S7) indicating these binding sites were predominantly located around TSSs indicating these regions are critical for TF binding and transcription regulating. Histone modifications, function as a prerequisite for dynamic chromatin state changes allow perpetual diversification of epigenome[33]. We found that H3K4me3 and H3K27ac modifications were associated with relatively higher peak density, compared to H3K4me1 and H3K27me3 with low density (Fig. 1F, Fig. S8). Because previous studies have demonstrated that H3K4me3 and H3K27ac were commonly associated with the activation of transcription[34] and mark spot of active enhancers[35], respectively, however, H3K4me1 and H3K27me3 were associated with transcriptional silencing and downregulation of nearby genes[36,37].

Taken together, these findings showed comparable open chromatin landscapes in early embryos, as early embryo samples tend to enrich more accessible signals compared to mature tissue samples.

Chromatin accessibility extends the dictionary of *cis*-regulatory elements.
In a comparison of open chromatin among epigenomes of human, mouse, *Drosophila*, worm, *Arabidopsis*, and yeast, we found the genomic distribution of THSs in each were highly similar, as majority of the peaks were enriched in promoter regions, except for human and mouse samples (Fig. 2A, Fig. S9). However, more than 90% of THSs lie outside of transcribed regions, and the majority of these THSs were found within 3 kb upstream of TSS in all species but for human, mouse, and fruit fly. The differences in reads distribution between advanced organisms and relatively lower livings may due to the fact that transcriptional regulatory elements (TREs) in plants and microbes are generally less numerous and closer to the genes they regulate than those of advanced genomes. For example, the
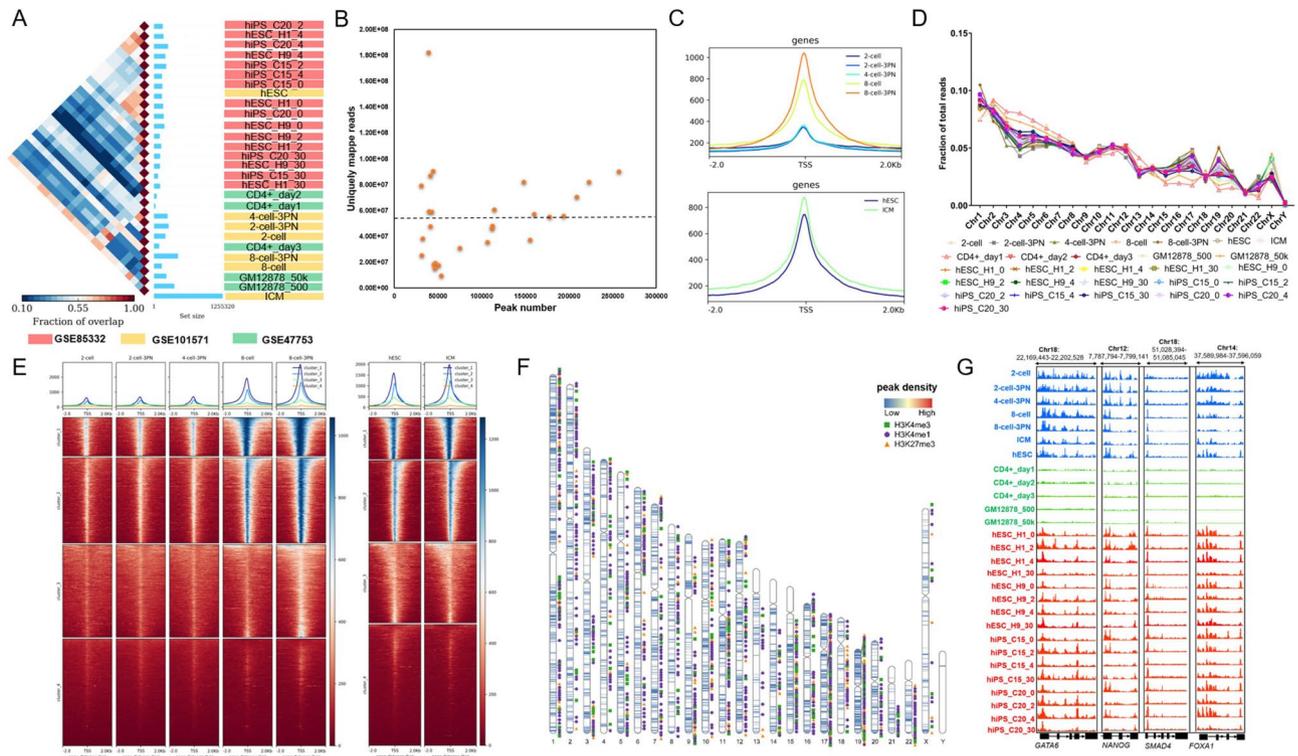
Figure 1. Accessible chromatin demonstrates the epigenetic dynamics across different developmental stages. (**A**) Pairwise analysis of peaks presented in ATAC-seq samples. Left, heatmap demonstrates the overlapping rate between peaks in each sample, right, histogram showing peak number identified in each sample. (**B**) Called peak counts for 28 human ATAC-seq datasets as a function of the number of uniquely mapped reads used for peak calling. (**C**) The average ATAC-seq enrichment of active genes around TSS region. The center of accessible regions was used to produce the distribution plots. The upstream and downstream regions (2 kb) of TSS are mappable. Only part of human samples was presented. Full of the enrichment plots were presented in Fig. S7. (**D**) Distribution of reads mapped to genome of human samples. (**E**) Average plots and heatmaps of ATAC-seq signals at ATAC-seq transposase hypersensitive sites (THSs). The regions in the heatmaps are ranked from highest ATAC-seq signal (top) to lowest (bottom). The cluster manually set to 4. (**F**) Distribution of peaks and DNA methylation marks in chromosomes. Peak density was calculated by average peak counts divided by peak length (kb). Only human plot was showed, the plots of other species were presented in Fig. S8 (**G**) The IGV views showing the ATAC-seq enrichment near key cell-fate-determined TFs during early embryogenesis.

median distance the enhancer and the TSSs of their target genes in fruit fly was reported to be 10 kb[38]. It was also uncovered that in human T cells, 91% of the enhancers, with a median distance of 130 kb between promoters, far greater than the distance across plant and fungi genome[39]. Interestingly, compared to embryonic stem cells, more peaks were enriched in the CD4+T cell promoter regions. Previous research has demonstrated that more promoters marked by H3K4me3 during early phase of CD4 T cell activation, which could enhance accessible chromatin status and reinforce activation-induced upregulation of gene expression[40]. Overall, it is clear that in all species the majority of THSs are located in promoter regions, suggesting that accessible chromatin overlaps extensively with putative *cis*-regulatory sequences.

While assessment of open chromatin plays an important role in predicting regulatory element in the genome[29,41], it does not provide direct information of functional activity. To address this issue, we asked whether shared open chromatin regions exist within all the datasets in each species, and what the role of regulatory elements that bound to these shared regions played. The results showed that there were 1082 shared peaks for *A. thaliana*, 7519 for *C. elegans*, 7970 for *D. melanogaster*, 1995 for *S. cerevisiae*, 8036 for *H. sapiens*, and 4151 for *M. musculus*. Functional enrichment annotation of these overlapped peaks shows that they were almost enriched in transcription and regulation of transcription (Fig. 2B, Fig. S10), indicating that majority of TFs that bound to accessible regions are involved in basal transcriptional activity. Then, we evaluated the ATAC-seq enrichment profiles of all samples in each species, the results showed differential ATAC-seq signals between different tissues/strains or developmental stages, indicating tissue or developmental heterogeneity (Fig. 2C, Fig. S11), except for *C. elegans*, which showed comparable signals in different developmental stages. And the high correlation of ATAC-seq signal between each of the sample (Fig. 2C, Fig. S11) demonstrates the high reproducibility.

