Chromatin reprogramming and transcriptional regulation orchestrate embryogenesis in hexaploid wheat

Long Zhao[1,2*], Xuelei Lin[1#], Yiman Yang[1,4#], Xiaomin Bie[3], Hao Zhang[1,2], Jinchao Chen[1,2], Xuemei Liu[1,2], Hao Wang[1,2], Jiafu Jiang[4], Xiangdong Fu[1,2], Xiansheng Zhang[3], Jun Xiao[1,2,5*]

[1]Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Shandong Agriculture University, Tai'an, Shandong, China

[4]Nanjing Agriculture University, Nanjing, Jiangsu, China

[5]CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, 100101, China

[#] These authors contributed equally to this work

*Correspondence to zhaolong@genetics.ac.cn; jxiao@genetics.ac.cn

ABSTRACT

Embryogenesis represents the beginning of life cycle, but our understanding of the regulatory circuitry in plants is far lagged to animals. Here, we draw a transcriptome trajectory and chromatin landscape profile during embryogenesis of most cultivated crop hexaploid wheat, highlighting large-scale chromatin reconfiguration and distinct proximal and distal transcriptional regulation in defining cell fate transition. Upon fertilization, H3K27ac and H3K4me3 resetting were correlated with maternal genome silence, while *de novo* building of chromatin accessibility activated zygotic genome. Global depletion of H3K27me3 in pre-embryo results in a permissive chromatin environment with gain-of-chromatin accessibility, allowing subsequent hierarchical cis- and trans-regulation network mediated by key factors, such as LEC1, MYB, ZHD, LEC2, governing embryo pattern formation. By contrast, H3K27me3 restoration coordinating with chromatin compaction in developmental genes attenuated totipotency and prohibited extensive organogenesis during embryo maturation. In addition, dynamic biased expression of homeolog triads and diverse expression profiles after polyploidization were observed. This is correlated with asymmetric transposon elements insertion in accessible proximal and distal regions. Thus, our study revealed a plant-specific chromatin reprogramming process in facilitating the

34   hierarchical transcription regulation circuits mediated "inverse hourglass model" and

35   unveiled epigenetic regulation of evolutionary divergence among different

36   sub-genome in shaping embryogenesis in polyploidy wheat.

Introduction

38   Embryogenesis, during which fusion of parental terminally differentiated

39   gametes following fertilization generates an entirely new organism, represents the

40   beginning of development and ensures the continuation of the life cycle for both

41   plants and animals[1-4]. Plant embryogenesis can be generally divided into three steps:

42   early-embryogenesis, during which a totipotent zygote was formed following

43   fertilization; mid-embryogenesis, during which the major cell lineages and embryo

44   body pattern was built; and late-embryogenesis, during which nutrition was

45   accumulated in mature embryo followed by seed dormancy[1,2]. Comparisons between

46   animals and plants' embryonic development highlight both conservation and

47   difference. Animals and plants share a similar cell fate determination process-cell

48   division followed by cell identity specification during embryonic development. In

49   addition, both the transcriptional patterns of animals and plants go through the same

50   transcriptomic hourglass model where a conserved phylotypic period during

51   mid-embryogenesis punctuating divergent early- and late-stage between species with

52   a phylum and the inverse hourglass model where gene expression patterns during

53   early- and late-stage are more comparable than mid-stage[5,6]. Nevertheless, there are

54   also differences between the two kingdoms: maternal dominant versus both parental

55   contribution to early-embryogenesis, completed versus partial organogenesis, with

56   dormancy or not, etc[1,7,8]. Besides the morphological and transcriptome similarity and

57   difference, a large gap is the chromatin dynamics comparison between two kingdoms,

58   due to the lack of comprehensive epigenomic study in plants.

59   Epigenetics enables one genome to give rise to a spectrally and phenotypically

60   diverse collection of cells and tissues in response to developmental and environmental

61   cues[9,10]. The technological advances for epigenetics landscape profiling, such as

62   histone modifications, DNA accessibility and DNA methylation, etc., greatly facilitate

63   the discovery of cis-regulatory elements and other non-coding genomic features

64   associated with development, environment, disease and cancer[11-13]. Epigenetic

65   regulatory logic ensures the precise progress and timing of maternal-to-zygotic

66   transition (MZT), zygotic gene activation (ZGA), cell-fate decision, and

67   lineage-specific differentiation during animal embryogenesis[14,15]. A large-scale

68   epigenetic remodeling occurred during embryogenesis, including histone

69   modifications, chromatin accessibility and 3D chromatin organization, as well as

70   DNA methylation[4,16]. H3K4me3 is rapidly depleted after fertilization but

71   re-established during major ZGA and its active removal is important for ZGA

72   initiation[17]. Besides, a noncanonical form of H3K4me3 was present in both humans

73   and mice at distal intergenic regions, though the exact function is not clear[18]. Global

74   erasure of H3K27me3 occurred during early embryogenesis, but with parental

75   genomic allelic difference manner[4]. As well, H3K27me3 could be inter-generationally

76   inherited from the maternal genome to regulator enhancer activity during MZT[19]. Plus,

77   H3K4me3 and H3K27me3 co-occupy lineage-specific genes at the blastocyst stage

78   suggest a critical role of bivalency in poised state maintenance and gene expression

79   regulation[20]. Major chromatin accessibility reorganization is critical for epigenetic

80   reprogramming to convert terminally differentiated gametes into a totipotent state[21].

81   Open chromatin was progressively established during early embryogenesis and

82   showed a major increase in 8-cell embryos in mouse[22,23], correlated with gene

83   activation. However, previous studies on plant embryogenesis have mainly focused on

84   the transcriptional change and functional identity of specific genes at certain embryo

85   development stages[8,24-28]. A systematic view of how epigenomic landscape contributes

86   to genes regulation in cell fate transition, lineage specification and cellular metabolic

87   accumulation during plant embryogenesis is lacking, especially for crops[29].

88   Wheat, as the most cultivated crop worldwide, provides 20% of the calories and

89   protein daily consumed and is important for human sustention. The majority of global

90   wheat production comes from two species, the hexaploid bread wheat (*Triticum*

91   *aestivum*), which accounts for 95% of global wheat production and tetraploid pasta

92   wheat (*Triticum turgidum* var *durum*). Hexaploid wheat (AABBDD) is the result of

93   two times hybridization and polyploidization events of three putative diploid wild

94   grass progenitor species, *Triticum monococcum* (AA), an *Aegilops speltoides*-related

95   species (BB), and *Aegilops tauschii* (DD)[30,31]. A recent finding investigated the

96   transcriptional landscape of polyploidy wheat and diploid ancestors during

97   embryogenesis[27]. Through comparison among different ploidy wheat, the

98   evolutionary divergence of gene expression and contribution of A, B, D subgenome to

99   grain development in polyploidy wheat were characterized[27]. Yet, without further

100   information such as epigenomic profiles, it would be hard to dissect the complex

101   transcription regulation network governing such programmed cell fate transition

102   process and elucidation of driving force for the ontogenetic divergence along with the

103   evolution of polyploidy wheat species. Besides, the elucidation of transcriptional

104   complexity and epigenome atlas during embryonic development could be helpful to

105   study in other developmental processes and plant breeding applications, such as

106   parthenogenesis and somatic embryogenesis[32,33].

107   Here, we investigate various histone modifications and chromatin accessibility of

108   eight-stage samples across embryonic development of hexaploid wheat, providing a

109   systematic view of transcriptome trajectory and chromatin landscape profiles.

110   Thousands of distal genomic regulatory elements and putative enhancer-gene linkages

111   are identified to drive embryogenesis. A plant-specific large-scale chromatin state

112   reconfiguration and transcriptional regulation circuitry were characterized during

113   embryonic phase transition. Further, we provide evidence for an explanation of the

114   regulatory mechanisms underlining evolutionary divergence among different

115   sub-genomes and the 'inverse hourglass model' of transcription variability during

116   hexaploid wheat embryogenesis. This data thus provides a comprehensive resource of

117   genetic and epigenomic annotations for researchers studying transcriptional regulation,

118   embryonic development and polyploid evolution.

119   **Result**

120   **Charting the chromatin landscapes during wheat embryogenesis**

121   Despite the importance of transcription (re)programming and epigenetic

122   regulation during plant embryogenesis, a comprehensive survey is limited for both

123   model plants and crops[29]. This is partly due to the difficulty to get robust epigenomic

124   data from a limited sample of developing embryos. Here, the Cleavage Under Targets

125   and Tagmentation (CUT&Tag) and Assay for Transposase Accessible Chromatin with

126   high-throughput sequencing (ATAC-seq) were used to tackle this challenge[34,35]. In a

127   pilot run, CUT&Tag showed not only high correlation with published Chromatin

128   immunoprecipitation (ChIP)-seq data (correlation >0.8) but also high repeatability and

129   sensitivity with lower nuclei input (~1,000 nuclei/reaction) and sequencing depth

130   (~1/3-1/20 of ChIP-seq) (Supplementary Fig. 1a-c). ATAC-seq from about 5,000 nuclei

131   showed logical peak numbers, good signal-to-noise ratio, reasonable fragment size and

132   genomic distribution, and fully meet the demand of transcription factors (TFs) footprint

133   calling (Supplementary Fig. 1d-h).

134   With this capability, the 'reference epigenome' of eight embryonic

135   developmental stages were generated (DPA 0, 2, 4, 6, 8, 12, 16, 22, days post

136   anthesis), including histone modifications, histone variant and RNA Polymerase II

137   profile and chromatin accessibility, as well as transcriptome (Fig. 1a). In general,

138   various chromatin marks showed featured distribution patterns and logical correlation

139   with transcription (Supplementary Fig. 2a, b). As well, pair-wise correlations analysis

140   showed that active markers (H3K27ac, H3K9ac, H3K4me3, Pol II, H3K36me3,

141   H3K4me1, H2A.Z and ATAC) correlate positively with one another but showed

142   varying degrees of exclusivity with repressive markers (H3K27me3, H3K9me2,

143   H3K9me3) (Supplementary Fig. 2c). Unsupervised PCA and hierarchical clustering of

144   RNA-seq and ATAC-seq revealed a continuous trajectory which can be divided into

145   five major groups roughly corresponding to fertilization, pre-embryo, transition,

146   differentiation and maturation of embryo (Fig. 1b, e). In contrast, core histone

147   modifications showed a phase transition pattern, which revealed a global

148   reconfiguration of chromatin at a certain stage (Supplementary Fig. 2d). GO

149   enrichment of time-lapse correlated modules revealed that cell cycle and DNA

150   replication-related genes were activated during early embryogenesis, organ and tissue

151   specification-related processes were distinguished during mid-development, and

152   nutrient accumulation and dormancy were built during the late stage (Fig. 1c). This is

153   evident in the evolutionary conservation of monocots and dicots embryogenesis[27,28].

154   In particular, some known factors involved in embryo pattern formation, such as

155   *WOX11*, *LEC2, ARR6* and *ABI5*, showed similar expression patterns during

156   embryogenesis among wheat, *Brachypodium* and *Arabidopsis* but not the other genes,

157   such as *YUC1* (Fig. 1d). Furthermore, chromatin accessibility was progressively built

158   until DPA 8, transition stage for embryo patterning, whereas gradually closed during

159   differentiation stage (Fig. 1f). Notably, the majority of ATAC-seq peaks are more than

160   3Kb even 20Kb away from transcriptional start site (TSS) of genes (Fig. 1g), which is

161   different from plants of small genome size[34,36]. Interestingly, more ATACseq peaks

162   distributed even further away from TSS in B subgenome as compared to A or D

163   subgenome (Fig. 1g), coincidently with the fact that B subgenome is larger than A or

164   D[31].

165   Though it is hard to isolate purified embryo at early embryonic development

166   stages (DPA 0, 2, 4), the progressive change of transcription and chromatin

167   accessibility presented here indicates a logical and meaningful cellular process. Thus,

168   our 'reference epigenome' could capture the dynamics of transcription and potential

169   transcriptional regulation features during wheat embryogenesis process.

Distinct feature of proximal and distal chromatin accessible regions and correlation with transcription regulation

Accessible chromatin can expose regulatory DNA sequences to transcription factors, and thus emerges as an accurate proxy for transcription regulatory potential and development[37]. In this study, we identified 1,315,547 accessible chromatin regions (ACRs) from ATAC-seq in total, which could be further categorized into genic (g), promoter (p) and distal (d)ACRs based on the location relative to genes, with a better correlation between dACRs, pACRs and gene expression than gACRs (Fig. 2a). Interestingly, the gain and loss of pACRs and dACRs varied largely during embryogenesis (Fig. 2b and Supplementary Fig. 3a, b). Lots of genes gained pACRs after fertilization at DPA2 but dramatically lost during late embryogenesis at DPA1 and DPA16. Whereas dACRs, accounting for about 75% of total ACRs, showed a sharp 'burst' at DPA8, the transition stage between cell division and differentiation, then quickly declined at DPA12 (Fig. 2b). Genome-wide ACRs distribution showed that centromere region and distal arms of chromosome including telomere are the 'hot spot' for chromatin loose-tight regions, with highly overlapped with Gypsy and CACTA types of TEs. In addition, there was a highly synchronized distribution pattern between ACRs and H3K9me3, but not other histone modifications, at those chromatin eruptions hot spots (Fig. 2b and Supplementary Fig. 3c). This may be correlated with transient expression of TEs just like the activation of retrotransposons in mouse and human 2-cell stage embryo which is critical for the activation of ZGA genes during early embryogenesis[4]. Besides the temporal dynamics, subgenomes of hexaploid wheat showed slight differences for the chromatin accessibility, especially at centromere and telomere regions in addition to the low-colinear chromosome regions (Fig. 2d and Supplementary Fig. 3d).