Collectively, these results suggest that TREs tend to be focused near the promoter rather than at more distal regions. The hypothesis implicit that open chromatin site near a TSS reflect TREs that regulate TSS rather than more distal regions, and that promoter-binding upstream elements contribute the majority of regulatory effects. And the TREs bound to promoters generally are in basal transcription and regulation of transcription function. Interestingly, these assumptions were previously validated in wet lab showing that an upstream fragment of
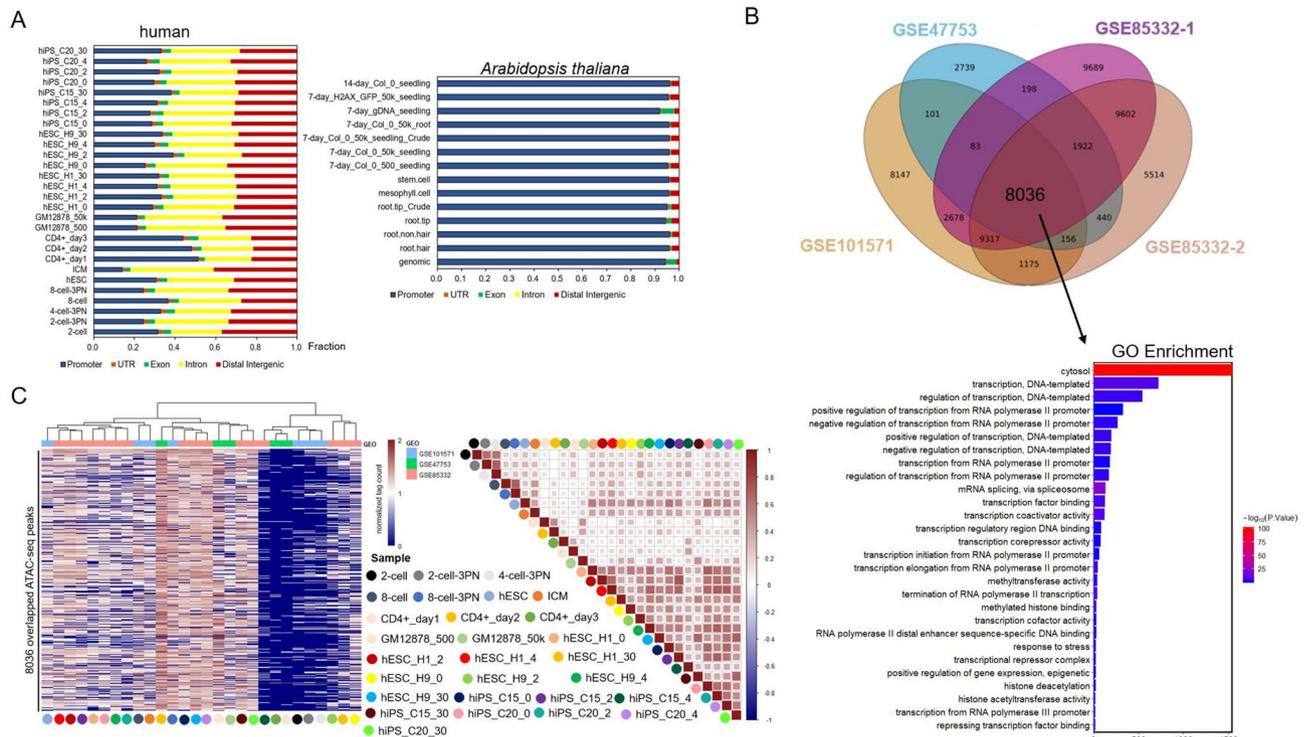
Figure 2. Genomic and functional annotation of accessible regions. (**A**) Genomic distributions of enriched accessible regions identified in ATAC-seq samples. THS peaks within TSS ± 3 kb are considered as promoter THS, and those not located in promoters, exons, introns, or UTRs are labeled as distal intergenic. (**B**) GO functional enrichment analysis of overlapped peaks. Upper part integrates Upset and Venn method to identify overlapping peaks across all human samples. Bottom part is a bar plot of GO enrichment of overlapping peaks. (**C**) Heatmaps showing the ATAC-seq enrichment (RPKM) (left) and the comparison of ATAC-seq signal within consensus ATAC-seq peaks by Pearson's Coefficient Correlation algorithm. The colored bubbles represent different samples. The ATAC-seq enrichment signals were normalized by log2(FPKM + 1).

several kilobases is capable of recapitulating native transcription patterns[42–44], which are consistent to our findings that upstream THSs are the most abundant category of accessible chromatin sites.

Taken together, these data indicated that THSs in animal and plant genomes showed largely comparable landscapes and *cis*-regulatory elements that bound to open chromatin regions mainly play roles in transcription and regulation of transcription.

Identification of cell-fate TFs during early embryogenesis. We applied HOMER *findmotif* to determine what TFs that bound to these open chromatin regions. And 400 for *A. thaliana*, 38 for *C. elegans*, 93 for *D. melanogaster*, 109 for *S. cerevisiae*, 414 for *H. sapiens*, and 398 for *M. musculus* were identified (Fig. 3A, Table S3). Gene family classifications showed that majority of the identified TFs were enriched in Homeobox and C2H2 zinc finger family (Fig. 3B, Fig. S12). Motif discovery indicated that, *PIF4*, *PCF*, *BIM1*, and *JKD* genes were highly enriched for root and seedlings in *Arabidopsis*, *elf-1*, *hlh-30*, *dpl-1*, *eor-1*, *pha-4*, and *pqm-1* were highly enriched during larva development in *C. elegans*, *zld*, *Dref*, and *Trl* for *D. melanogaster* during nuclear cycle period, *ABF1*, *REB1*, *AZF1*, *OPI1*, and *RSC3* for different strains of *S. cerevisiae*, *CTCF*, *BORIS*, *SOX2*, *NFYA*, *SP1*, *OCT4*, and *NANOG* for *H. sapiens* during embryonic stem cell development, and *JunB*, *Batf*, *Nanog*, and *AP-1* for *M. musculus* during induced pluripotent stem cell development (Fig. 3C, Fig. S13). To investigate the functions of these TFs, we performed functional GO (gene ontology) analysis. The results showed that these TFs were almost involved in transcription and regulation of transcription. Interestingly, we also found that some TFs were TFs involved in cell fate decisions (cell fate commitment, and cell fate specification) (Fig. 3C, Fig. S13).

To investigate the expression patterns of these TFs across different tissues or strains, we collected expression profiles from public databases. For *Arabidopsis*, the TFs such as *GATA1*, *TCP3*, *CDF3*, *PIF4*, *CCA1*, *LHY*, *SPL1*, and *MYB38* were highly expressed in different mature tissues, and *ABI5*, *WUS*, *HB5*, *WIP5*, and *SHP1* were lowly expressed even unexpressed (Figure S13), as these genes are involved in nuclear cycle or early embryonic development[8,45–48]. For *C. elegans*, all identified TFs were deemed to differentially expressed across multiple strains (Figure S13). For *Drosophila*, some TFs such as *Kr*, *bcd*, *zen*, *cad*, and *twi* hardly expressed in various mature tissues, as these TFs previously supposed to play major role in early embryonic development of *Drosophila*[49–53]. For *S. cerevisiae*, it is obvious that the identified TFs were differentially expressed across all yeast strains and were higher in strain w303a than other strains (Figure S13). For *H. sapiens*, we found some TFs, *PAX6*, *SOX2*, *POU3F1*, *HOXA10*, *CDX2*, *NANOG*, *TEAD4*, and *OCT4* were scarcely expressed across mature tissues, because they function in early stem cell development[54–57]. For *M. musculus*, we also found some TFs,
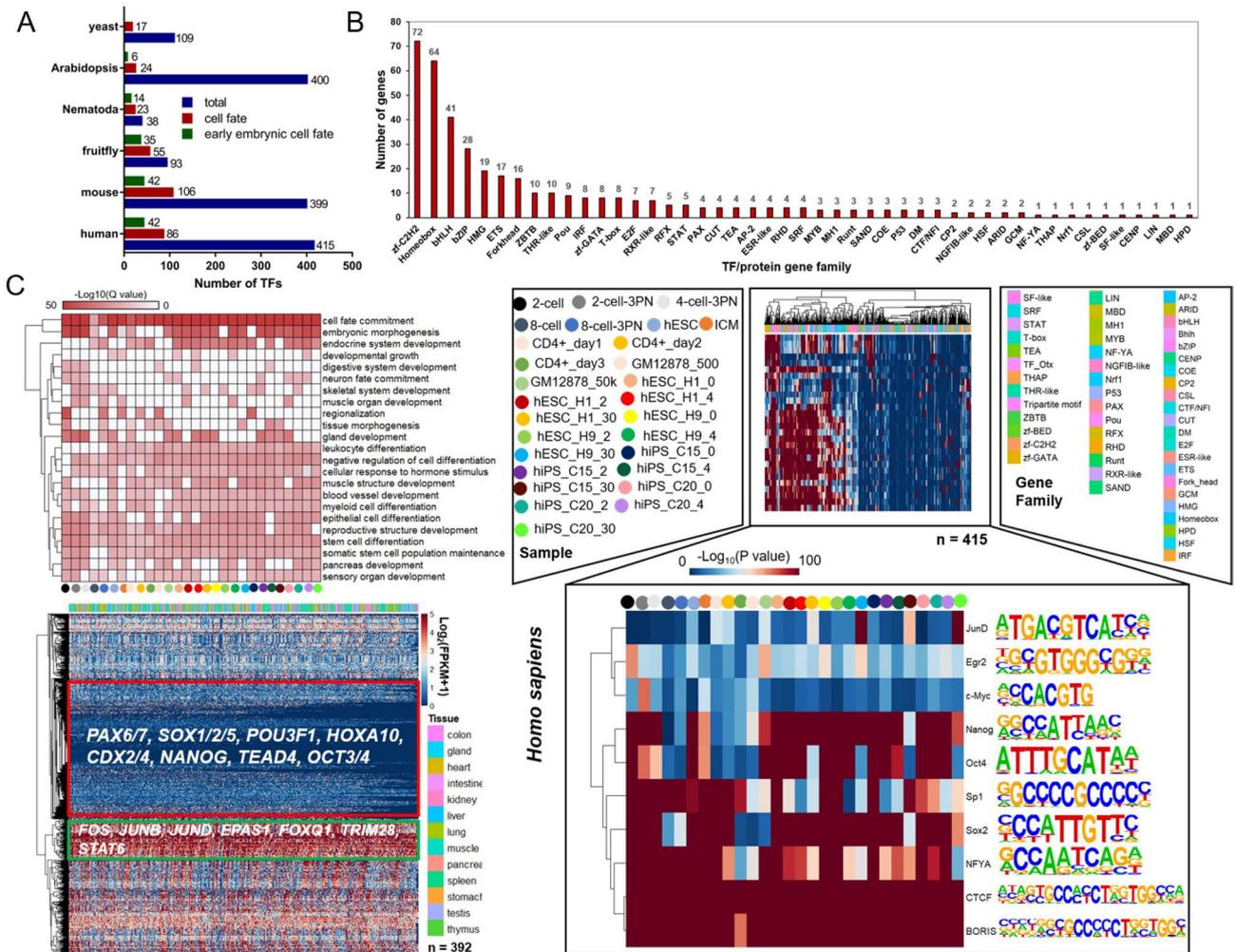
Figure 3. Classification and functional annotation of TFs identified from ATAC-seq samples. (**A**) Classification of TFs that identified from all ATAC-seq samples. (**B**) Gene family classification of the identified TFs in human. The gene family information were collected from JASPAR (http://jaspar.genereg.net/). The information of other species were listed in Fig. S12. (**C**) Functional and expression pattern analysis of TFs identified from ATAC-seq samples. Right part, annotation of all TFs and randomly selected 10 for motif analysis. Top left, heatmaps showing the top 20 enriched GO terms of all TFs using Metascape enrichment. Bottom left, expression patterns of all TFs across different tissues. The expression profiles were obtained from ENCODE database (https://www.encodeproject.org/). And the raw expression matrices were normalized by log2(FPKM + 1).

Cdx2, *Oct4*, *Eomes*, *Esrrb*, *Gsc*, and *Nanog*, were scarcely expressed in mature tissues, as these TFs constitute an important reservoir for early embryonic development[57–59].

Additionally, we found a set of TF complexes which were pertinent to cell proliferation (*Oct4::Sox17*, promotes cell development and differentiation[60], and *OCT4-SOX2-TCF-NANOG*, forms core regulatory circuitry of ES cells, critical for pluripotency and self-renewal[61]), cell differentiation (RAR/RXR, triggers pluripotent cell differentiation[62], *NF1::FOXA1*, mediate gene expression and cell differentiation in prostate[63]), tumorigenesis, and immunogenesis (Figure S14), demonstrating that even in early embryos, these TFs that involved in oncogenesis, and tumor suppress are also expressed to maintain normal cell divisions and differentiation of early embryo.

To further validate some key TFs that function in early embryos, we visualized the ATAC-seq signal enrichment near them across all samples in each species using IGV (Integrative Genomics Viewer). The results showed that for the developmentally regulated genes, such as *GATA6*, *NANOG*, *SMAD4*, and *FOXA1* in human, were found elevated ATAC-seq enrichment at annotated or putative enhancers and promoters during embryonic development instead of in mature cells (Fig. 1G). For TFs such as, *Oct4*, *Sox4*, *Eomes*, and *Gata4*, we also observed increased signals during mouse embryonic development (Figure S15), which comparable to *Su(H)*, *zen*, *Abd-B*, and *twi* during *Drosophila* nuclear cycle (Fig. S15). However, for *WUS*, *ATML1*, *JKD*, and *KAN* in *Arabidopsis*, they showed distinct signal intensity over different tissues or under different treatments (Figure S15). Nevertheless, *FKH1*, *STE12*, *MSN2*, and *DIG1* in *S. cerevisiae* showed comparable signal intensity over different strains (Figure S15).

Overall, by integrating the information of cell-fate-determined TFs and the transcriptomes, we delineated that these cell-type-specific TFs showed high tissue or developmental heterogeneity.
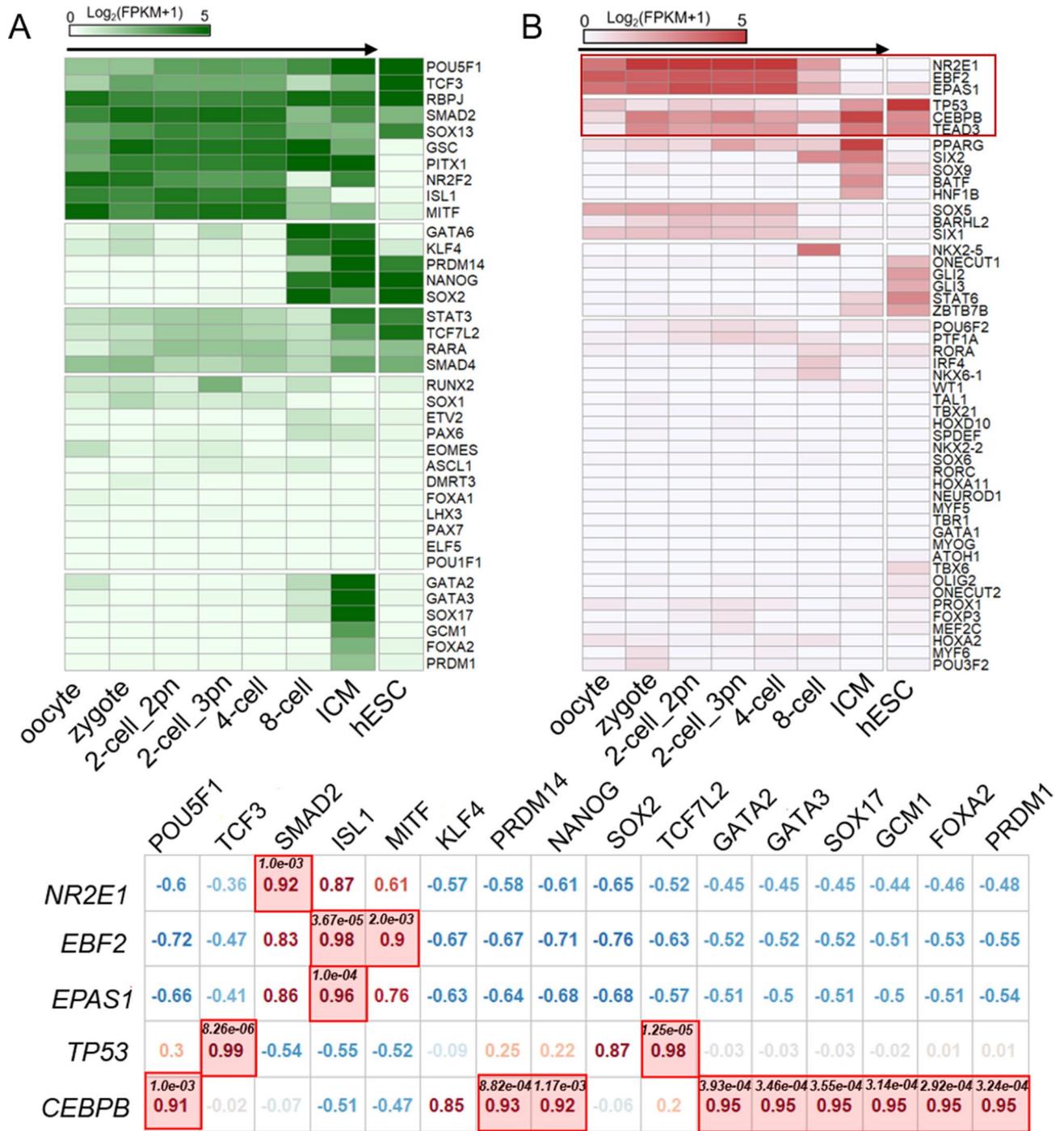
Figure 4. Expression profiles of TFs involved in early cell fate determination. (**A**) The expression patterns of stat-of-the-art early cell fate determined TFs during early embryonic development. (**B**) Exhibiting the expression profiles of cell-fate TFs during early embryogenesis. Bottom, showing the correlations of expression pattern between public-accepted TFs of early cell-fate determination and some that we identified in this research. The normalized expression matrices were collected from NCBI database from GSE101571 accession.

Regulatory networks of cell-fate decision in early embryo. Cell fate decisions play a key role in crucial processes such as tissue repair, immune response, or embryonic development[64–66]. Here, we identified numerous TFs that are involved in cell fate control (Fig. 4, Fig. S16). For each species, using public expression profiles of early embryonic development, we have not only verified high expression values of some previously widely accepted early embryonic TFs (Fig. 4A), but also found some cell-fate determining TFs that were highly expressed that previously unreported during early embryonic development (Fig. 4B, Fig. S16), indicating they may play roles in early embryos. However, we also found some TFs previously reported to play major roles during early embryogenesis in *Drosophila*, had a low expression pattern during early embryonic stage, such as *pnr*[67],

vnd[68], and *Ubx*[69] (Figure S16). The expression profiles of some previously unreported TFs that function in early embryogenesis were also have high correlations with some early embryonic TFs, such as *Jra*, *Blimp-1*, *hth*, and *Tk* in *Drosophila*, *NR2E1*, *EBF2*, *EPAS1*, *TP53*, and *CEBPB* in human.

To comprehensively resolve the mystery of regulatory mechanisms of cell fate control during early embryogenesis, we combined cell-fate TFs of six species to construct TF regulatory network to predict the regulatory circuit based on their homology relationships (Figure S17). And, we investigated some homolog TFs in other five species for human TFs (Table S4). Surprisingly, these homolog TFs are also previously reported to be involved in cell fate control during early embryonic development.