The different remodeling pattern of pACR and dACR suggests a distinct regulation weight of gene expression during embryogenesis. Indeed, the

197   stage-abundant pACRs did not perfectly fit corresponding genes, as well as functional

198   pathway enrichment (Fig. 2e, f and Supplementary Fig. 4a). We thus assigned dACRs

199   to corresponding genes using a correlation strategy as the previous report (Method;

200   Supplementary Fig. 4b)[38]. In total, we identified 128,212 regulatory dACR (termed

201   reg-dCAR) linked to 43,901 genes, which accounts for 60% of expressed genes

202   during embryo development and shows higher expression variation (Fig. 2g and

203   Supplementary Fig. 4c). High interaction frequency between reg-dACRs and targets

204   measured by Hi-C and conservation among various species emphasized the accuracy

205   of this distal assignment strategy (Supplementary Fig. 4d, e). Only 35% reg-dACRs

206   are assigned to the nearest genes (Supplementary Fig. 4f, g). Most reg-dACRs have

207   only one target, while one gene could be regulated by multiple dACRs

208   (Supplementary Fig. 4h, i). Genes with few reg-dACRs regulations were enriched for

209   essential biological processes, whereas multiple reg-dACR regulated genes are

210   preferred in stress response and organ development (Supplementary Fig. 4j). The

211   assigned target genes of stage-specific reg-dACRs are well associated with specified

212   developmental stages during embryogenesis (Fig. 2g). Of interest, about 60% of the

213   reg-dACRs were located within TEs of different types (Fig. 2h). Interestingly, the

214   accessibility of reg-dACRs and expression of linked genes were sensitive to TEs

215   classes where they are located (Supplementary Fig. 4k, l). Besides, reg-ACRs were

216   highly overlapped with H2A.Z, a histone variant with looser DNA wrapping as

217   compared to canonical H2A[39] (Fig. 2i). Indeed, more than 80% H2A.Z signal located

218   within reg-ACRs, and reg-dACRs with H2A.Z showed higher 'enhancer activity' with

219   more accessibility and higher expression of targeted genes (Fig. 2i).

220   Thus, proximal and distal chromatin accessibility undergoes distinct

221   reprogramming patterns during the process of embryogenesis based on genomic

222   features underlying, with different functional output. Meanwhile, a large proportion of

223   long-range distal regulation and asymmetry of chromatin accessibility dynamics

224   among subgenomes exhibit uniqueness for the huge allohexaploid wheat genome.

Chromatin reprogramming during wheat embryogenesis

To generally classify chromatin states, we integrated eight histone modification marks and Pol II profiles from four embryonic developmental stages (DPA 0, 2, 4, 8) using ChromHMM[40,41]. Totally 15 chromatin states were categorized, covering 0.02% to 3.5% of the genome respectively (exclude No-signal states) (Fig. 3a). They further grouped into five major functional classes: Promoter (Pr), Enhancer-like (EnL), Transcriptional (Tr), Polycomb group (PcG), and Heterochromatin (Hc) (Fig. 3a). Promoter and enhancer-like classes are enriched for active histone modifications such as H3K27/9ac, H3K4me1/3, chromatin accessibility and histone variant H2A.Z (Fig. 3a). As expected, chromatin states and histone modifications underlying were dramatically altered across developmental stages, with the enhancer-like class varied the most than promoters, just like the case in mouse fetal development[42] (Fig. 3b, c). This is correlated with the dramatic transcriptome fluctuation (Fig. 1b, c) and regulatory role of dACRs and pACRs (Fig. 2e, h). In addition, the chromatin state pattern at DPA0 is distinguished from other stages (Fig. 3d), revealing a drastic histone modification reprogramming after fertilization.

In mice and humans, histone modifications such as H3K4me3, H3K27me3, and H3K9me3 are extensively reprogrammed after fertilization and pivotal for early embryogenesis[4,16]. Thus, we take a close look at their behaviors in wheat embryonic development. Expect for H3K36me3 and H3K4me1, most of histone modifications examined show a heavy proportion distributed in the distal intergenic regions (Supplementary Fig. 2a). We investigated the dynamic profiles of each histone modification with separation of proximal and distal regions. Different from animals, genome-wide H3K4me3 showed a slight change during embryogenesis of wheat at both proximal and distal regions (Fig. 3e and Supplementary Fig. 5a). Whereas, H3K27ac presented a quick decline after fertilization, then restore and maintained high till mid-embryo followed by gradually decreasing at late and mature embryo

252   stages (Fig. 3e and Supplementary Fig. 5a). For H3K27me3, instead of being erased

253   quickly after fertilization like the case in animals, a sharp erasure occurred at

254   pre-embryo stage then gradually restored and kept high at late and maturation stages.

255   H3K9me3 fluctuates slightly after fertilization and during pre-embryo, and then

256   increased through mid-embryo till maturation (Fig. 3e and Supplementary Fig. 5a).

257   Besides the individual histone modification, H3K4me3 and H3K27me3 co-occupied

258   regions are referred as bivalent chromatin, which is correlated with and function in

259   lineage-specific genes activation or silencing during embryonic development in

260   animals[4,16]. During wheat embryogenesis, bivalency reached its peak at transition

261   stage (DPA 8) and gradually decreased during tissue differentiation (Fig. 3f). This is

262   mostly due to the turbulence of H3K27me3, since H3K4me3 generally didn't change

263   much (Fig. 3f). In both proximal and distal bivalent regions, we observed widespread

264   ATAC-seq, H3K27ac, H3K4me1, H3K9me3, Pol II and H2A.Z signals, but not

265   transcript products and transcription-associated H3K36me3, suggesting a poised state

266   (Fig. 3g). Of note, genes encoding transcription factors are more enriched than others

267   in bivalent regions at every embryonic stage, indicating that H3K27me3 and

268   bivalency are indispensable for accurate transcriptional regulatory circuitry (Fig. 3h).

269   Besides, subgenomes of hexaploid wheat showed a different abundance of histone

270   modifications even at collinear regions. Interestingly, H3K27me3 and H3K9me3

271   exhibited higher variation among different sub-genomes than active histone marks

272   such as H3K27ac and H3K4me3, as well as histone variant H2A.Z (Fig. 3i and

273   Supplementary Fig. 5b).

274   Taken together, the histone modification reprogramming is distinct in plants as

275   compared to their counterparts in animals (Fig. 3j), indicating a divergent function of

276   individual histone marks in driving the embryogenesis process in two kingdoms. As

277   well, different subgenomes have varied histone modification, especially H3K27me3.

Chromatin regulation of maternal gene silence and zygotic gene activation upon fertilization

Upon fertilization, differentiated parental gametes fused and converted into a totipotent zygote, including maternal gene silencing and zygotic genes initiation with epigenetic reconfiguration in both animals and higher plants[8,15,43]. A large amount of differential expressed genes (DEG) was captured upon fertilization (DPA2/ DPA0), with 8,079 down-regulated genes, mainly involved in maternal tissue identity maintaining such as AP2 and NF-YC9, as well as 8,380 up-regulated genes, including cell cycle and cytokine (CK) signaling genes associated with zygotic initiation (Fig. 4a). We next investigate the chromatin landscapes for seeking of driving force for those DEGs. Consistently, active histone modification H3K27ac and H3K4me3 suffered an overall lopsided decrease at down-regulated genes, while nearly no change at up-regulated genes (Fig. 4b). By contrast, chromatin accessibility didn't change much for down-regulated genes but gained at up-regulated genes (Fig. 4b). As for H3K27me3, regardless of up or down-regulated genes, a reduction upon fertilization was observed. Such profiles indicated that loss of active histone modification might contribute to genes down-regulation, while gain of chromatin accessibility may trigger gene activation. Indeed, attenuation of either or both H3K27ac and H3K4me3 is extensively enriched for down-regulated genes (Fig. 4c, d), with more remarkable down-regulation for loss of both. H3K27ac reprogramming is more determinable compared with H3K4me3 (Fig. 4c, d). Genes subjected to both active histone marks decreasing and transcription levels attenuation are mainly maternal silenced genes, such as orthologs of floral homeotic gene AP2, signaling-related gene MAPKKK17, and stigma expressed gene IAA1 (Fig. 4e, h). Whereas, *de novo* building of pACRs showed a significant overlap and correlation with genes up regulated (Fig. 4f, g). These genes were enriched for pathways in DNA replication, cell cycle, heterochromatin, etc., which are the cornerstones of the early

305   zygotic building (Fig. 4e, h).

306   Besides, there are still 6,737 genes, nearly 80% of up-regulated genes upon

307   fertilization, which can't be explained by either chromatin accessibility or histone

308   modification change (Fig. 4f). We thus speculate a hierarchy transcription regulation

309   network (TRN) might govern the sequential activation of ZGA genes, like the case in

310   animals[44,45]. We constructed such TRN based on top-down and bottom-up methods

311   via integrating the ATACseq and RNAseq data (Method; Fig. 4i and Supplementary

312   Fig. 6b). Three TFs, one TCP and two MYBs were on the top layer of the pyramid

313   (Fig. 4i). The footprint of TCP and MYBs showed not only an augmenting binding

314   strength after fertilization but also transcriptional activators' character, generating

315   robust flanking accessibility without change of the footprint depth (Fig. 4j)[38]. In

316   addition, the basal targets, which were activated at DPA2 or DPA4, were extremely

317   conserved with more than 80% of which have orthologs in *Arabidopsis* ZGA genes

318   set[8], including genes involved in heterochromatin, cell wall, and cytokinesis (Fig. 4i).

319   We further verified the specific expression pattern of TCP during MZT process by in

320   situ hybridization (Fig. 4k), and its activation of downstream targets *GRAS/NF-YB* by

321   reporter assay (Fig. 4l).

322   Taken together, the silence of the maternal genome is attributed to the

323   diminishing of active modification on genes, while the trigger of zygotic may own to

324   *de novo* building of chromatin accessibility and hierarchical regulation of TFs, with

325   several potential key TFs such as TCP being engaged.

Reprogramming of H3K27me3 coordinates with chromatin accessibility to ensure pre-embryo development

328   Following fertilization, several cell fate determination genes were gradually

329   activated in pre-embryo, such as genes in WOX and ARR families (Fig. 1c, d). In

330   parallel, we observed a dramatic drop of H3K27me3 at DPA4, but progressively

331   re-build until DPA8 (Fig. 3e, f, j, Fig. 5a and Supplementary Fig. 7a). For

332   developmental essential genes such as *LEC1*, *BBM*, *WOX11* and *ARR12*, local

333   chromatin environments were permissive in pre-embryo (DPA4), with decreasing of

334   repressive mark H3K27me3 and increasing of active marks H3K27ac and ATAC

335   signal, while a limited change of H3K4me3 (Fig. 5b). What is more, about 90% of the

336   H3K27me3 erasure regions at pre-embryo were re-built in transition stage, along with

337   a large amount of *de novo* gain of H3K27me3 regions (Fig. 5c). The erasure of

338   H3K27me3 was associated with CK pathway genes, embryo essential genes, and

339   tissue specification genes, while the few augment was involved in auxin-related gene

340   sets (Fig. 5c and Supplementary Fig. 7b, c). Correlated with the massive gain of

341   H3K27me3 at DPA8, TaCLFs, a methyltransferase for H3K27, were temporally

342   activated around DPA8 (Supplementary Fig. 7d), and several functional tested 'PRE'

343   type motifs[46,47] were enriched at H3K27me3 restoration loci with their cognate

344   binding TFs highly expressed at DPA8 (Supplementary Fig. 7e, f). Thus, cis-motif and

345   trans-factors mediated H3K27me3 deposition may be involved in embryonic

346   development in wheat.

347   H3K27me3 intensity generally showed a negative correlation with transcription

348   (Fig. 5d and Supplementary Fig. 2b), however, decreasing of H3K27me3 alone could

349   not guarantee gene activation from DPA2 to DPA4 (Fig. 5e). However, H3K27me3

350   reduction was highly overlapped with the addition of H3K27ac or chromatin

351   accessibility (Fig. 5e). As expected, genes with a gain of active chromatin

352   modification in addition to H3K27me3 reduction showed significant activation, with a

353   more dramatic effect by a gain of chromatin accessibility (Fig. 5e). Furthermore, there

354   were 561 genes whose transcriptional activation showed highly synchronous patterns

355   with the gain-of-chromatin after DPA4 (Fig. 5f and Supplementary Fig. 7g). This

356   synchronization ensured the proper timing of genes activation, with embryo essential

357   gene *LEC1* firstly activated followed by pluripotency gene *ARR12*, and later

358   differentiation genes *XMXL5* and *IPS2* (Fig. 5f and Supplementary Fig. 7g). Together,

359   this suggested that resetting of H3K27me3 lifted the chromatin layer barrier on

360   embryo patterning genes, which was progressively activated along with the gain of

361   chromatin accessibility. Alternatively, a global regain of H3K27me3 was observed at

362   DPA8, a transition stage with the maximal open chromatin and bivalent regions as

363   well (Fig. 1f, 3f, 5a, c, g and Supplementary Fig. 7a). Indeed, among the 133,716

364   gained H3K27me3 peaks, 40% show bivalency status with enriched H3K4me3 and

365   opened chromatin status (Fig. 5g, h). The expression levels of genes with H3K27me3

366   marked promoters were significantly lower compared with bivalency promoters at

367   transition stage, while slight change was detected between genes with distal

368   H3K27me3 and bivalency (Fig. 5g, i). Genes got bivalency or H3K27me3 at both

369   promoter and distal regions were silenced at transition stage, with stronger

370   suppression function of modifications at promoter compared with those at distal (Fig.