For *Arabidopsis*, we analyzed several key TFs, which play key roles during root epidermis patterning, seeding, leaf, and QC (quiescent center) development in details (Fig. 5A). Four cell-fate-determining TFs, *JKD*, *GL2*, *GL3*, and *EGL3*, which are homolog to *PRDM14*, *GSC*, and *MITF* in human, respectively, are indispensable for controlling the patterns of epidermis in the *Arabidopsis* root meristem[70]. *HAT3*, homologs to *PAX6* and *NANOG*, combined with *HAT2*, *BZR1*, and *BIM1* to promote seedling development[70–72]. We assume that *BIM1* may play a role as a signal integrator to integrate signals from *HAT2*, *HAT3*, and *BZR1* to promote seedling development (Fig. 5A). Another cluster of TFs, *KAN*, *PHB*, *PHV*, and *BIM1*, in which *PHB* and *PHV* are homolog to *ISL1*, contribute to promote *Arabidopsis* leaf development and leaf adaxial polarity[73,74]. We hypothesize that *BIM1* may function as downstream target genes of *PHV* to regulate leaf development (Fig. 5A). The last cluster TFs that we found have homologs of human early embryonic TFs are *HDG11*, *KAN*, *WUS*, *PLT1*, and *WIP4*, in which *HDG11* and *WUS* are homolog to *POU5F1* and *ASCL1*, respectively. Previous studies have demonstrated that *WUS*, *PLT1*, and *WIP4* all contribute to the cell-fate determination of QC[47,75,76]. Therefore, we conject that *HDG11* and *KAN* may function as upstream target genes of *WUS* to form the *HDG11-KAN-WUS-PLT1-WIP4* complex to control the cell-fate determination of QC. The expression value of all cell-fate-determined TFs mentioned above all keep high levels during *Arabidopsis* early embryonic development (Fig. 5B), indicating they may play major roles in *Arabidopsis* early embryos.

For *C. elegans*, we identified several early embryonic TFs, *hlh-2*, *pha-4*, *elt-1*, *hlh-1*, and *pal-1*, in which the former three are homolog to *TCF3*, *FOXA1/2*, and *GATA2/3/6*, respectively (Fig. 5C). Previous demonstrated *elt-1* and *pal-1* are critical for the specification of epidermal cell fates[77,78], furthermore, in our study, we presumed that *hlh-2* and *hlh-1* may act as upstream target TFs of *pal-1*, and *pha-4* functions as binding protein of *elt-1*, and these five TFs function together to control epidermal cell fate. The expression patterns of these five TFs showed highest levels at 16-cell stage (Fig. 5D), indicating 16-cell stage may be a critical timepoint for epidermal cell fate determination.

As the propagating method of *S. cerevisiae* is budding reproduction without embryo development, we analogously regarded the cell cycle period as embryonic development stage. We identified several TFs, *RIM101*, *FKH1*, *FKH2*, *MSN2*, *MSN4*, *ABF2*, *DAL80*, and *CBF1*, which homolog to *FOXA1*, *FOXA2*, *KLF4*, *PRDM14*, *SOX2*, *SOX17*, *GATA2/3/6*, and *MITF*, respectively (Fig. 5E). Previous studies showed that these TFs were all involved in stress responses[79–81]. So, we proposed a regulatory circuit that regulate the progressive process of stress responses based on their interaction relationships (Fig. 5E). And, we noticed these TFs kept high expression values during the full stages of cell cycle (Fig. 5F), indicating that yeast is susceptible to external or internal damages, TFs that regulate the defense systems need to be constantly functioning.

For *D. melanogaster*, we identified several clusters of TFs that involved in eye, gland, and nerve system cell fate determination (Fig. 5G). Firstly, a cluster of TFs, *zen*, *tll*, *toy*, *pnr*, *Mad*, *Med*, and *lz*, which homolog to *NANOG*, *NR2F2*, *PAX6*, *GATA2/3/6*, *SMAD2*, and *SMAD4*, respectively, were previously reported to be involved in pattern formation of eye cell fate[82–84], we proposed the model for eye cell fate decisions, *lz-Med-Mad-pnr-toy-tll-zen*, in which *lz* bound to *Med*, and *Med* bound to *Mad*, to promote the expression of *Mad*, as a research have shown that *lz* encodes a TF involved in prepatterning photoreceptor precursors in the *Drosophila* eye[85]. Then, *prd*, *fkh*, *pnr*, and *pan*, in which the latter three are homolog to *FOXA1/2*, *GATA2/3/6*, *TCF7L2*, and *SOX13*, respectively, were demonstrated to be indispensable for gland cell fate determination[86–90], we proposed a regulatory model, *pan-pnr-fkh-prd*, in which pnr served as a binding protein, which bound to *fkh*, and prd may function as terminal target gene. Overall, these TFs function together to promote gland cell fate determination and cell development. Thirdly, *su(Hw)*, *ac*, *Su(H)*, *Kr*, *tll*, *D*, *pnr*, *pnt*, and *gcm*, homolog to *PRDM14*, *ASCL1*, *RBPJ*, *KLF4*, *NR2F2*, and *SOX1/2*, were reported to constitute important reservoirs for nervous system cell fate decisions[91–93]. Based on the interactions of these TFs, we proposed a regulatory circuit, which is required for the cell fate control and development of nervous system (Fig. 5G). Expression profiles of these TFs showed that they started to play functions from nuclear cycle stage 14, regardless of some had high expression values from the very beginning of the nuclear cycle (Fig. 5H).

We have identified 24 human homologue TFs in mouse (Fig. 5I). Similar to humans, these TFs also function to determine cell fate or promote cell development during early embryonic developmental stage. Furthermore, the interactions of TFs were almost the same, and the expression patterns of these TFs were slightly different (Fig. 5J).

Taking together, by integrating the information about *cis*-regulatory elements and the transcriptomes, we scratched the surface of cell-fate-determining regulatory networks during early embryonic development that is orchestrated by a set of TFs and their targets.

Evolution of cell fate decision in early embryos. To further investigate the evolution characteristics of cell fate decisions, we focused on *YOX1*, a key G1/S transition regulator in yeast. We discovered a cluster of homeobox TFs in human, mouse, fruit fly, nematode, and *Arabidopsis* that orthologous to *YOX1*. The orthologous TFs in former four species are involved in neuron cell fate determination, while in *Arabidopsis* promote epidermis cell fate decision (Fig. 6A). Expression profiles of these TFs showed relatively high expression values during early stage of embryonic development, indicating cell fate decision events may occur during early embry-

Figure 5. Transcriptional regulatory networks underlying early embryonic cell fate determinations. (**A**) TFs mainly play roles in *Arabidopsis* root epidermis, seedling, leaf, and QC (quiescent center) development, orange, blue, and red lines indicated interactions that control seedling, leaf, and QC cell fates, respectively. Green dashed lines represent homology relationship between *Arabidopsis* TFs and human TFs. (**B**) Expression profiles of these *Arabidopsis* TFs during early embryogenesis. The x-axis represents different stages of *Arabidopsis* early embryonic development, which were annotated in Figure S16. (**C**) TFs determine the cell fate of *C. elegans* epidermis. Orange dashed lines represent homology relationship between nematode TFs and human TFs. (**D**) Expression patterns of these five TFs during *C. elegans* early embryogenesis. (**E**) TFs involved in pattern formation of stress responses in yeast. Purple dashed lines represent homology relationship between yeast TFs and human TFs. (**F**) Expression patterns of these eight TFs during *S. cerevisiae* cell cycle. (**G**) TFs participate in cell fate determination of eye, gland, and nerve system in fruit fly. Orange lines demonstrate TF interactions that control nerve system, red lines demonstrate TF interactions that control gland development, green lines demonstrate TF interactions that control eye cell fate. The red dashed lines represent homology relationship between fruit fly TFs and human TFs. (**H**) Expression patterns of these 21 TFs during *D. melanogaster* cell cycle. (**I**) Regulatory circuits determine mouse and human early embryonic cell fate. (**J**) Expression patterns of these TFs during early embryogenesis of mouse and human. The interaction relationships were all predicted by STRING database. The normalized expression profiles were collected from NCBI database accessions of GSE101571 (human), GSE66582 (mouse), GSE25180 (fruit fly), GSE77944 (nematode), GSE123010 (*Arabidopsis*), and GSE104904 (yeast).

ogenesis. While distinct expression patterns across species during embryonic development might be a cue for the differences in the determination of different cell fates (Fig. 6B). Phylogenetic analysis of protein sequence of these homologous TFs indicated that TFs in yeast and Arabidopsis are more ancient in evolution, compared to that of advanced organisms (human, mouse, fruit fly, and worm) (Fig. 6C). Because yeast and Arabidopsis have experienced an ancient whole-genome duplication event[94]. We further investigated whether consensus sequence of these TFs shared. And the result showed a motif in homeobox domain from residues 20 to 59 was conserved, indicating the critical functions in cell fate decision events (Fig. 6D).

Then, we constructed an integrated network to investigate the transcriptional regulatory functions of above TFs in cell fate decision events (Fig. 7). *YOX1*, a TF expressed in mid-G1 through early S phage, interplays with S-specific TF—*YHP1*, function as transcriptional repressor to negatively regulate *MCM1-FKH2-NDD1*-mediated G2/M-G1 transition during cell cycle progression[95]. Recently, a research has reported that *ROX1* is in promotion of *RAP1-HAP1-MSN4* module, which is an important branch for G2/M to G1 phase transition in yeast[96]. By homology analysis, we found *YOX1* was homologous to three TFs (*Pou5f1*, *Nanog*, and *Pax6*) in mouse, which were previously reported to be involved in restriction of a cluster of neuron identity maintainers (*Sox1*, *Sox2*, *Sox17*, and *Tcf7l2*), which were homologous to *ROX1*[97,98]. These maintainers, in turn, suppress the expression of neural differentiation effectors, including *Irx1*, *Irx2*, *Zic1*, and *Zic2*[99–101]. Nevertheless, the ortholog of *YOX1* in worm, *unc-86* were involved in activation of and interplay with several TFs, including *vab-3*, *ttx-3*, and *mec-3*, to define neuron identity[102,103]. For fruit fly, *YOX1* was orthologous to *vnd*, which were reported to interplay with *ind*, *D* (*Dichaete*), and *msh* to regulate neuroblast cell fate[104]. There are two models were proposed to regulate neuroblast cell fate determination, achaete–scute complex[105] and 'neuroblast clock'[106], in which former one acts as proneural cluster and was activated by *vnd* to promote neuroblast formation[107]. We hypothesized that interactions of *vnd*, *ind*, *D*, and *msh* may also positive regulate neuroblast clock model in the manner as achaete–scute complex. Meanwhile, homologs of *YOX1* in *Arabidopsis*, *GL2* and *HDG11*, interact with each other, play an intermediate role of a positive feedback loop to promote epidermis cell fate determination[108]. *GL2* was positively regulated by upstream complex, called *WER-GL3/EGL3-TTG* transcriptional complex[109,110], in turn, leading to the activation of downstream target gene *MYB23*[109]. Then, *MYB23* interact with *WER-GL3/EGL3-TTG* complex to form a positively regulatory loop[110].

Above all, different (epidermis) or similar (neuron) cell fate decision events among different species not only depend on sequence characteristics and expression patterns of TFs but also the roles they played in regulatory networks. And conserved motifs may contribute to their conserved functions in different species.

Methods

ATAC-seq data acquisition.
The raw ATAC-seq datasets were collected from NCBI Gene Expression Omnibus (GEO). We selected data from *Arabidopsis thaliana*, *Drosophila melanogaster*, *Homo sapiens*, *Mus musculus*, *Saccharomyces cerevisiae*, and *Caenorhabditis elegans*, as these species are well-annotated models for transcriptional regulation researches. For *Arabidopsis* samples, ATAC-seq data were collected from GSE101940[111], GSE89346[112], GSE101482[113], and GSE85203[31]. For *Drosophila* samples, ATAC-seq data were collected from GSE104957[114] and GSE83851[115]. For human samples, ATAC-seq data were collected from GSE101571[116], GSE85332[117], and GSE47753[28]. For mouse samples, ATAC-seq data were collected from GSE110264[118], GSE79230[119], GSE82010[120], and GSE67298[121]. For yeast samples, ATAC-seq data were collected from GSE111815[122], GSE101290[123], and GSE66386[124]. For worm samples, ATAC-seq data were collected from GSE114439[125] and GSE98758[126].

ATAC-seq data analysis.
The raw ATAC-seq datasets from six difference species (human, mouse, *A. thaliana*, fruitfly, *C. elegans*, and yeast) were trimmed via trim-galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), with parameters – q 20 – phred33 – nextera – length 20 – e 0.1 – stringency 3. Then the clean reads were quality-controlled by FastQC (v0.11.7, https://www.bioinformatics.babraham.ac.uk/proje
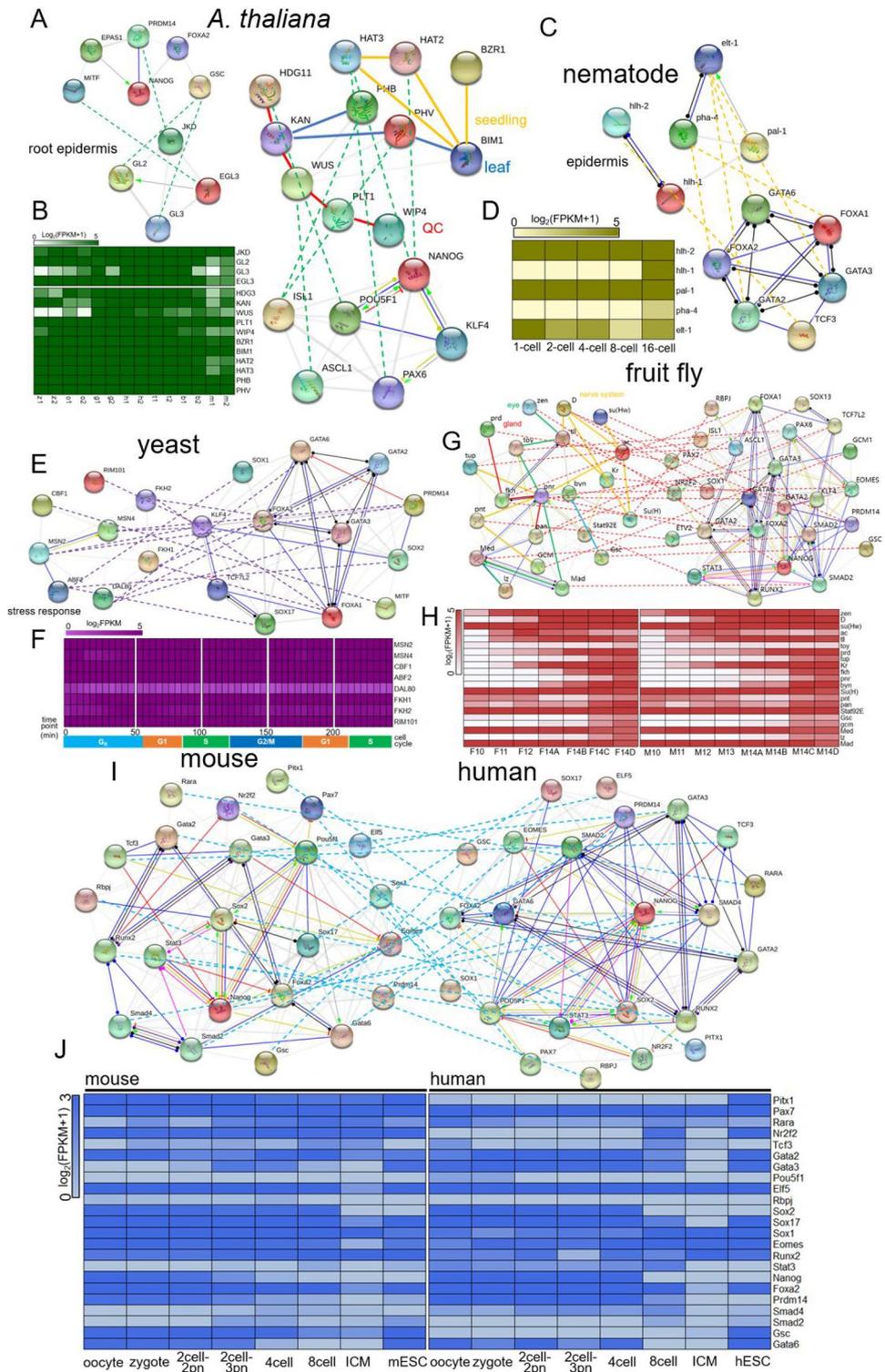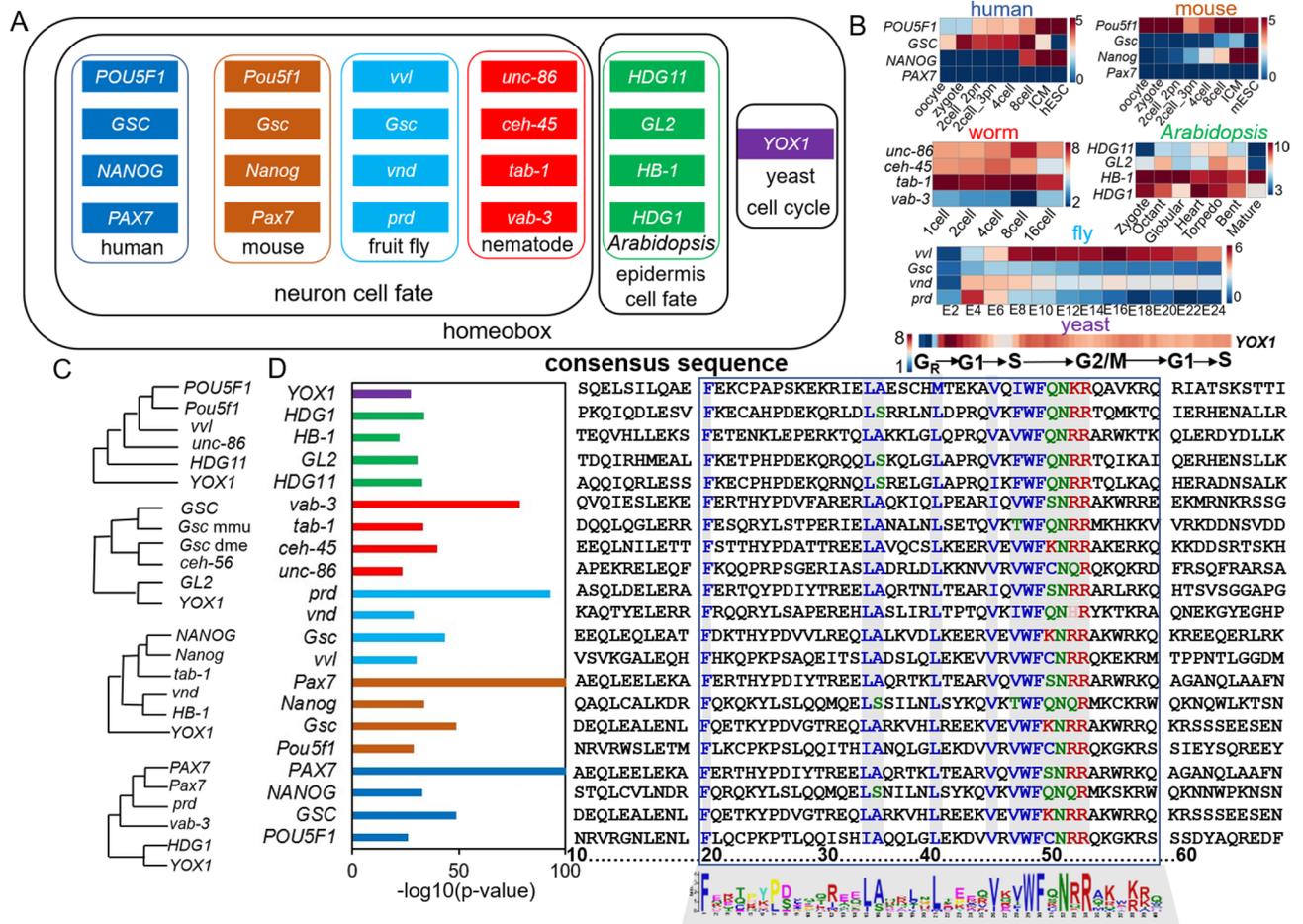
Figure 6. Evolution basis of members of homeobox TFs and their roles in cell fate decisions. (**A**) Key cell cycle transition TFs, YOX1, and its orthologs in five other species play roles in cell fate decision. Homologous TFs of YOX1 in human, mouse, fruit fly and nematode are involved in neuron cell fate determination, while in Arabidopsis, control epidermis cell fate. All of the TFs are members of homeobox family. (**B**) Heat maps showing the expression patterns of the cell-fate TFs during early embryogenesis. E2–24 representing 2–24 embryo stages. The normalized expression matrices were collected from NCBI GSE101571 (human), GSE66582 (mouse), GSE77944 (nematode), GSE123010 (*Arabidopsis*), modENCODE (fruitfly), and GSE104904 (yeast). (**C**) Phylogenetic trees of the homeobox TFs across species. Neighbor joining and 500 bootstrap runs were carried out using the protein sequence. (**D**) Multiple alignment of homeobox TFs across species. The left bar showing the adjusted p value of motif corresponding to each TFs. The right part representing the consensus sequences alignment of the corresponding TFs. The standard numbering of a typical HD (homeodomain) with 60 residues starting from 10 (the upstream 9 residues were not shown) is given at the bottom, and the blue-lined box denote the conserved regions from 20 to 59 of HDs.