371   5g, i, j). Yet, genes with bivalent promoters still have the opportunity to re-activation,

372   such as *BBM*, but not genes with H3K27me3, such as *ARR12* (Fig. 5i, j).

373   Thus, the global depletion of H3K27me3 in pre-embryo resulted in a permissive

374   chromatin environment, where the timely opening of chromatin can trigger the

375   activation of developmental genes. Bivalency modification generated by H3K27me3

376   restoring marked cell fate determination genes, whose dysregulation was related to

377   transition of totipotent pre-embryo to differentiation.

Gene regulatory network orchestrates embryo pattern formation

379   After transition stage, embryo pattern formation was initiated, along with

380   precisely programmed transcription regulation pyramid[48]. The complex relationship in

381   TF-genes and TF *per se* could be well exemplified by the continuous transcription

382   trajectories (Fig. 6a, b). To test how the architecture of embryo patterning from

383   totipotent stem cell to maturation, we calculated the pseudotime of stages covering

384   early- to mature-embryo based on the PCA distance (Fig. 6b). As expected, an

385   excellent association between gene expression and both pACR and dACR was

386   observed (Fig. 6c), which indicated that the expression pattern of these genes was
387   governed by individual ACRs and TFs binding. Comparison of genes expressed in the
388   early time point (Cluster 1 and 2) and the later ones (Cluster 3 and 4) presented a
389   functional diversity, of which the early elevated genes were involved in cell division,
390   while the later ones in cell specification (Fig. 6d). What is more, motifs of TFs
391   binding sites, whose orthologs can facilitate inducement of cell totipotency in
392   *Arabidopsis*, were enriched in early build ACRs, such as the binding motifs of LEC1,
393   MYB118, and WUS, etc. Whereas the TFs, which function in seed dormancy, mainly
394   occupied the later elevated ACRs, such as ABI5 (Fig. 6e). This result demonstrated
395   evolutionary conservation of several key factors in embryo patterning in monocot and
396   eudicot.

397   We further dissect the architecture of the gene regulation network (GRN). In
398   addition to binding motif specificity, we filtered target genes to TFs by co-expression
399   pattern as well (Method; Fig. 6f). As a result, 1,158 paired TF-cis motif interactions
400   were identified including 154 target genes, 695 ACRs and 191 TFs (Fig. 6f). To better
401   understand how TFs are coordinated during embryo pattern formation, we next extract
402   the TF-TF regulation network (Fig. 6g). We found four pivotal classes of TFs located
403   at the center of the network, TCPs, ARFs, MYBs, and WOXs. Some of those TFs
404   such as WOXs and ARFs are reported to regulate embryo patterning in other
405   species[49-51]. Several known genes regulating organ and tissue specification in
406   Arabidopsis were under the governing of these TF modules, such as I*AA14, LAX2,*
407   *PLT1,* and *SMXL5*, etc. To further estimate this GRN, we use in situ hybridization to
408   verify the spatiotemporal expression pattern of key TFs module including
409   LEC1-Myb118-ZHD5-LEC2-BBM. As shown, the *LEC1, Myb118* and *ZHD15* genes
410   were highly expressed in pre-embryo stages of overlapped tissue, while *LEC2* was
411   activated later, with a peak at transition stage, and *BBM* gradually induced at
412   mid-embryo, indicating a sequential pattern (Fig. 6h). The luciferase activity from the
413   reporter assay validated the regulatory circuit along LEC1, MYB118, ZHD5, LEC2

414 and BBM module (Fig. 6i). Interestingly, TaLEC2 could bind the pACR of *BBM*
415 directly and activate it *in vitro* (Fig. 6i), which is not reported in the model plant
416 *Arabidopsis*.

417 Thus, these data described a cohesive, sequential wave of TFs regulation in
418 governing wheat embryo pattern formation.

Chromatin regulation facilitates embryo maturation and prohibits extensive organogenesis

421 During seed maturation, nutrient was accumulated in both endosperm and
422 embryo, meanwhile, the dormancy program was progressively established with
423 increased ABA levels accompanied by desiccation[52]. Indeed, genes elevated
424 specifically in late embryo development were enriched in carbohydrate, lipid, protein,
425 amino acid, secondary metabolic pathway, and ABA signaling and homeostasis (Fig.
426 7a). As expected, H3K27ac increased progressively at those genes during late embryo
427 development, but H3K27me3 declined to a relatively low level, while chromatin
428 accessibility was slightly changed (Fig. 7b). Next, we applied ATAC footprint to seek
429 candidate TFs associated with the regulatory regions of the above genes. ABA-related
430 TFs were enriched for regulating ABA pathway genes, which verified the logic of this
431 analysis (Fig. 7c). In addition, most nutrient storage pathways shared at least one
432 upstream TF with ABA pathway, such as ABI5 and ABF1 (Fig. 7d), which are known
433 to regulate a subset of late embryogenesis abundant genes in *Arabidopsis*[53]. This
434 result suggested that several key regulators could synchronize nutrient accumulation
435 and embryo dormancy in cooperation with chromatin landscape dynamics.

436 Unlike animals, extensive organogenesis is prohibited during late embryo
437 development, which enables plants to generate post-embryonic lateral organs in
438 adaptation to changeable environments. Meanwhile, immature embryo served as
439 excellent explants for wheat regeneration, a process including regaining of totipotency

440   and organogenesis[53,54], however such capability was reduced upon embryo maturation

441   (Supplementary Fig. 8a). We wonder how the totipotency capability is attenuated and

442   why the terminal organ is not formed during embryo maturation. First, we identified

443   organ identity genes by their specific expression in terminal organs such as root and

444   leaf, for comparison totipotent genes were characterized with high expression in

445   mid-embryo or embryonic callus (Fig. 7e). GO enrichment of different groups of

446   genes verified the accuracy of genes identity (Supplementary Fig. 8b). Next, we

447   evaluated the chromatin landscape of each cluster genes in different organs. Depletion

448   of H3K27me3 on organ identity genes was observed at the respective organ, whereas,

449   an obvious ATAC signal presented for both totipotency and organ identity genes in its

450   corresponding organ (Fig. 7f, Supplementary Fig. 8c, d). Surprisingly, low/no

451   H3K27me3 was enriched on totipotent genes even in differentiated organs (Fig. 7f).

452   In embryonic callus, of which new plant will regenerate, the promoter region of both

453   totipotent genes and organ identity genes were opened with high chromatin

454   accessibility and low H3K27me3 level (Fig. 7f). Thus, chromatin accessibility and

455   H3K27me3 coordinately regulated organ identity gene's activation potential while

456   chromatin accessibility but not H3K27me3 mediated repression is indispensable for

457   totipotency gene silencing. Indeed, several regeneration-related factors, such as REV,

458   SAUR41, CKX7 and DDM1 were gradually diminished during embryo maturation

459   along with closing chromatin status but activated during callus induction process (Fig.

460   7g). Last, we investigated the gene-regulatory logic underlying plant organ formation.

461   WRKYs and AtHB TFs footprint were enriched in regulatory regions of root and leaf

462   identity genes respectively, whereas, bHLH and GATA TFs footprint were enriched in

463   both gene-sets (Supplementary Fig. 8e). Remarkably, nearly half of the identified TFs

464   did not show organ-specific expression, but instead a wide-spectrum pattern,

465   indicating that TFs expression *per se* cannot determine the specificity of organ

466   identity (Supplementary Fig. 8f). For example, although WRKY75 was expressed

467   during embryo maturation and it owns the capability to bind and active *AMT1;1,* an

468   ammonium transporter coding gene in root (Fig. 7h, Supplementary Fig. 8f), AMT1;1

469   was silenced due to the H3K27me3 repression and compaction chromatin

470   environment at the promoter (Fig. 7i).

471   Taken together, chromatin layer regulation ensures the proper embryo maturation

472   process, with an accumulation of nutrients and reduced totipotency ability but not

473   extensive organogenesis.

Epigenetic regulated ontogenetic divergence and inverse hourglass pattern during embryogenesis

476   As wheat is an allohexaploid plant, we further extend the understanding of

477   epigenetic regulated embryogenesis to the evolution and polyploidization perspectives.

478   Genes were clustered based on their evolutionary age (Supplementary Fig. 10a). Most

479   of the genus *Triticum*-specific genes, especially D-subgenome unique genes, were

480   located within PcG chromatin state and did not participate in embryonic development

481   (Supplementary Fig. 10a-c). For the homeolog triads that are present in all three

482   subgenomes, chromatin accessibility is related to the bias expression pattern

483   (Supplementary Fig. 11) in addition to histone modifications as reported[55]. In detail,

484   suppressed triads are generally more than dominant triads, in addition, D-subgenome

485   suppressed triads are relatively less than A or B-subgenome suppressed. More

486   balanced triads were expressed in the middle period compared with early and late

487   embryogenesis (Supplementary Fig. 10d). Such observation was generally consistent

488   with the previous report but different in details[27].

489   Polyploidy may drive genes neo- or sub-function to confer phenotypic plasticity,

490   which can be represented by the varied expression pattern. Instead of identifying

491   genes of differential expression at a specific embryonic stage by comparing hexaploid

492   wheat with tetraploid and diploid[27], we look for genes showing varied expression

493   profiles during the process of embryo development. Indexing by Pearson correlation

494 degree, genes are categorized into dysfunction, middle, and conserved clusters (Fig 8a

495 and Supplementary 9a). As expected, conserved genes were mostly homeologs triads,

496 with sub-genome balanced expression, and showed higher sequence conservation and

497 stronger selection pressure (Fig. 8b, c and Supplementary Fig. 9b-f). However, a

498 considerable proportion of dysfunctional genes changed expression profile in

499 hexaploid wheat as compared to ancestors, but with slight sequence variation and

500 subject to strong negative selection (Fig. 8c). Promoters of these genes, especially

501 around 3Kb upstream of TSS, were preferably inserted by TEs (Fig. 8d and

502 Supplementary Fig. 9g). Intriguingly, TE insertion regions tended to be accessible

503 chromatin and enriched for H3K27ac and H2A.Z (Fig. 8e). For example,

504 TaUGT91C1 showed a B-subgenome dominance expression in hexaploid wheat,

505 while silenced in all ancestors (Fig. 8f). Accordingly, the promoter and dACR region

506 harbored TEs with higher chromatin accessibility as compared to A or D counterparts.

507 Therefore, the TEs insertion within proximal and distal active chromatin regions may

508 drive the novel expression pattern of offspring genes following polyploidization.

509   A previous study on embryogenesis revealed a conserved hourglass model in

510 hexaploid wheat as well as its tetraploid, and diploid ancestry, but a slight difference

511 for the phylotypic stage[27]. Here, our transcriptome data in hexaploid wheat (Chinese

512 Spring) also supported that transition stage of DPA8 represents the phylotypic stage

513 (Supplementary Fig. 9h). In addition, we find an inverse hourglass model for

514 transcriptional variability. A sharp mid-developmental transition was presented, which

515 separated two conserved phases, early and late developmental stages in both the

516 comparison between homoeologs in hexaploid *per se* and between counterparts of

517 hexaploid gene and its ancestors (Fig. 8a, g and Supplementary Fig. 9a). We

518 hypothesized that epigenetic regulation may account for this middle-stage transition.

519 Indeed, H3K27ac was mainly contributed to the middle stage rather than early- and

520 late-stage, whereas H3K27me3, H3K4me3, and ATAC behaved oppositely (Fig. 8h

521 and F Supplementary Fig. 9i). In addition, chromVAR analysis indicated that the key

522   TFs during embryogenesis functioned mainly in the middle stage (Fig. 8h and F

523   Supplementary Fig. 9j). Thus, the cis- and trans- regulation driven by TFs was more

524   meaningful in mid-stage compared with early- and late-stage. In sum, transcriptional

525   variance showed an inverse hourglass model, which may be governed by chromatin

526   layer regulation at early and late stage, while cis-trans regulation at the middle stage

527   (Fig. 8i).

528   **DISCUSSION**

529   Despite containing the same genome, different cell types possess unique

530   identities with specific gene expression patterns. This is partly because some unique

531   portions of the genome are accessible in each cell type, which is controlled at the level

532   of chromatin. During embryogenesis, parental differentiated gametes fuse and convert

533   into a totipotent zygote, and then to a differentiated embryo, cells undergo extensive

534   identity switch, and thus provide an ideal system for studying chromatin-based

535   transcription regulation. Indeed, embryogenesis is extensively studied in mammalians.

536   However, our understanding of embryogenesis is far lagged in plants, especially in

537   crops. In this study, we generated comprehensive RNAseq, CUT&Tag and ATACseq

538   data in three early embryo sac and five later embryo samples of hexaploid wheat

539   Chinese Spring. It provides a genome-wide annotation of dynamic 'reference

540   epigenome', which enables the comparative analysis of large-scale chromatin

541   reprogramming and its function for embryogenesis between plants and animals, and

542   the understanding of multi-layer orchestrated transcriptional regulation circuit at

543   different stages of embryonic development in wheat. Integration with the previous

544   report on the transcriptome of embryogenesis from different ploidy wheat[27], our study

545   further revealed the potential driving force for evolutionary divergence among

546   different sub-genome in shaping embryogenesis in polyploidy wheat.

547   In this study, we observed large-scale chromatin modification reprogramming

548   during wheat embryogenesis, especially at MZT, pre-embryo and transition stages

549 (Fig. 3e, j, Fig. 4b, Fig. 5a and Supplementary Fig. 5a). Though the histone

550 modification resetting pattern shows general similarity to animals[4,16], there was a large

551 variation in terms of modification types and timing of reprogramming. In wheat,

552 widespread H3K27ac dropped off during MZT and contributed to the silence of

553 maternal genes, slightly for H3K4me3 (Fig. 3e, j and Fig. 4b), which however is

554 dramatically reset upon fertilization and re-build during major ZGA and is importan

555 for ZGA in mice[4,15]. H3K27me3 sharply dropped at pre-embryo but not right after

556 fertilization like its counterpart in mammals[4] (Fig. 3e, j, Fig. 5a and Supplementary

557 Fig. 5a). Both H3K27ac and H3K27me3 were momentarily and universally

558 re-established, of note, regions of H3K27me3 decreased in DPA4 is a great part of that

559 increased in DPA8 (Fig. 3j, Fig. 5a and Supplementary Fig. 5a). This suggests the

560 epigenome 'rebooting' model is conserved for both animals and plants[15]. Importantly,

561 H3K27me3 dynamics marked embryonic essential genes (Fig. 3h and Fig. 5b),

562 indicating its pivotal role in embryo development. Indeed, we found aberrant seed

563 development when knocking out H3K27me3 writer complex component

564 (Supplementary Fig. 12), though a detailed mechanism still needs to address. Recent

565 findings characterize the direct involvement of PRC2 component, writer of

566 H3K27me3, for embryo development in *Arabidopsis*[56]. Besides histone modification,

567 chromatin accessibility also shows dynamics during embryogenesis. Interestingly,

568 dACRs (account for 75% total ACRs) and pACRs (16%) show unique fluctuation

569 pattern, correlated with distinct transcriptional regulation respectively (Fig. 2b). For

570 genic regions, pACRs were significantly gained upon fertilization, strongly correlated

571 with zygotic gene activation (Fig. 4b, e-h), and continued gain till pre-embryo (DPA4)

572 then gradually declined later on. Whereas, dACRs gradually gained after fertilization,

573 with 'transient' maximize at transition stage (DPA8) then quickly declined later,

574 where the opened intergenic regions are highly overlapped with multiple TEs, such as

575 Gypsy and CACTA, as well as 'burst' of H3K9me3 region at centromere and telomere.

576 Further analysis of global DNA methylome and enhancer RNA profiling around

577 transition stage would give a clearer conclusion of the functional consequence of such

578 dACR eruptions. Nevertheless, our data provide a novel comparison between animals

579 and plants for the whole embryogenesis process driven by chromatin landscape

580 dynamics.

581   Embryogenesis is governed by precise transcriptional regulation program[2,3,57].

582 However, how such transcription regulation is orchestrated by different types of

583 regulation, such as histone modification, chromatin accessibility, DNA methylation,

584 transcription factor-cis motif modules, is not clear. Based on our analysis, a dynamic

585 'three-layer regulation model' is proposed for different embryonic development stages

586 (Fig. 8i). In this model, de-repression of local histone modification is a prerequisite of

587 permissive chromatin environment and open chromatin is the cornerstone of

588 transcription factors free binding. At pre-embryo, H3K27me3 was globally reset,

589 which provides a flexible chromatin environment, but yet, not sufficient to active genes

590 and gain-of accessibility at the promoters are the indispensable triggers. After this,

591 freestanding TFs coordinate was gradually built in a subtle and specific manner (Fig. 5).

592 The attenuation of totipotency and prohibition of extensive organogenesis at

593 late-embryo development stage can also support this model in an opposite way (Fig.

594 7e-i). A representative case is the WKRY75- *AMT1;1*. Despite the transcriptional

595 activation of WKRY75 for *AMT1;1* in reporter assay, *AMT1;1* was still repressed in

596 late-embryo because of the compacted promoter region along with enriched

597 H3K27me3. Besides the stage-specific cases, similar multi-layer regulation could

598 apply to the entire embryogenesis process in general. Transcription data for

599 embryonic development demonstrate a cross-kingdom and ploidy conserved 'inverse

600 hourglass model', in which a transcriptional divergent mid-developmental transition

601 punctuated the conserved early- and late-stage (Fig. 8a, g and Supplementary Fig. 9a).

602 Accordingly, a panorama of genetic and epigenetic dynamic presents that global

603 histone modification reprogramming at early-stage, while the chromatin accessibility

604 burst and TF-binding at mid-stage as well as the relative compacted genome at

605   late-stage (Fig. 4-7, Fig. 8h-i and Supplementary Fig. 9i, j). Thus, the early- and

606   late-stage were dominated by chromatin level restriction, while mid-stage was

607   governed by TF-motif regulatory circuits along with open chromatin, of which the

608   chromatin level regulation effect is minimized. This regulation pattern fits the

609   transcriptional divergent 'inverse hourglass' in general, in combination with the

610   genetic diversity 'hourglass model' for mediating conserved and diversified

611   embryogenesis in different species.

612   Bread wheat (AABBDD) undergoes polyploidization and domestication, along

613   with evolutionary divergence among different sub-genomes. Indeed, we observed

614   dynamic biased expression of homeolog triads during embryogenesis, correlated with

615   histone modification variations as well as chromatin accessibility differences among

616   A,B, D sub-genome (Fig. 2d, Fig. 3i, Supplementary Fig 3d and Supplementary Fig

617   5b). In addition, a considerable proportion of genes showed varied expression profiles

618   during embryogenesis in allohexaploid bread wheat compared with the diploid

619   ancestors[27] (Fig. 8a and Supplementary Fig 9a). Part of them possesses DNA

620   variations within the genic regions, which may alter gene function and indirectly

621   change their expression profile as an adaptation for the 'new' roles. Whereas other

622   genes with slight or nearly no sequence variation at genic region still show altered

623   expression patterns. Of note, promoters and distal regulatory regions of those genes

624   contain asymmetric TE insertions among different sub-genomes, with a preference for

625   sites of open chromatin region and enriched for active histone marks such as

626   H3K27ac and H2A.Z. Thus, epigenetic landscape variation among different

627   sub-genome, especially TE inserted active chromatin regions mediated transcriptional

628   alteration, may drive the evolutionary divergence in shaping embryogenesis in

629   polyploidy wheat.

630   Besides, our data also elucidated detailed transcriptional regulation circuits

631   during ZGA and embryo pattern formation, as well as the maturation of embryo (Fig.

632   4, Fig. 6, Fig. 7). The hierarchy transcription regulation pyramids and underlying

633 epigenetic regulation of key transcription factors during ZGA would be worth to test

634 for the automatic embryo development program without fertilization. As well, the

635 knowledge about programmed embryo pattern formation and gradually attenuation of

636 totipotency would be valuable for studying wheat regeneration, during which

637 immature embryo serves as the excellent explants. As a comprehensive study of

638 embryonic chromatin landscape dynamics and transcriptome, our research is

639 anticipated to facilitate gene functional study of wheat embryogenesis and plant

640 breeding application, such as parthenogenesis and somatic embryogenesis.

641 **Methods**

642 **Plant materials and growth conditions**

643 A spring wheat (Triticum aestivum) Chines spring was used in this study. All plant was

644 grown in a growth room under a 16h light/8h dark cycle with light intensity 1,000 μmol

645 m−2 s−1, and temperature 15-26°C depending on growth stage. The stamens are

646 removed before the pollen matures. Then we conduct artificial pollination and record

647 the number of days to ensure the accurate time of seed development. Embryo and

648 embryo sac at the specific developmental stage were sampled using RNA and nuclear

649 extraction.

650 **RNA extraction**

651 Total RNA was extracted using HiPure Plant RNA Mini Kit according to the

652 manufacturer's instructions (Magen, R4111-02), and libraries were sequenced using

653 an Illumina Novaseq platform.

654 **CUT&Tag experiment**

655 The CUT&Tag experiment was performed as reported with minor modifications[35]. For

656 each replicate, the fresh samples were soaked in the HBM buffer (25 mM Tris-Cl pH

657   7.6, 0.44 M sucrose, 10 mM MgCl2, 0.1% Triton-X, 10 mM Beta-mercaptoethanol, 2

658   mM spermine, 1 mM PMSF, EDTA-free protease inhibitor cocktail), after chopped

659   with a razor blade, the mixture was filtered through a 40-μm cell strainer. The crude

660   nuclei was washed twice by HBB buffer (25 mM Tris-Cl pH 7.6, 0.44 M sucrose, 10

661   mM MgCl2, 0.1% Triton-X, 10 mM Beta-mercaptoethanol). Then the crude nuclei

662   were stained with 4,6-diamidino-2-phenylindole and counted by hemocytometer.

663   Extracted nuclei (10,000) was resuspend softly in 50μl antibody buffer (20 mM

664   HEPES pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1× Protease inhibitor cocktail; 2

665   mM EDTA; 0.1% BSA) containing a 1:50 dilution of corresponding antibody. After

666   overnight incubation in 4℃, the primary antibody was removed by centrifuged

667   (1,200g,3min), the nuclei were incubated in 50μl wash buffer (20 mM HEPES pH 7.5;

668   150 mM NaCl; 0.5 mM Spermidine; 1× Protease inhibitor cocktail) with secondary

669   antibody (1:100; Guinea Pig anti-Rabbit IgG antibody) at 4℃ for around 1-2 hour and

670   then wash twice with wash buffer. A 1:100 dilution of pA-Tn5 complex was prepared in

671   CT-300 buffer (20 mM HEPES pH 7.5; 300 mM NaCl; 0.5 mM Spermidine; 1×

672   Protease inhibitor cocktail), after nuclei centrifuged, add 100μl mix and incubate 2-3h

673   in 4℃. After incubation with pA-Tn5, wash twice with CT-300 buffer. Then incubate

674   nuclei in 300 μl Tagmentation buffer (20 mM HEPES pH 7.5; 300 mM NaCl; 0.5 mM

675   Spermidine; 1× Protease inhibitor cocktail; 10 mM MgCl2) in 37℃ for 1h. To stop

676   tagmentation reaction, add 10 μl 0.5M EDTA, 3 μl 10% SDS and 2.5 μl 20mg/ml

677   Protease   K,   incubate   1h   in   50 ℃.   The   DNA   was   extracted   with

678   phenol:chloroform:isoamyl alcohol, precipitated with ethanol and resuspended in

679   ddH2O. The library was amplified 17 cycles by Q5 high fidelity polymerase (NEB,

680   M0491L), and purified by AMPure XP beads (Beckman, A63881). Finally, the library

681   was sequenced using an Illumina Novaseq platform.

ATAC-seq experiment

682
683 The method of nuclei extraction was performed as before. After checking the nuclear
684 integrity, the nuclei extracted (50,000 per reaction) were incubated with the Tn5
685 transposase and tagmentation buffer at 37 °C for 30 min (Vazyme Biotech,TD501-01)[34].
686 After tagmentation, the DNA is purified by PCR purification kit (QIAGEN, 28106).
687 PCR was performed to amplify the library for 9-12 cycles using the following PCR
688 conditions: 72 °C for 5 min; 98 °C for 30 s; and thermocycling at 98 °C for 15 s, 63 °C
689 for 30 s and 72 °C for 40 s; following by 72 °C 5 s. After the PCR reaction, libraries
690 were purified with AMPure beads (Beckman, A63881).

Bioinformatics Data Preprocessing and Alignment

692 All fastq data, including DNA and RNA sequencing, were generated based on Illumina
693 Hiseq-PE150. Raw data were filtered by fastp (v0.20.0) with "--detect_adapter_for_pe"
694 parameter for reads filter, low-quality bases trimming, and adapters removing[58].
695 Furthermore, the clean data was evaluated by fastqc software (v0.11.8)
696 (https://github.com/s-andrews/FastQC) to ensure the high quality of reads.

697 Both DNA sequencing including resequencing, CUT&Tag, and ATAC-seq and
698 RNA sequencing data were aligned based on *Triticum aestivum* (Chinese Spring)
699 genome assembly (IWGSC RefSeq v1.0)[31], which was downloaded
700 from https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/.
701 The IWGSC Annotation v1.1 was used as the gene annotation reference. For DNA
702 sequencing data, BWA-MEM (v0.7.17) algorithm was used for alignment with "-M"
703 parameter to avoid shorter split hits[59]. For RNA sequencing data, hisat2 (2.1.0) was
704 applied for reads mapping with default parameters[60].

RNA-seq Data Processing and Expression Clustering

706 Sam files generated from hisat2 were converted to bam files without deduplication.

707 FeatureCount v1.6.4 was used for reads quantity per gene[61]. An R package edgeR was

708 used for DEGs (differentially expressed genes) examination, with a threshold absolute

709 value of Log2 Fold Change ≥ 1 and FDR ≤ 0.05[62]. The raw matrix was further

710 normalized to TPM (Transcripts Per Kilobase Million) for gene expression

711 quantification. Genes expression data from different tissue was downloaded from a

712 previous publication[55].

713 TPM values of genes were clustered by k-means method in the heatmaps in Fig.

714 5e and 5h. Modules information generated by WGCNA[63] was used for key gene set

715 selection.