cts/fastqc/) and MultiQC v 1.5[127]. After quality control, the clean reads from all samples were mapped to corresponding reference genome (hg38, mm10, TAIR10, BDGP6, IRGSP-1.0, and R64-1-1). For paired-end reads longer than 50 bp, bowtie2 v 2.3.4.2[128] was applied with parameter—very-sensitive—maxins 2000, conversely, single-end reads were set to—very-sensitive. Paired-ends short or equal to 50 bp, we adopt bowtie v 1.2.2[129] with − X 2000 − m 1 parameters to allow up to 2 kb fragments to align and only uniquely mapped reads to retain[28]. All unmapped, MAPQ < 30, and PCR duplicates were removed (samtools − F 0 × 4; samtools view q 30; sambamba markdup). Bam files were then converted to bed files and shifted using a custom shell script to reflect a 4 bp increase on the plus strand and a 5 bp decrease on the minus strand as recommended by[28]. Replicate samples were merged by samtools merge[130].

Genomic tracks generation. For normalization and visualization, the sorted, filtered and merged .bam files from each sample were converted to bigwig format using bamCoverage utility in deepTools v3.3.0[32] with parameters –binSize 1 –ignoreDuplicates –skipNonCoveredRegions –normalizeUsing RPKM. The normalized ATAC-seq signal for a scaled region representing each of the genes in our gene subsets plus/minus 2 kb were compiled and plotted using the computeMatrix and plotHeatmap programs from deepTools package. All genomic track visualization was performed using Integrative Genomics Viewer (IGV) v2.4.16[131].
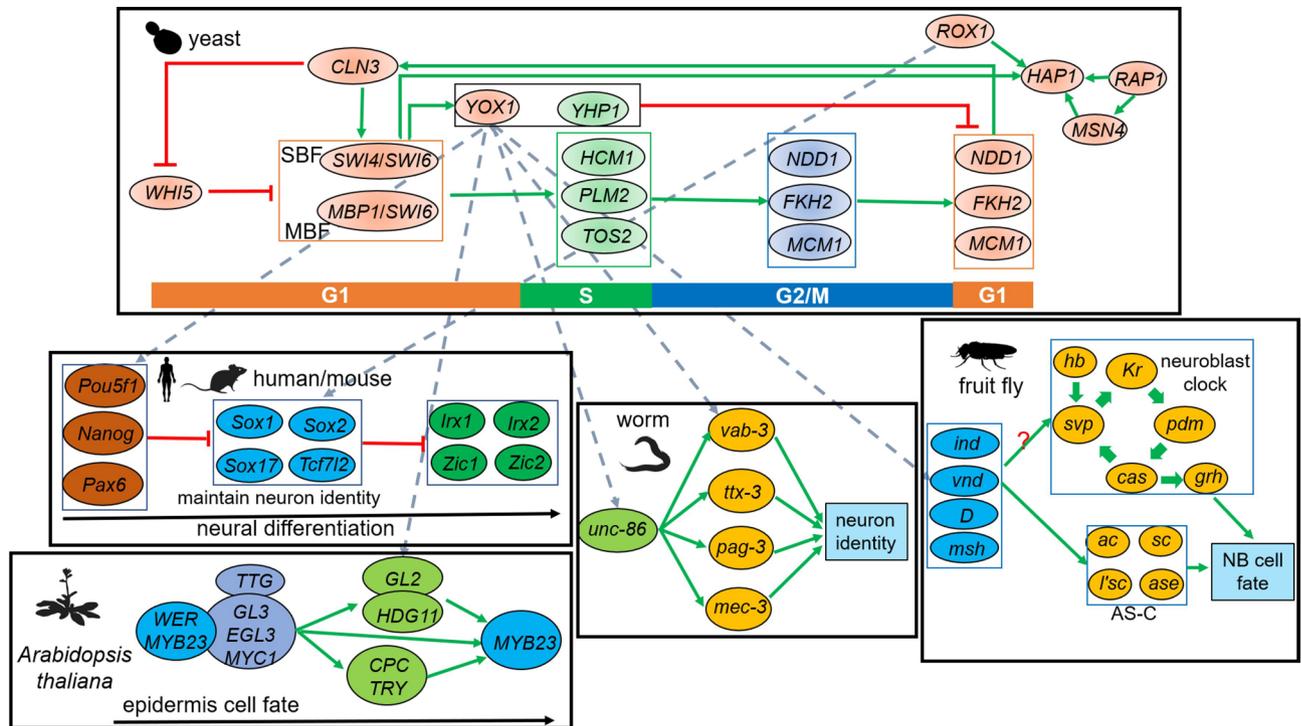
Figure 7. An integrated regulatory network model for the control of cell fate decision events during early stage of embryogenesis/cell cycle. This diagram depicts the regulatory interactions of cell fate determination occurring early embryogenesis/cell cycle. networks in black lined boxes represent regulatory interactions corresponding to each species. Transcription factors are denoted by ovals. Dashed grey lines represent orthologs. Green solid lines represent positive regulation and red lines represent negative regulation. Genes in colored boxes represent gene complexes/interactions. The question mark represents the regulatory relationship was unknown. *AS-C* achaete–scute complex.

Peak calling. Peak calling on ATAC-seq data was performed using MACS2[132] *callpeak* with parameters -g (tair10: 1.1e8, dm6: 1.4e8, hg38: 2.8e9, IRGSP-1.0: 3.7e8, mm10: 2.5e9, sacCer3: 1.2e7) − q 0.05 − extsize 200 − nomodel − shift − 100 − nolambda − keep-dup all. These parameters set a smoothing window of 200 bp between peaks before they are merged into a single peak and allow identification of variable length peaks, respectively.

Expression and correlation of overlapped accessible regions. After peak calling, we summarized the peaks called from each species by Intervene[133]. We counted the number of reads that were enriched in overlapped peak regions by using featureCounts[134]. Peak counts were normalized to $\log_{10}(FPKM + 1)$. Heatmaps of the expression of overlapped peaks were plotted to show differentially expressed peaks in all samples of each species. The count matrix of all the ATAC samples in six species was used to calculate and visualize the Spearman correlation for every sample pair by corrplot[135] package in R.

Peak distribution and functional enrichment annotation. We randomly selected 10,000 peaks and histone modification sites (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K27me3, H3K36me3) in all samples of each species to show the distribution patterns of peaks in chromosomes by RIdeogram[136]. The UCSC genomic annotation was used to associate peaks with different genomic regions. Then we called the annotatePeak function from the R/Bioconductor ChIPseeker[137] package for genomic annotation. Promoters were considered to be ± 3 kb from TSS and all the regions that did not fall within exons, introns, UTRs or promoters were classified as distal intergenic regions. The annotated peaks from ChIPseeker above were functionally enriched by compareCluster function from clusterProfiler[138] package with default parameters.

Transcript factor motif discovery and gene ontology. The peaks generated from ATAC-seq datasets were used for de novo motif analysis using HOMER v4.10[139] against the JASPAR, DMMPMM, Yeast, AthaMap, and Homer databases with parameters − size 400 − len 8,10,12. De novo motifs were retained if the p value < 0.01 and (< percent of target >/< percent of background >) > 1.0. Gene Ontology enrichment for these motifs/transcription factors was performed using Metascape[140]. Those GO terms had a false discovery rate (FDR) of 0.05 or less were considered significant.

Transcriptional regulatory network construction. To explore the transcriptional regulatory basis of six species, we used BLAST to find ortholog genes of human TFs that play cell-fate-choice function, after which

we used Cytoscape to construct a comprehensive network of six species. The regulatory relationships of different TFs were predicted based on STRING and TF2Network databases.