716 For functional enrichment, GO annotation files were generated from IWGSC

717 Annotation v1.1 and an R package clusterProfiler was used for enrichment analysis[64].

Cut&Tag and ATAC-seq Data Processing

719 Cut&Tag data analysis was largely based on the previously provided pipeline[65]. In

720 brief, two replicates sam files were converted to bam files and sorted by samtools,

721 respectively[59]. We further filter the reads mapped with "samtools view -bS -F 1,804 -f

722 2 -q 30" to filter the low-quality mapped reads. The high-quality mapped reads were

723 reduplicated using Picard-2.20.5-0 ("Picard Toolkit." 2019). Two replicate bam files

724 were merged by samtools. For IGV browser visualization, merged bam files were

725 converted to RPKM (Reads Per Kilobase per Million mapped reads) normalized

726 bigwig files with 10 bp bin size for browser visualization by bamCoverage provided

727 by deepTools (3.3.0) with parameters "-bs 10 --effectiveGenomeSize 14,600,000,000

728 --normalizeUsing RPKM --smoothLength 50"[66]. For Cut&Tag data comparison, scale

729 factors were calculated by ChIPseqSpikeInFree (v1.2.4)[67]. For peak calling, both

730 SEACR v1.3 and MACS2 v 2.1.2 were used[68,69]. We performed SEACR with numeric

731 threshold 0.05 and normalized stringent model. For narrow histone marker (H3K27ac

732 and H3K4me3) and broad histone markers (H3K27me3), parameters "-p 1e-3" and

733 "--broad --broad-cutoff 0.05" provided by MACS2 were used, respectively. Finally,

734   only peaks generated by MACS2 which overlapped with that generated by SEACR

735   were retained for downstream analysis by "intersect -wa" parameters of bedtools

736   v2.27.1[70].

737   The bam file process and bigwig conversion steps in ATAC-seq are the same as

738   that in Cut&Tag. For ATAC-seq peak calling, only MACS2 was used with parameters

739   "--cutoff-analysis --nomodel --shift -100 --extsize 200". ChIPseqSpikeInFree was not

740   applied in ATAC-seq data process.

741   For both Cut&Tag and ATAC-seq peaks, if a peak overlapped with the proximal

742   of a gene, including 3kb upstream and 2.5 kb downstream, we assigned the peak to

743   the gene. If multiple genes meet the condition, a position priority strategy (promoter >

744   exon > intron > 5'UTR > 3'UTR > downstream) and nearest gene principle was used

745   for target genes assign. An R package ChIPseeker was used for this peaks annotation

746   process[71].

747   **DERs detection**

748   Reads count under special peaks of Cut&Tag and ATAC-seq were calculated by

749   FeatureCount[61]. For Cut&Tag data, scale factors generated by ChIPseqSpikeInFree

750   were used for differential analysis following the suggested usage of DESeq2 method

751   in ChIPseqSpikeInFree manual[67,72].

752   **Chromatin state analysis**

753   For chromatin state analysis, chromHMM was used[40,41]. "BinarizeBam" and

754   "LearnModel" commands were used for chromatin-state annotation. We used 9

755   CUT&Tag data as input and got 15 chromatin-states (Fig. 3a left panel).

756   **ATAC-seq foorprints identity**

757   HINT (Hmm-based IdeNtification of Transcription factor footprints) was used for

758   ATAC-seq footprints identity[73]. JASPAR Plantae database (https://jaspar.genereg.net/)

759 was used as motifs set[74].

Distal ACRs annotation

761 The distal ACRs annotation strategy was largely based on a previous study with a

762 small modification[38]. In brief, genes within 0.5M from a distal ATAC-seq peak are

763 considered candidate target genes (Supplementary Fig. 4b). Then if there is a

764 significant positive correlation between gene expression (TPM) and ATAC-seq signal

765 (FPKM) in eight samples, this gene is considered to be the target of the peak

766 (Supplementary Fig. 4b). A small difference from the Corces et al. (2018) is that we

767 use this strategy only for distal ATAC-seq peaks, and proximal peaks annotation was

768 based on the nearest gene principle[38].

GRNs construction in early embryogenesis

770 GRNs construction in early embryogenesis was based on top-down and bottom-up

771 methods (Fig. S5a). We believe that genes whose expression was up-regulated but not

772 promoter ATAC signal may be largely regulated by TFs that both expression and

773 pACRs up-regulated (Fig. 4b left panel). Thus, the GRN contained four layers, in

774 which the lowe two layers were genes firstly up-regulated in DPA2 and DPA4

775 compared with DPA0, respectively. Firstly TFs motif was enriched for the

776 up-regulation (DPA2/DPA0) promoter ATAC-seq peaks that were assigned to

777 up-regulated (DPA2/DPA0) genes using homer software (v4.11)[75]. Two types of

778 transcription factors were enriched, including one TCP and two MYB. We used the

779 top-down method to find their downstream target genes through the footprint of the

780 three TFs. Then we used the bottom-up method to find the upstream TFs of genes in

781 the lower two layers through motif enrichment by clusterProfiler[64]. Finally, we got the

782 TFs in the second layer through the intersection of the targets of the top layer and the

783 enrichment result of the bottom layers. These TFs were both expression up-regulated

784 and pACR signal up-regulated. In most conditions, one motif corresponds to multiple

785 TFs. To get the specific TFs, we overllaped our GRN result with previous Genie3

786 network[55] and got the final GRN in Fig. 4h. EMB genes were got from a previous

787 review in Arabidopsis[76], and we used diamond (v2.0.13) with parameter "--id 50 -e

788 1e-20 -f 6"[77,78].

Psuedotime indexing and GRNs construction in embryo body formation

791 Only six-module genes generated from WGCNA were used for the analysis in Fig. 6.

792 The psuedotime indexing method is the same as previous studies[79-81]. Briefly,

793 principal component analysis

794 PCA was used for both RNA- and ATAC-seq data for specific genes. The

795 psuedotime was calculated based on the sample distance between neighbor samples

796 and was scaled to a range from 0 to 10. For one gene, the expression model was fit

797 based on expression level and psuedotime using the "loess" function in R, and

798 500-time points were generated between 0 to 10, as well as the corresponding

799 expression levels based on the fitted curve.

800 For one gene, the expression level was normalized by Z-score. We further

801 calculated the PC1 and PC2 for every gene using the expression values of eight

802 samples. Because the standard expression values of all genes can form a circle, the

803 atan2 function in R was used to return the angle in radians for the tangent PC2/PC1,

804 which were further used for gene expression order ranking. As a result, the

805 psuedotime expression of genes was ranked and visualized by complexHeatmap in

806 R[82]. GO and motif (generated from footprint analysis) enrichment was calculated by

807 clusterProfiler[64].

808 For GRN building, the TFs for significantly enriched motifs with p-value < 0.05

809 were used. We further calculated the Pearson correlation between TFs and

810 corresponding targets, and only significant pairs (p-value < 0.05) were retained. To

811 simplify the GRN, we focused on the TFs-TFs network. As a result, several TFs

812 contained multiple genes, we combine those as TF modules, including TCPs, ARFs,

813 MYBs, and woxs.

814 Data availability

815 The raw sequence data were deposited in the Genome Sequence Archive

816 (https://bigd.big.ac.cn/gsa) under accession number xxx.

817 Acknowledgements

822 Author contributions

823 J.X. and L.Z. designed and supervised the research and wrote the manuscript. H.Z.,

824 L.Z. and X.-L.L. performed Cut&Tag, ATAC-seq and RNA-seq experiments; L.Z.

825 performed data analysis; X.-L.L., Y.-M.Y and H.W. performed the rest of the

826 experiments. All authors discussed the results and commented on the manuscript.

827 Competing interests

828 The authors declare no competing interests.

References

1. Armenta-Medina, A. *et al.* Developmental and genomic architecture of plant embryogenesis: from model plant to crops. *Plant Commun* **2**, 100136 (2021).

2. Dresselhaus, T. & Jurgens, G. Comparative Embryogenesis in Angiosperms: Activation and Patterning of Embryonic Cell Lineages. *Annu Rev Plant Biol* **72**, 641-676 (2021).

3. Gerri, C., Menchero, S., Mahadevaiah, S.K., Turner, J.M.A. & Niakan, K.K. Human Embryogenesis: A Comparative Perspective. *Annu Rev Cell Dev Biol* **36**, 411-440 (2020).

4. Xu, R., Li, C., Liu, X. & Gao, S. Insights into epigenetic patterns in mammalian early embryos. *Protein & Cell* **12**, 7-28 %U https://link.springer.com/10.1007/s13238-020-00757-z (2021).

5. Drost, H.G., Gabel, A., Grosse, I. & Quint, M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol* **32**, 1221-31 (2015).

6. Quint, M. *et al.* A transcriptomic hourglass in plant embryogenesis. *Nature* **490**, 98-101 (2012).

7. Nodine, M.D. & Bartel, D.P. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**, 94-7 (2012).

8. Zhao, P., Zhou, X., Zheng, Y., Ren, Y. & Sun, M.-X. Equal parental contribution to the transcriptome is not equal control of embryogenesis. *Nature Plants* **6**, 1354-1364 (2020).

9. Trevino, A.E. *et al.* Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, eaay1645 %U https://www.sciencemag.org/lookup/doi/10.1126/science.aay1645 (2020).

10. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642-54 (2013).

11. Sharma, S., Kelly, T.K. & Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).

12. Baulcombe, D.C. & Dean, C. Epigenetic regulation in plant responses to the environment. *Cold Spring Harb Perspect Biol* **6**, a019471 (2014).

13. Cheng, Y. *et al.* Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct Target Ther* **4**, 62 (2019).

14. Corces, M.R. *et al.* Lineage-specific and single-cell chromatin accessibility

863   charts human hematopoiesis and leukemia evolution. *Nat Genet* **48**, 1193-203
864   (2016).

865   15.   Xia, W. *et al.* Resetting histone modifications during human parental-to-zygotic
866   transition. *Science* **365**, 353-360 (2019).

867   16.   Xu, Q. & Xie, W. Epigenome in Early Mammalian Development: Inheritance,
868   Reprogramming and Establishment. *Trends Cell Biol* **28**, 237-253 (2018).

869   17.   Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3
870   in early mammalian development. *Nature* **537**, 553-557 (2016).

871   18.   Liu, X. *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains
872   in pre-implantation embryos. *Nature* **537**, 558-562 (2016).

873   19.   Zenk, F. *et al.* Germ line-inherited H3K27me3 restricts enhancer function
874   during maternal-to-zygotic transition. *Science* **357**, 212-216 (2017).

875   20.   Huang, J.J. *et al.* Impairment of Preimplantation Porcine Embryo Development
876   by Histone Demethylase KDM5B Knockdown Through Disturbance of
877   Bivalent H3K4me3-H3K27me3 Modifications. *Biology of Reproduction*
878   **92**(2015).

879   21.   Burton, A. & Torres-Padilla, M.E. Chromatin dynamics in the regulation of cell
880   fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol* **15**, 723-34
881   (2014).

882   22.   Wu, J. *et al.* The landscape of accessible chromatin in mammalian
883   preimplantation embryos. *Nature* **534**, 652-7 (2016).

884   23.   Lu, F. *et al.* Establishing Chromatin Regulatory Landscape during Mouse
885   Preimplantation Development. *Cell* **165**, 1375-1388 (2016).

886   24.   Radoeva, T. & Weijers, D. A roadmap to embryo identity in plants. *Trends Plant*
887   *Sci* **19**, 709-16 (2014).

888   25.   Yi, F. *et al.* High Temporal-Resolution Transcriptome Landscape of Early
889   Maize Seed Development. *Plant Cell* **31**, 974-992 (2019).

890   26.   Radoeva, T. *et al.* A Robust Auxin Response Network Controls Embryo and
891   Suspensor Development through a Basic Helix Loop Helix Transcriptional
892   Module. *Plant Cell* **31**, 52-67 (2019).

893   27.   Xiang, D. *et al.* The Transcriptional Landscape of Polyploid Wheats and Their
894   Diploid Ancestors during Embryogenesis and Grain Development. *Plant Cell*
895   **31**, 2888-2911 (2019).

896   28.   Hofmann, F., Schon, M.A. & Nodine, M.D. The embryonic transcriptome of
897   Arabidopsis thaliana. *Plant Reprod* **32**, 77-91 (2019).

898 29. Kawakatsu, T., Nery, J.R., Castanon, R. & Ecker, J.R. Dynamic DNA
899   methylation reconfiguration during seed development and germination.
900   *Genome Biol* **18**, 171 (2017).

901 30. El Baidouri, M. *et al.* Reconciling the evolutionary origin of bread wheat
902   (Triticum aestivum). *New Phytol* **213**, 1477-1486 (2017).

903 31. Consortium, T.I.W.G.S. *et al.* Shifting the limits in wheat research and breeding
904   using a fully annotated reference genome. *Science* **361**(2018).

905 32. Vijverberg, K., Ozias-Akins, P. & Schranz, M.E. Identifying and Engineering
906   Genes for Parthenogenesis in Plants. *Front Plant Sci* **10**, 128 (2019).

907 33. Kumar, V. & Van Staden, J. New insights into plant somatic embryogenesis: an
908   epigenetic view. *Acta Physiologiae Plantarum* **39**, 1-17 (2017).

909 34. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method
910   for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*
911   **109**, 21 29 1-21 29 9 (2015).

912 35. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small
913   samples and single cells. *Nat Commun* **10**, 1930 (2019).

914 36. Lu, Z. *et al.* The prevalence, evolution and chromatin signatures of plant
915   regulatory elements. *Nat Plants* **5**, 1250-1259 (2019).