Conclusion

Study of the cell-fate decision across multiple species is still a long way to go, and epigenomic research seems to contribute to some extent. The findings in this study proposed possible molecules for further research of cell-fate determination. We speculate that both the TFs and motifs identified in the integration analysis of this study can be further investigated. Furthermore, the findings presented herein can be correlated with single-cell strategies, such as single-cell RNA-seq and single-cell ATAC-seq in order to uncover the mysterious veil of the evolutionary basis of cell-fate decision events.

References

1. Hamatani, T. *et al.* Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131 (2004).
2. Wang, Q. T. *et al.* A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev. Cell* **6**, 133–144 (2004).
3. Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: Multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* **10**, 467–477 (2009).
4. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
5. Tintori, S. C., Osborne Nishimura, E., Golden, P., Lieb, J. D. & Goldstein, B. A transcriptional lineage of the early *C. elegans* embryo. *Dev. Cell* **38**, 430–440 (2016).
6. Vergassola, M., Deneke, V. E. & Di Talia, S. Mitotic waves in the early embryogenesis of Drosophila: Bistability traded for speed. *Proc. Natl. Acad. Sci USA* **115**, E2165–E2174 (2008).
7. ten Hove, C. A., Lu, K. J. & Weijers, D. Building a plant: Cell fate specification in the early Arabidopsis embryo. *Development* **142**, 420–430 (2015).
8. Boscá, S., Knauer, S. & Laux, T. Embryonic development in *Arabidopsis thaliana*: From the zygote division to the shoot meristem. *Front. Plant Sci.* **2**, 93 (2011).
9. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decision. *Nature* **569**, 345–354 (2019).
10. Rhee, C., Kim, J. & Tucker, H. O. Transcriptional regulation of the first cell fate decision. *J. Dev. Biol. Regen. Med.* **1**, 102 (2017).
11. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
12. Tsankov, A. M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).
13. Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102 (2005).
14. Russ, A. P. *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95–99 (2000).
15. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
16. Chew, J. L. *et al.* Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol. Cell Biol.* **25**, 6031–6046 (2005).
17. Zhang, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol.* **8**, 1114–1123 (2006).
18. Nishioka, N. *et al.* Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283 (2008).
19. Kuckenberg, P. *et al.* The transcription factor TCFAP2C/AP-2γ cooperates with CDX2 to maintain trophectoderm formation. *Mol. Cell Biol.* **30**, 3310–3320 (2010).
20. Ralston, A. *et al.* Gata3 regulates trophoblast development downstream of Tead4 in parallel to Cdx2. *Development* **137**, 395–403 (2010).
21. Ng, R. K. *et al.* Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* **10**, 1280–1290 (2008).
22. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
23. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837 (2016).
24. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
25. Volpato, V. *et al.* Reproducibility of molecular phenotypes after long-term differentiation to human iPSC-derived neurons: A multi-site omics study. *Stem Cell Rep.* **11**, 897–911 (2018).
26. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. USA* **102**, 4936–4942 (2005).
27. Uzkudun, M., Marcon, L. & Sharpe, J. Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. *Mol. Syst. Biol.* **11**, 815 (2015).
28. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
29. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
30. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).
31. Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* **45**, e41 (2017).
32. Ramirez, F. *et al.* deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
33. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
34. Chen, K. *et al.* Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47**, 1149–1157 (2015).
35. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).

36. Cheng, J. *et al.* A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol. Cell* **53**, 979–992 (2014).

37. Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).

38. Kvon, E. Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–95 (2014).

39. Mumbach, M. R. *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).

40. LaMere, S. A., Thompson, R. C., Komori, H. K., Mark, A. & Salomon, D. R. Promoter H3K4 methylation dynamically reinforces activation-induced pathways in human CD4 T cells. *Genes Immun.* **17**, 283–297 (2016).

41. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).

42. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structures and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2005).

43. Andersen, P. K., Lykke-Andersen, S. & Jensen, T. H. Promoter-proximal polyadenylation sites reduce transcription activity. *Genes Dev.* **26**, 2169–2179 (2012).

44. Mikhaylichenko, O. *et al.* The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).

45. Lopez-Molina, L., Mongrand, S., McLachlin, D. T., Chait, B. T. & Chua, N. H. ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J.* **32**, 317–328 (2002).

46. De Smet, I. *et al.* Transcriptional repression of BODENLOS by HD-ZIP transcription factor HB5 in *Arabidopsis thaliana*. *J. Exp. Bot.* **64**, 3009–3019 (2013).

47. Crawford, B. C. *et al.* Plant development. Genetic control of distal stem cell fate within root and embryonic meristems. *Science* **347**, 655–659 (2015).

48. Ehlers, K. *et al.* The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in *Arabidopsis thaliana*. *PLoS One* **11**, e0165075 (2016).

49. Knipple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A. & Jäckle, H. Spatial and temporal patterns of Krüppel gene expression in early Drosophila embryos. *Nature* **317**, 40–44 (1985).

50. Ali-Murthy, Z. & Kornberg, T. B. Bicoid gradient formation and function in the Drosophila pre-syncytial blastoderm. *Elife* **5**, e13222 (2016).

51. Rushlow, C., Colosimo, P. F., Lin, M., Xu, M. & Kirov, N. Transcriptional regulation of the Drosophila gene zen by competing Smad and Brinker inputs. *Genes Dev.* **15**, 340–351 (2001).

52. Kimelman, D. & Martin, B. L. Anterior-posterior patterning in early development: Three strategies. *Wiley Interdiscip. Rev. Dev. Biol.* **1**(2), 253–266 (2012).

53. Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. & Levine, M. dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev.* **6**, 1518–1530 (1992).

54. Terzić, J. & Saraga-Babić, M. Expression pattern of PAX3 and PAX6 genes during human embryogenesis. *Int. J. Dev. Biol.* **43**, 501–508 (1999).

55. Wu, X. *et al.* The POU Domain transcription factor POU3F1 is an important intrinsic regulator of GDNF-induced survival and self-renewal of mouse spermatogonial stem cells. *Biol. Reprod.* **82**, 1103–1111 (2010).

56. Kumar, R. P. *et al.* Regulation of energy metabolism during early mammalian development: TEAD4 controls mitochondrial transcription. *Development* **145**, dev162644 (2018).

57. Göke, J. *et al.* Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. *PLoS Comput. Biol.* **7**, e1002304 (2011).

58. Toyooka, Y., Oka, S. & Fujimori, T. Early preimplantation cells expressing Cdx2 exhibit plasticity of specification to TE and ICM lineages through positional changes. *Dev. Biol.* **411**, 50–60 (2016).

59. Arnold, S. J., Hofmann, U. K., Bikoff, E. K. & Robertson, E. J. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development* **135**, 501–511 (2008).

60. Ying, L., Mills, J. A., French, D. L. & Gadue, P. OCT4 coordinates with WNT signaling to pre-pattern chromatin at the SOX17 locus during human ES cell differentiation into definitive endoderm. *Stem Cell Rep.* **5**, 490–498 (2015).

61. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Gens Dev.* **22**, 746–755 (2008).

62. Chatagnon, A. *et al.* RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Res.* **43**, 4833–4854 (2015).

63. Grabowska, M. M. *et al.* NF1 transcription factors interact with FOXA1 to regulate prostate-specific gene expression. *Mol. Endocrinol.* **28**, 949–964 (2014).

64. de Lázaro, I. & Kostarelos, K. Engineering cell fate for tissue regeneration by in vivo transdifferentiation. *Stem Cell Rev. Rep.* **12**, 129–139 (2016).

65. Petrocca, F. & Lieberman, J. Micromanagers of immune cell fate and function. *Adv. Immunol.* **102**, 227–244 (2009).

66. Takaoka, K. & Hamada, H. Cell fate decisions and axis determination in the early mouse embryo. *Development* **139**, 3–14 (2012).

67. Sandler, J. E. & Stathopoulos, A. Stepwise progression of embryonic patterning. *Trends Genet.* **32**, 432–443 (2016).

68. Reeves, G. T. & Stathopoulos, A. Graded dorsal and differential gene regulation in the Drosophila embryo. *Cold Spring Harb. Perspect. Biol.* **1**, a000836 (2009).

69. Qian, S., Capovilla, M. & Pirrotta, V. The bx region enhancer, a distant cis-control element of the Drosophila Ubx gene and its regulation by hunchback and other segmentation genes. *EMBO J.* **10**, 1415–1425 (1991).

70. Hassan, H., Scheres, B. & Blilou, I. JACKDAW controls epidermal patterning in the Arabidopsis root meristem through a non-cell-autonomous mechanism. *Development* **137**, 1523–1529 (2010).

71. Liu, K. *et al.* ERF72 interacts with ARF6 and BZR1 to regulate hypocotyl elongation in Arabidopsis. *J. Exp. Bot.* **69**, 3933–3947 (2018).

72. Zhang, D. *et al.* Transcription factors involved in brassinosteroid repressed gene expression and their regulation by BIN2 kinase. *Plant Signal Behav.* **9**, e27849 (2014).

73. Prigge, M. J. *et al.* Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. *Plant Cell* **17**, 61–76 (2005).

74. Liang, T. *et al.* UVR8 interacts with BES1 and BIM1 to regulate transcription and photomorphogenesis in Arabidopsis. *Dev. Cell* **44**, 512–523 (2018).

75. Kamiya, N., Nagasaki, H., Morikami, A., Sato, Y. & Matsuoka, M. Isolation and characterization of a rice WUSCHEL-type homeobox gene that is specifically expressed in the central of cells of a quiescent center in the root apical meristem. *Plant J.* **35**, 429–441 (2003).