916 37. Klemm, S.L., Shipony, Z. & Greenleaf, W.J. Chromatin accessibility and the
917   regulatory epigenome. *Nat Rev Genet* **20**, 207-220 (2019).

918 38. Corces, M.R. *et al.* The chromatin accessibility landscape of primary human
919   cancers. *Science* **362**, eaav1898 %U
920   https://www.science.org/doi/10.1126/science.aav1898 (2018).

921 39. Giaimo, B.D., Ferrante, F., Herchenrother, A., Hake, S.B. & Borggrefe, T. The
922   histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* **12**, 37 (2019).

923 40. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with
924   ChromHMM. *Nat Protoc* **12**, 2478-2492 (2017).

925 41. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
926   characterization. *Nat Methods* **9**, 215-6 (2012).

927 42. Gorkin, D.U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal
928   development. *Nature* **583**, 744-751 (2020).

929 43. Long, J. *et al.* Nurse cell--derived small RNAs define paternal epigenetic
930   inheritance in Arabidopsis. *Science* **373**, eabh0556 (2021).

931 44. Anderson, S.N. *et al.* The Zygotic Transition Is Initiated in Unicellular Plant
932   Zygotes with Asymmetric Activation of Parental Genomes. *Dev Cell* **43**,

933   349-358 e4 (2017).

934   45.   Schulz, K.N. & Harrison, M.M. Mechanisms regulating zygotic genome
935   activation. *Nat Rev Genet* **20**, 221-234 (2019).

936   46.   Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb
937   repressive complex 2 in Arabidopsis. *Nat Genet* **49**, 1546-1552 (2017).

938   47.   Bieluszewski, T., Xiao, J., Yang, Y. & Wagner, D. PRC2 activity, recruitment,
939   and silencing: a comparative perspective. *Trends Plant Sci* **26**, 1186-1198
940   (2021).

941   48.   Willemsen, V. & Scheres, B. Mechanisms of pattern formation in plant
942   embryogenesis. *Annu Rev Genet* **38**, 587-614 (2004).

943   49.   Horstman, A., Bemer, M. & Boutilier, K. A transcriptional view on somatic
944   embryogenesis. *Regeneration (Oxf)* **4**, 201-216 (2017).

945   50.   Wang, F.X. *et al.* Chromatin Accessibility Dynamics and a Hierarchical
946   Transcriptional   Regulatory   Network   Structure   for   Plant   Somatic
947   Embryogenesis. *Dev Cell* **54**, 742-757 e8 (2020).

948   51.   Su, Y.H., Tang, L.P., Zhao, X.Y. & Zhang, X.S. Plant cell totipotency: Insights
949   into cellular reprogramming. *Journal of Integrative Plant Biology* **63**, 228-243
950   (2021).

951   52.   Baud, S. & Lepiniec, L. Physiological and developmental regulation of seed oil
952   production. *Prog Lipid Res* **49**, 235-49 (2010).

953   53.   Skubacz, A., Daszkowska-Golec, A. & Szarejko, I. The Role and Regulation of
954   ABI5 (ABA-Insensitive 5) in Plant Development, Abiotic Stress Responses and
955   Phytohormone Crosstalk. *Front Plant Sci* **7**, 1884 (2016).

956   54.   Ikeuchi, M., Ogawa, Y., Iwase, A. & Sugimoto, K. Plant regeneration: cellular
957   origins and molecular mechanisms. *Development* **143**, 1442-51 (2016).

958   55.   Ramirez-Gonzalez, R.H. *et al.* The transcriptional landscape of polyploid wheat.
959   *Science*   **361**,   eaar6089   %U
960   https://www.sciencemag.org/lookup/doi/10.1126/science.aar6089 (2018).

961   56.   Simonini, S. *et al.* The Polycomb group protein MEDEA controls cell
962   proliferation and embryonic patterning in Arabidopsis. *Dev Cell* **56**, 1945-1960
963   e7 (2021).

964   57.   Armenta-Medina, A. *et al.* Developmental and genomic architecture of plant
965   embryogenesis: from model plant to crops. *Plant Communications* **2**, 100136
966   (2021).

967   58.   Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ

968   preprocessor. *Bioinformatics* **34**, i884-i890 (2018).

969   59.   Li, H. & Durbin, R. Fast and accurate short read alignment with
970   Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).

971   60.   Kim, D., Paggi, J.M., Park, C., Bennett, C. & Salzberg, S.L. Graph-based
972   genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
973   *Biotechnol* **37**, 907-915 (2019).

974   61.   Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose
975   program for assigning sequence reads to genomic features. *Bioinformatics* **30**,
976   923-30 (2014).

977   62.   Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor
978   package for differential expression analysis of digital gene expression data.
979   *Bioinformatics* **26**, 139-40 (2010).

980   63.   Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation
981   network analysis. *BMC Bioinformatics* **9**, 559 (2008).

982   64.   Yu, G.C., Wang, L.G., Han, Y.Y. & He, Q.Y. clusterProfiler: an R Package for
983   Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of*
984   *Integrative Biology* **16**, 284-287 (2012).

985   65.   Zheng, Y. CUT&Tag Data Processing and Analysis Tutorial. (2020).

986   66.   Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. & Manke, T. deepTools: a
987   flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**,
988   W187-91 (2014).

989   67.   Jin, H.J. *et al.* ChIPseqSpikeInFree: a ChIP-seq normalization approach to
990   reveal global changes in histone modifications without spike-in. *Bioinformatics*
991   **36**, 1270-1272 (2020).

992   68.   Meers, M.P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment
993   Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42
994   (2019).

995   69.   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**,
996   R137 (2008).

997   70.   Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing
998   genomic features. *Bioinformatics* **26**, 841-2 (2010).

999   71.   Yu, G., Wang, L.G. & He, Q.Y. ChIPseeker: an R/Bioconductor package for
1000   ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-3
1001   (2015).

1002   72.   Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and

1003   dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

1004   73.   Gusmao, E.G., Dieterich, C., Zenke, M. & Costa, I.G. Detection of active
1005   transcription factor binding sites with the combination of DNase
1006   hypersensitivity and histone modifications. *Bioinformatics* **30**, 3143-51 (2014).

1007   74.   Khan, A. *et al.* JASPAR 2018: update of the open-access database of
1008   transcription factor binding profiles and its web framework. *Nucleic acids*
1009   *research* **46**, D260-D266 (2018).

1010   75.   Heinz, S. *et al.* Simple combinations of lineage-determining transcription
1011   factors prime cis-regulatory elements required for macrophage and B cell
1012   identities. *Mol Cell* **38**, 576-89 (2010).

1013   76.   Meinke, D.W. Genome-wide identification of EMBRYO-DEFECTIVE (EMB)
1014   genes required for growth and development in Arabidopsis. *New Phytol* **226**,
1015   306-325 (2020).

1016   77.   Buchfink, B., Xie, C. & Huson, D.H. Fast and sensitive protein alignment using
1017   DIAMOND. *Nature Methods* **12**, 59-60 (2015).

1018   78.   Buchfink, B., Reuter, K. & Drost, H.G. Sensitive protein alignments at
1019   tree-of-life scale using DIAMOND. *Nature Methods* **18**, 366-+ (2021).

1020   79.   Hao, Z. *et al.* Conserved, divergent and heterochronic gene expression during
1021   Brachypodium and Arabidopsis embryo development. *Plant Reprod* **34**,
1022   207-224 (2021).

1023   80.   Leiboff, S. & Hake, S. Reconstructing the Transcriptional Ontogeny of Maize
1024   and Sorghum Supports an Inverse Hourglass Model of Inflorescence
1025   Development. *Curr Biol* **29**, 3410-3419 e3 (2019).

1026   81.   Levin, M. *et al.* The mid-developmental transition and the evolution of animal
1027   body plans. *Nature* **531**, 637-641 (2016).

1028   82.   Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and
1029   correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-9
1030   (2016).

1031
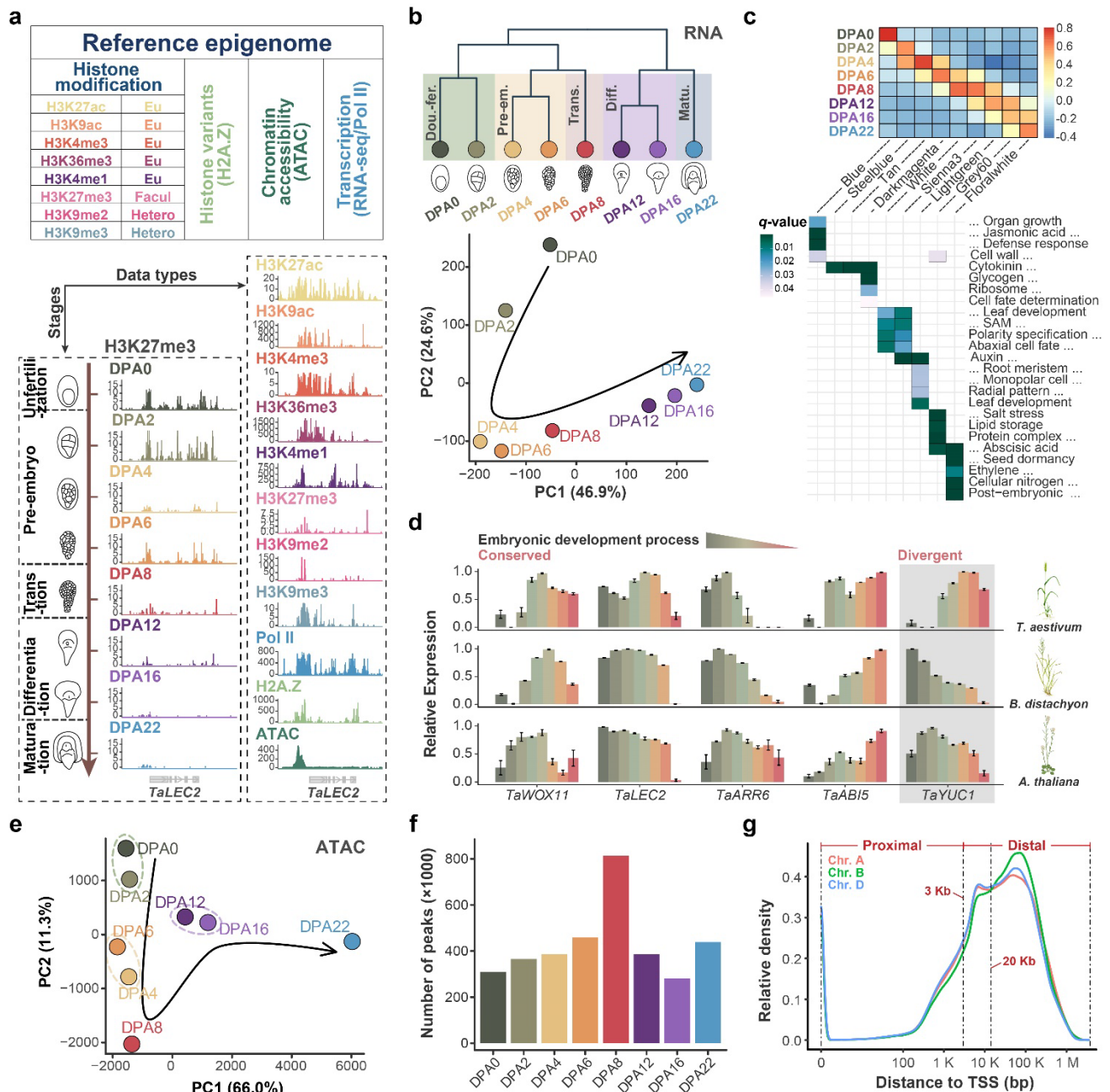
Fig. 1 | Charting chromatin landscapes of wheat embryogenesis

a, Experimental design and major axes of the data series: data types and sampling developmental stages, DPA: days post anthesis.

b, Cluster dendrogram and PCA of transcriptome showing five distinct development stages: around double fertilization (DPA 0/2), pre-embryo (DPA 4/6), transition (DPA 8), differentiation (DPA 12/16), and maturation (DPA 22).

c, Representative modules from WGCNA cluster and enrichment of corresponding GO.

d, The conserved and diversified expression pattern of representative genes between monocot (*T. aestivum* and *B. distachyon*) and dicot (*A. thaliana*) during embryogenesis.