76. Rovere, F. D., Fattorini, L., Ronzan, M., Falasca, G. & Altamura, M. M. The quiescent center and the stem cell niche in the adventitious roots of *Arabidopsis thaliana*. *Plant Signal Behav.* **11**, e1176660 (2016).

77. Page, B. D., Zhang, W., Steward, K., Blumenthal, T. & Priess, J. R. ELT-1, a GATA-like transcription factor, is required for epidermal cell fates in *Caenorhabditis elegans* embryos. *Genes Dev.* **11**, 1651–1661 (1997).

78. Baugh, L. R. *et al.* The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo. *Development* **132**, 1843–1854 (2005).

79. Serra-Cardona, A., Canadell, D. & Ariño, J. Coordinate responses to alkaline pH stress in budding yeast. *Microb. Cell* **2**, 182–196 (2015).

80. Postnikoff, S. D., Malo, M. E., Wong, B. & Harkness, T. A. The yeast forkhead transcription factors fkh1 and fkh2 regulate lifespan and stress response together with the anaphase-promoting complex. *PLoS Genet.* **8**, e1002583 (2012).

81. Sadeh, A., Movshovich, N., Volokh, M., Gheber, L. & Aharoni, A. Fine-tuning of the Msn2/4-mediated yeast stress responses as revealed by systematic deletion of Msn2/4 partners. *Mol. Biol. Cell* **22**, 3127–3138 (2011).

82. Friedrich, M. Evolution of insect eye development: First insights from fruit fly, grasshopper and flour beetle. *Integr. Comp. Biol.* **43**, 508–521 (2003).

83. Yeung, K. *et al.* Integrative genomic analysis reveals novel regulatory mechanisms of eyeless during Drosophila eye development. *Nucleic Acids Res.* **46**, 11743–11758 (2018).

84. Roignant, J. Y. & Treisman, J. E. Pattern formation in the Drosophila eye disc. *Int. J. Dev. Biol.* **53**, 795–804 (2009).

85. Daga, A., Karlovich, C. A., Dumstrei, K. & Banerjee, U. Patterning of cells in the Drosophila eye Lozenge, which shares syntenic domains with AML1. *Genes Dev.* **10**, 1194–1205 (1996).

86. Xue, L. & Noll, M. Dual role of the Pax gene paired in accessory gland development of Drosophila. *Development* **129**, 339–346 (2002).

87. Myat, M. M., Isaac, D. D. & Andrew, D. J. Early genes required for salivary gland fate determination and morphogenesis in *Drosophila melanogaster*. *Adv. Dent. Res.* **14**, 89–98 (2000).

88. Abrams, E. W. & Andrew, D. J. CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. *Development* **132**, 2743–2758 (2005).

89. Minakhina, S., Tan, W. & Steward, R. JAK/STAT and the GATA factor Pannier control hemocyte maturation and differentiation in Drosophila. *Dev. Biol.* **352**, 308–316 (2011).

90. Zhang, C. U., Blauwkamp, T. A., Burby, P. E. & Cadigan, K. M. Wnt-mediated repression via bipartite DNA recognition by TCF in the Drosophila hematopoietic system. *PLoS Genet.* **10**, e1004509 (2014).

91. Harris, R. M., Pfeiffer, B. D., Rubin, G. M. & Truman, H. W. Neuron hemilineages provide the functional ground plan for the Drosophila ventral nervous system. *Elife* **4**, e04493 (2015).

92. Romani, S. *et al.* Krüppel, a Drosophila segmentation gene, participates in the specification of neurons and glial cells. *Mech. Dev.* **60**, 95–107 (1996).

93. Soustelle, L. & Giangrande, A. Novel gcm-dependent lineages in the postembryonic nervous system of *Drosophila melanogaster*. *Dev. Dyn.* **236**, 2101–2108 (2007).

94. Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the evolution of complexity and diversity?. *Mol. Biol. Evo.* **23**, 887–892 (2006).

95. Cho, C. Y., Kelliher, C. M. & Haase, S. B. The cell-cycle transcriptional network generates and transmits a pulse of transcription once each cell cycle. *Cell Cycle* **18**, 363–378 (2019).

96. Panchy, N. L., Lloyd, J. P. & Shiu, S. H. Improved recovery of cell-cycle gene expression in Saccharomyces cerevisiae from regulatory interactions in multiple omics data. *BMC Genom.* **21**, 159 (2020).

97. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440–454 (2012).

98. Lee, H. K., Lee, H. S. & Moody, S. A. Neural transcription factors: From embryos to neural stem cells. *Mol. Cells* **37**, 705–712 (2014).

99. Becker, M. B., Zulch, A., Bosse, A. & Gruss, P. Irx1 and Irx2 expression in early lung development. *Mech. Dev.* **106**, 155–158 (2001).

100. McMahon, A. R. & Merzdorf, C. S. Expression of the zic1, zic2, zic3, and zic4 genes in early chick embryos. *BMC Res. Notes* **3**, 167 (2010).

101. Luo, Z. J. *et al.* Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol. Cell* **57**, 685–694 (2016).

102. Baum, P. D., Guenther, C., Frank, C. A., Pham, B. V. & Garriga, G. The *Caenorhabditis elegans* gene ham-2 links Hox patterning to migration of the HSN motor neuron. *Genes Dev.* **13**, 472–483 (1999).

103. Altun-Gultekin, Z. *et al.* A regulatory cascade of three homeobox genes, ceh-10, ttx-3 and ceh-23, controls cell fate specification of a defined interneuron class in *C. elegans*. *Development* **128**, 1951–1969 (2001).

104. Seibert, J. & Urbach, R. Role of en and novel interactions between msh, ind, and vnd in dorsoventral patterning of the Drosophila brain and ventral nerve cord. *Dev. Biol.* **346**, 332–345 (2010).

105. Skeath, J. B. & Carroll, S. B. The achaete–scute complex: Generation of cellular pattern and fate within the Drosophila nervous system. *FASEB J.* **8**, 714–721 (1994).

106. Homem, C. C. & Knoblich, J. A. Drosophila neuroblasts: A model for stem cell biology. *Development* **139**, 4297–4310 (2012).

107. Skeath, J. B., Panganiban, G. F. & Carroll, S. B. The ventral nervous system defective gene controls proneural gene expression at two distinct steps during neuroblast formation in Drosophila. *Development* **120**, 1517–1524 (1994).

108. Khosla, A. *et al.* HD-Zip proteins GL2 and HDG11 have redundant functions in Arabidopsis trichomes, and gl2 activates a positive feedback loop via MYB23. *Plant Cell* **26**, 2184–2200 (2014).

109. Schiefelbein, J., Huang, L. & Zheng, X. H. Regulation of epidermal cell fate in Arabidopsis roots: The importance of multiple feedback loops. *Front. Plant Sci.* **5**, 47 (2014).

110. Bruex, A. *et al.* A gene regulatory network for root epidermis cell differentiation in Arabidopsis. *PLoS Genet.* **8**, e1002446 (2012).

111. Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor network. *Plant J.* **94**, 215–231 (2018).

112. Jégu, T. *et al.* The Arabidopsis SWI/SNF protein BAF60 mediates seedling growth control by modulating DNA accessibility. *Genome Biol.* **18**, 114 (2017).

113. Maher, K. A. *et al.* Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* **30**, 15–36 (2018).

114. Haines, J. & Eisen, M. B. Patterns of chromatin accessibility along the anterior–posterior axis in the early Drosophila embryo. *PLoS Genet.* **14**, e1007367 (2018).

115. Blythe, S. A. & Wieschaus, E. F. Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. *Elife* **5**, e20148 (2016).

116. Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).

117. Liu, Q. *et al.* Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. *Circ. Res.* **121**, 376–391 (2017).

118. Cao, S. *et al.* Chromatin accessibility dynamics during chemical induction of pluripotency. *Cell Stem Cell* **22**, 529–542 (2018).

119. Jung, Y. H. *et al.* Chromatin states in mouse sperm correlate with embryonic and adult regulatory landscape. *Cell Rep.* **18**, 1366–1382 (2017).

120. Su, Y. *et al.* Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat. Neurosci.* **20**, 476–483 (2017).

121. Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769–774 (2015).
122. Dronamraju, R. *et al.* Spt6 association with RNA polymerase II directs mRNA turnover during transcription. *Mol. Cell* **70**, 1054–1066 (2018).
123. Gowans, G. J. *et al.* INO80 chromatin remodeling coordinates metabolic homeostasis with cell division. *Cell Rep.* **22**, 611–623 (2018).
124. Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770 (2015).
125. Jänes, J. *et al.* Chromatin accessibility dynamics across *C. elegans* development and aging. *Elife* **7**, e37344 (2018).
126. Kolundzic, E. *et al.* FACT sets a barrier for cell fate reprogramming in *Caenorhabditis elegans* and human cells. *Dev. Cell* **46**, 611–626 (2018).
127. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
128. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
129. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
130. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
131. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
132. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
133. Khan, A. & Mathelier, A. Intervene: A tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinform.* **18**, 287 (2017).
134. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
135. Wei, T. Y., & Simko, V. R. package "corrplot": Visualization of a Correlation Matrix (Version 0.84). https://github.com/taiyun/corrplot (2017).
136. Hao, Z. D. *et al.* RIdeogram: Drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* **6**, e251 (2020).
137. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: An R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
138. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
139. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
140. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

Acknowledgements

Author contributions

T.F. carried out the data analysis and drafted the manuscript. Y.H. participated in the design of the study and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86919-3.

Correspondence and requests for materials should be addressed to Y.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.