e-g, Dynamic pattern of genome-wide chromatin accessibility measured by ATACseq during embryogenesis. PCA analysis (**e**), peak numbers dynamics (**f**), and peak distribution in different sub-genomes (**g**).
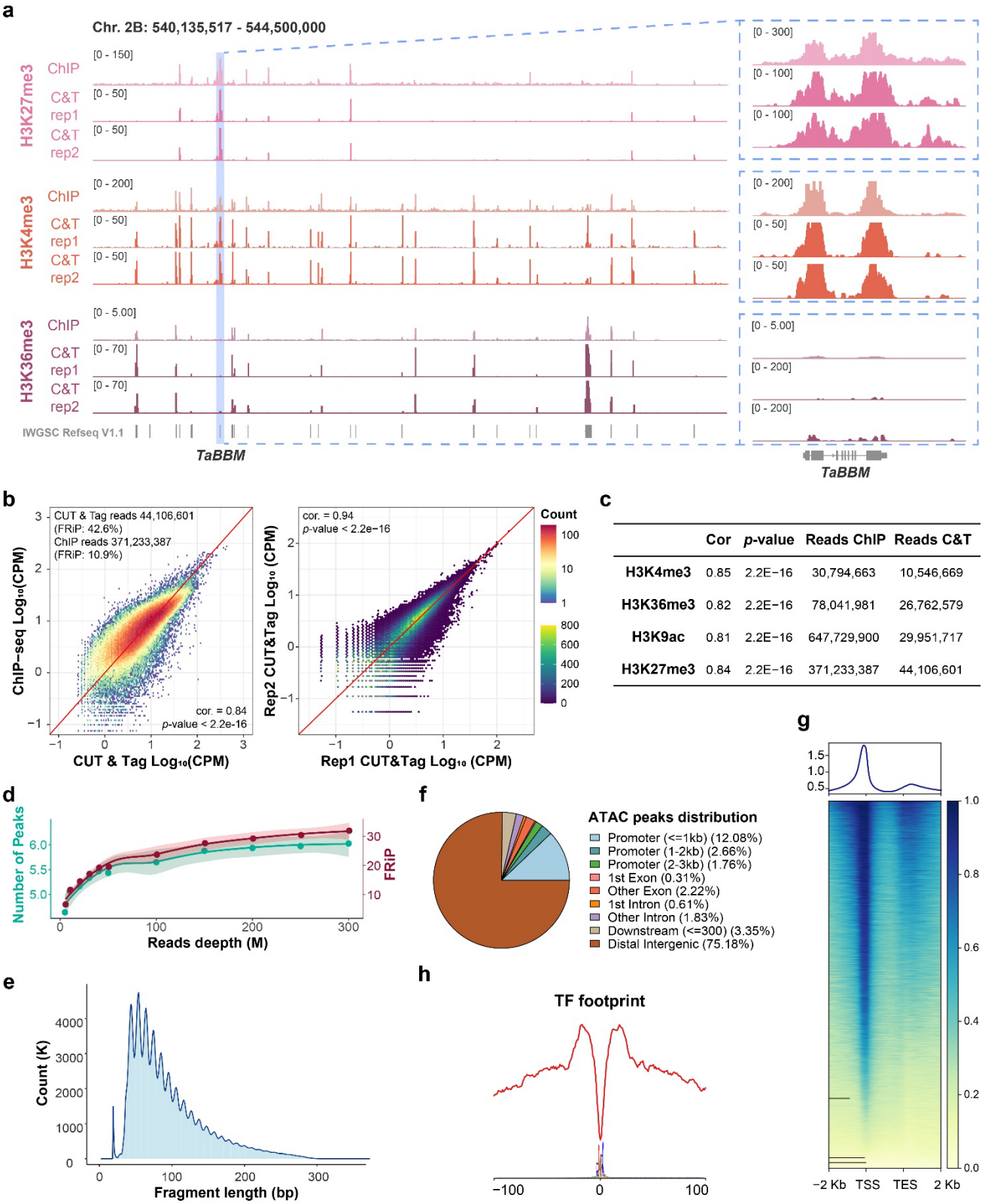
Fig. S1 | Quality control for CUT & Tag and ATAC-seq

a-c, CUT & Tag recapture ChIPseq signal with high reproducibility and low sequencing depth. IGV (Integrated Genome Browser) showing high reproducibility between ChIP-seq and CUT & Tag in seedlings (**a**). Correlation between ChIP-seq and CUT & Tag as well as two biological replicates (**b**). Sequencing depth comparison for various histone modifications between CUT&Tag and ChIPseq (**c**).

d-h, Quality control of ATAC-seq data. Peaks number, and FRiP (The fraction of reads in called peak regions) were calculated at different sequence depths (**d**). Fragment size distribution (**e**) and peak distribution within the genome (**f**). ATAC-seq signal distribution along with genes (**g**). Transcription factor footprint profile of ATAC-seq data (**h**).
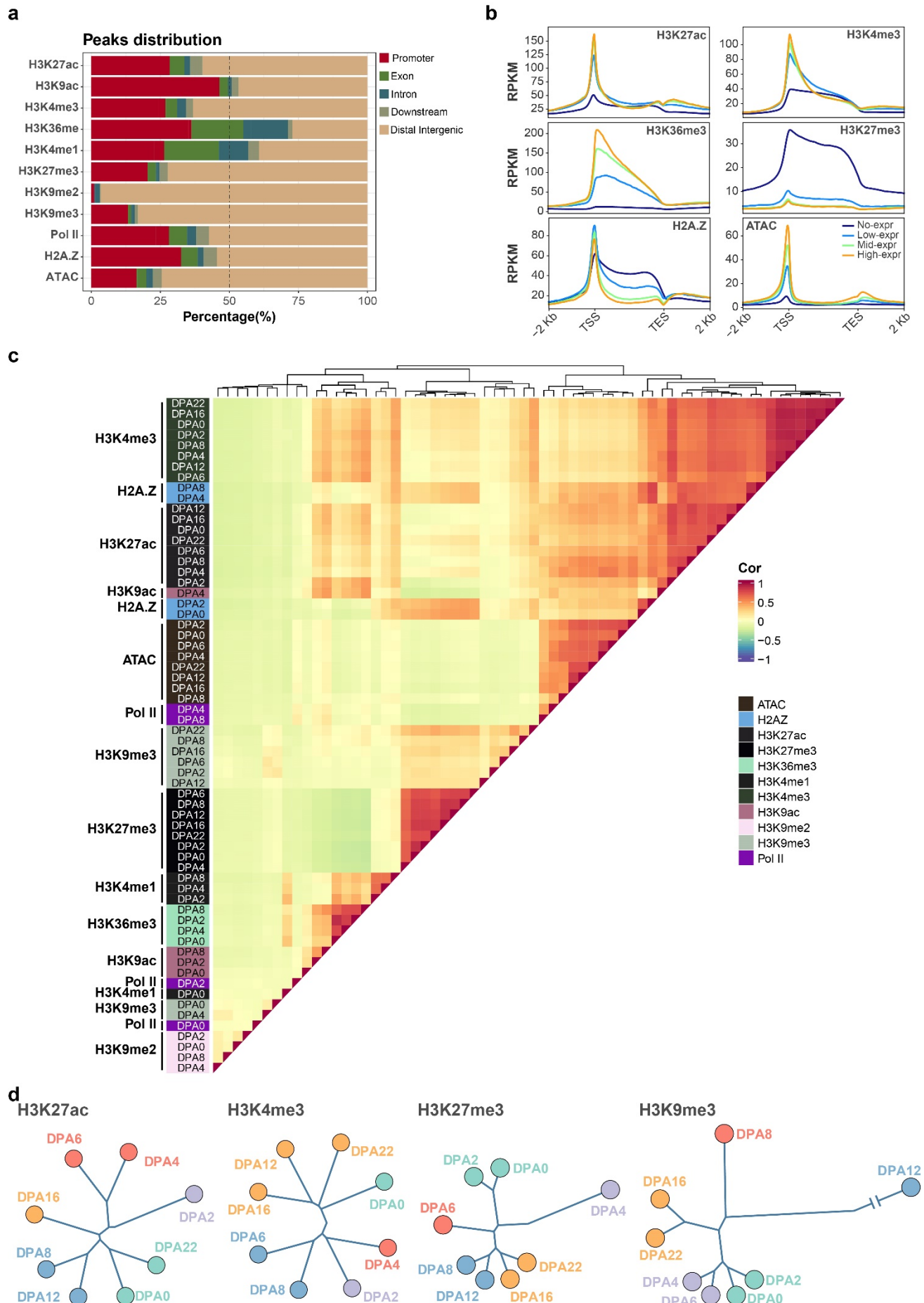
Fig. S2 | Features of various histone modifications

a, Peak distribution of different histone modifications relative to genes.

b, Correlation between different types of histone modification profiles and gene expression levels.

c, Pair-wise correlation map among different histone modification profiles. Jaccard index was calculated based on the peaks overlap, and then Pearson correlation scores were generated.

d, Cluster dendrogram of histone modifications. Reads count in peaks were normalized by CPM, and euclidean distances were calculated for tree clustering.
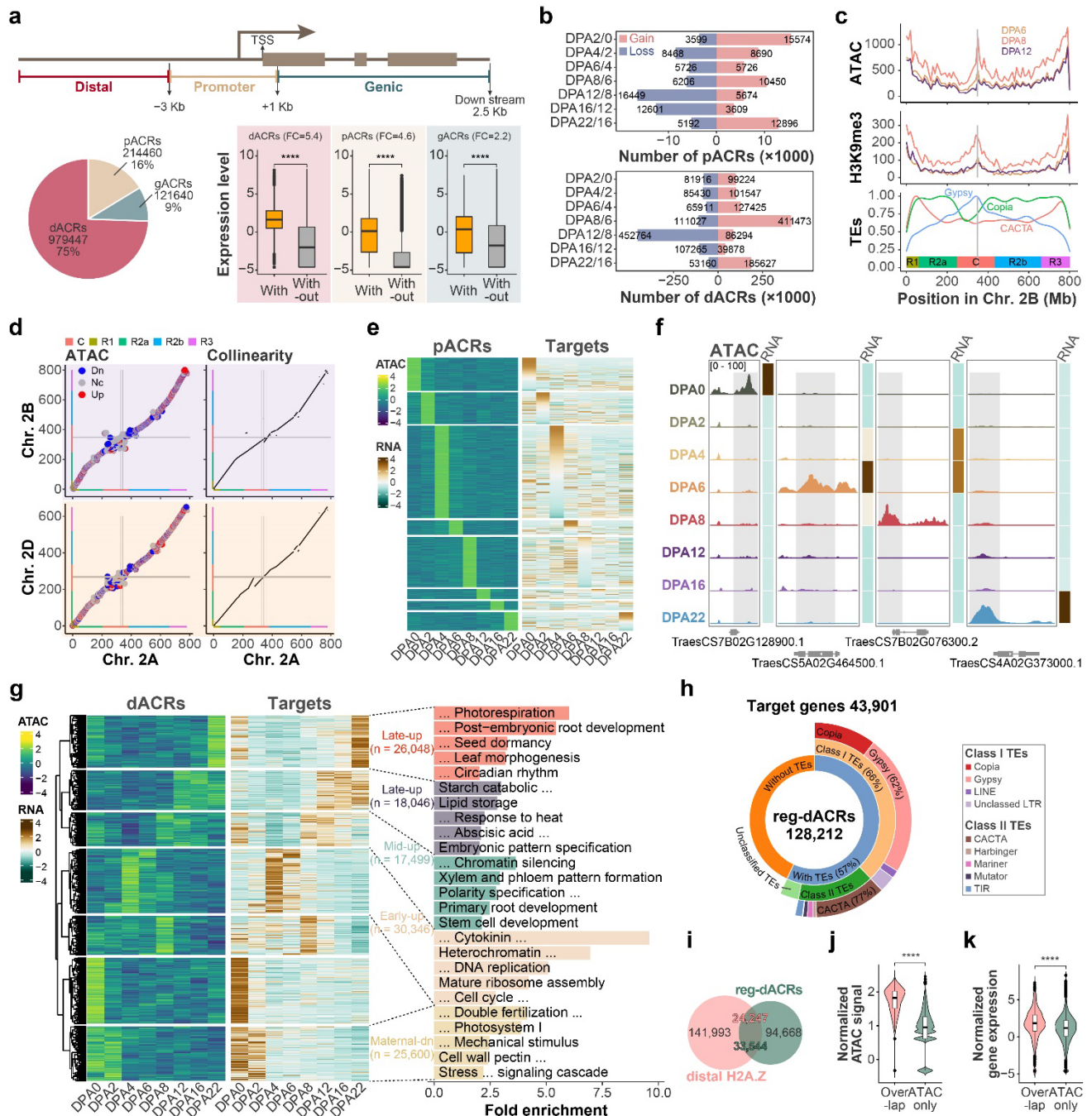
Fig. 2 | Distinct feature of different ACRs in transcriptional regulation in wheat

a, Subdividing ACRs based on distribution pattern around genes, and correlation between different types of ACRs and genes expression.

b, Distinct gain and loss of pACRs and dACRs between adjacent embryonic developmental stages.

c, ATAC peaks, H3K9me3 and various types of TEs distributions on chromosome 2B at different embryogenesis stages.

d, Comparisons of chromatin accessibility of collinear regions among subgenomes, chromosome 2B compared to 2A (up-left) and chromosome 2D with 2A (down-left). Collinear regions were identified based on sequence conservation (right). Color bars on x and y axis indicate the chromosome segments defined by **Consortium (IWGSC) et al., 2018. Grey lines indicate the centromere** for individual chromosomes.

e, Stage-specific pACRs and corresponding genes expression.

f, Representative genes in **e.**

g, Heatmap showing correlation between reg-dACRs and corresponding genes (left) and GO enrichment for different genes sets that were clustered based on the linking reg-dACRs dynamic (right).

h, Fraction of reg-dACRs overlapped with various types of TEs.

i, Overlap between reg-dACRs and distal H2A.Z histone variant.

j and k, Comparisons of ATAC signal intensity **(j) and target**s **expression (k) between** reg-dACRs overlapped with H2A.Z and those not overlapped. Mann-Whitney U test (two-sided) was used for **a**, **j** and **k**. * : $p \le 0.05$; ** : $p \le 0.01$; ***: $p \le 0.001$; ****: $p \le 0.0001$.
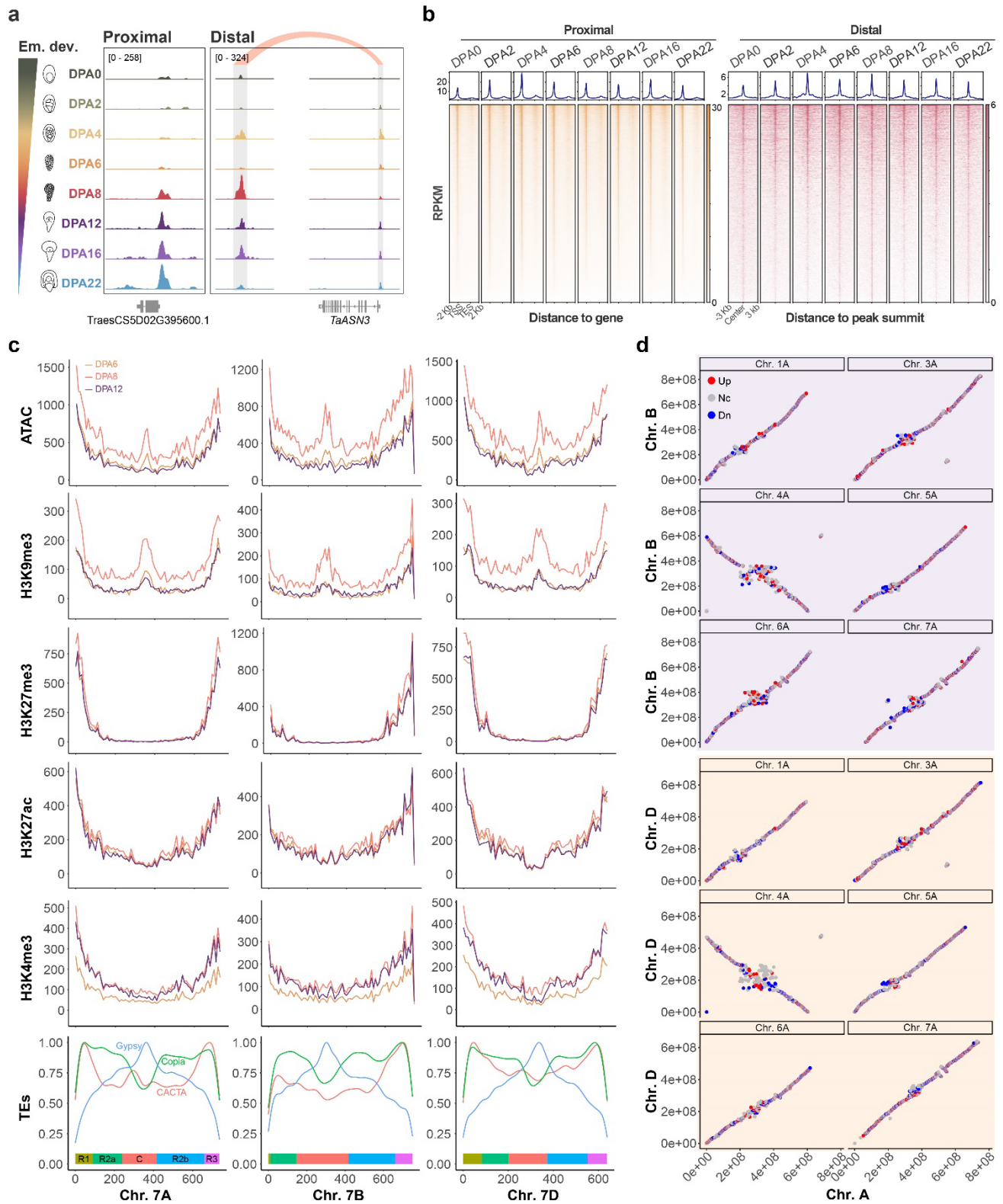
Fig. S3 | Dynamic ACRs during wheat embryogenesis and difference among sub-genomes

a, IGV showing chromatin accessibility change cross wheat embryogenesis in both proximal and distal regions.

b, ACRs dynamic in proximal (left) and distal regions (right) during wheat embryogenesis.

c, Chromatin accessibility (ATAC), histone modifications and various types of TEs distributions change on Chr. 7 of A-, B- and D-subgenome at DPA 6, 8 and 12. Color bars on x and y axis indicate the chromosome segments defined by Consortium (IWGSC) et al., 2018.

d, Comparisons of chromatin accessibility of collinear regions among subgenomes. Down- (Dn), and Up-regulated (Up) loci were define by the $\log_2$(fold change Chr. B/Chr. A) (left top-panel) and $\log_2$(fold change Chr. D/Chr. A) (left bottom-panel) values bigger and lower than 1 and -1. Nc indicates no change between subgenomes.
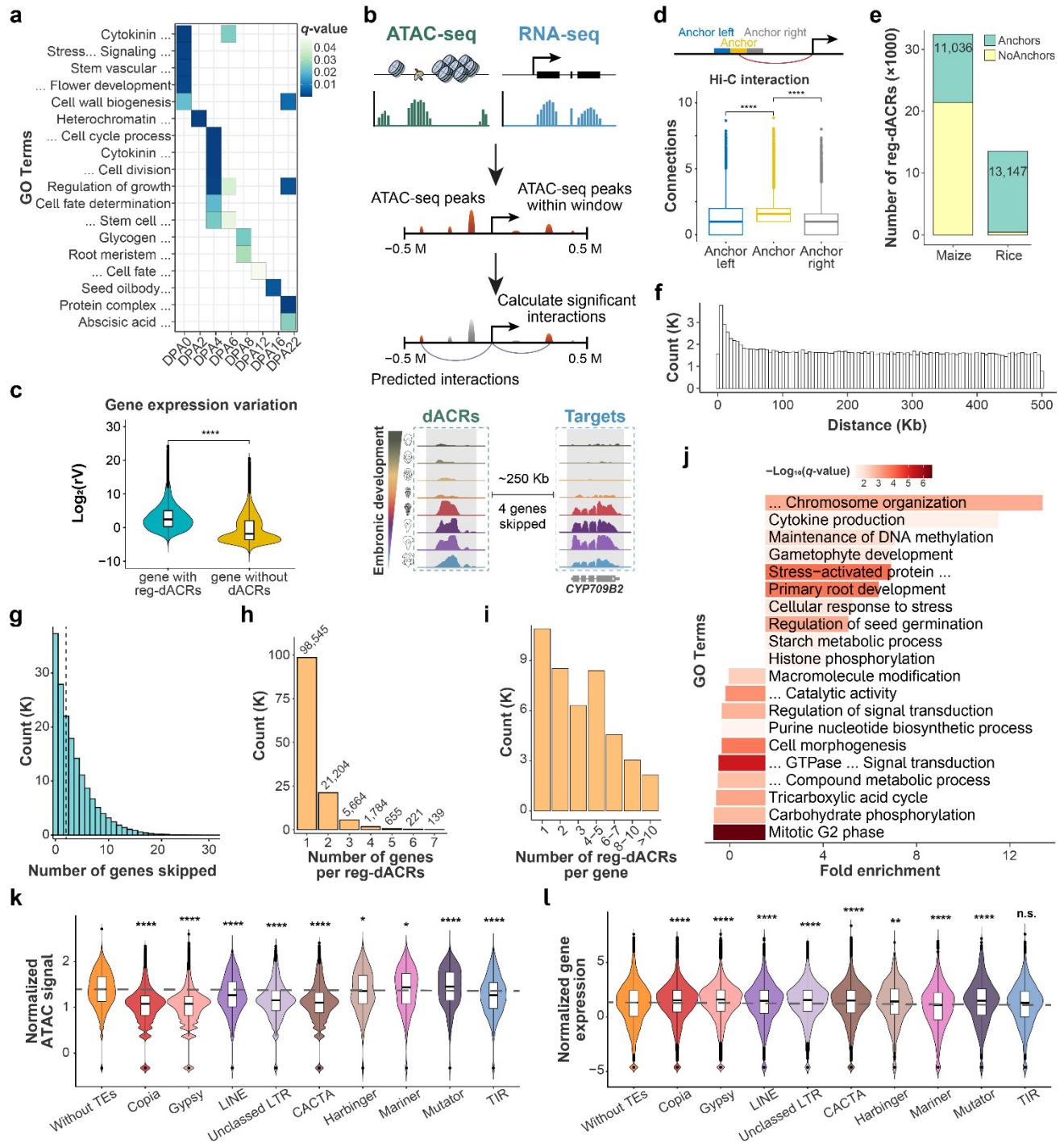
Fig. S4 | Association between reg-dACRs to genes

a, GO enrichment of stage-specific genes with pACRs.

b, Schematic of the strategy used to link distal ATAC-seq peaks to genes (see Method for detail).

c, Relative variation of expression level for genes with dACRs or without.

d, Hi-C anchors highlight the connection between reg-dACRs and target genes.

e, Number of dACRs homologous segment in maize and rice. Based on previously published Hi-C and ChIA-pet data, the homologous segment was grouped in anchors and NoAnchors.

f, g, Distance of reg-dACR to genes (**f**) and the number of genes skipped (**g**) between reg-dACR and target genes.

h, Number of genes can be targeted per reg-dACRs.

i, Number of reg-dACRs distribution for the individual target gene.

j, Corresponding GO enrichment for genes with solo or multiple dACRs.

k, l, Comparison of ATAC-seq signal and corresponding target genes expression among reg-dACRs overlapped by different types of TEs. Mann-Whitney U test (two-sided) was used for **c** and **k** (* : p <= 0.05; ** : p <= 0.01; ***: p <= 0.001).
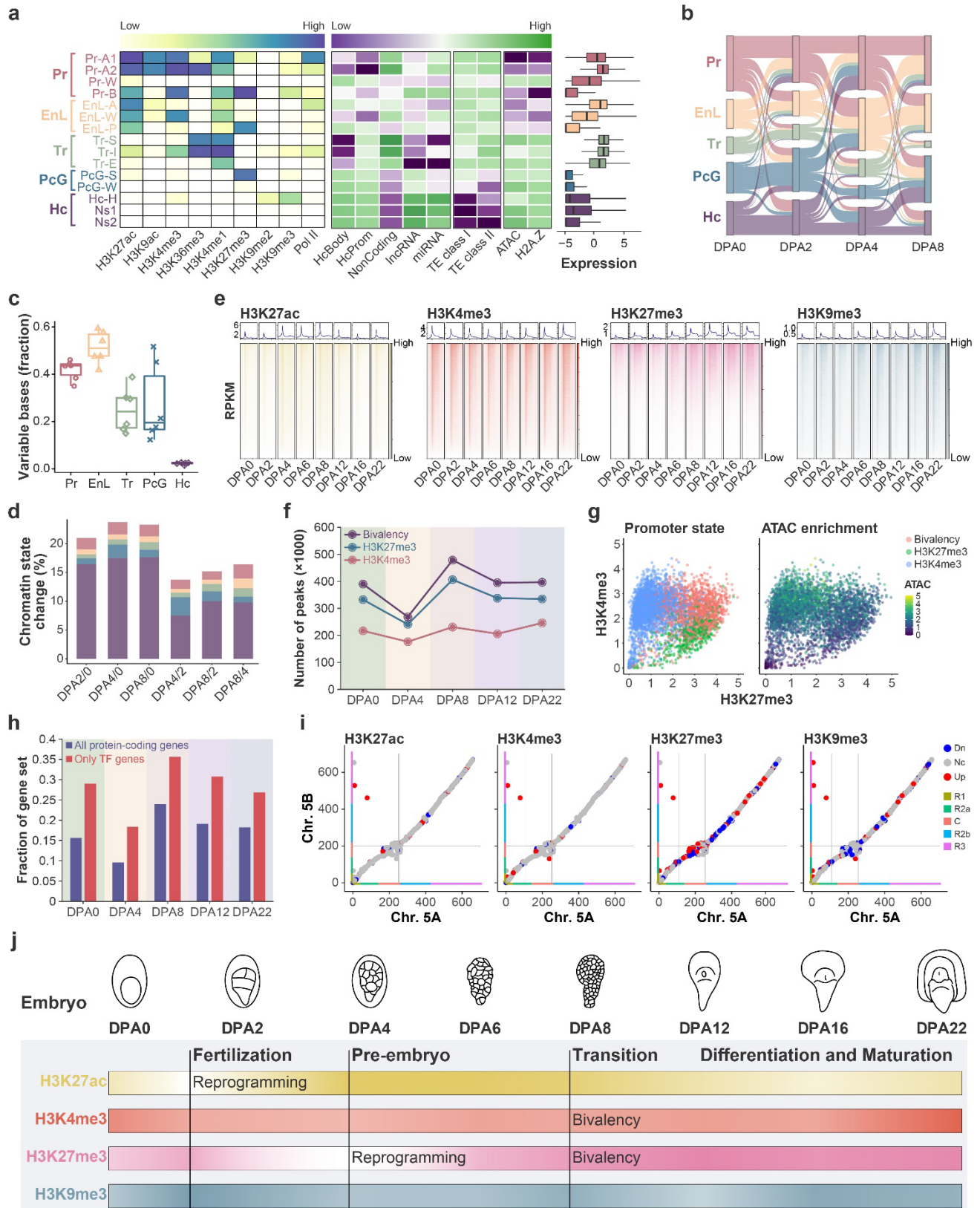
Fig. 3 | Reprogramming of histone modifications during wheat embryogenesis

a, Emission probabilities for histone modifications in 15 ChromHMM states, which could be mainly categorized in five groups: Pr (promoter), EnL (enhancer-like), Tr (transcription), PcG (polycomb group), and Hc (heterochromatin) state. The abbreviation following dash: A: active; W: weak; B: bivalency; P: poised; S: strong; I: initiation; E: elongation; H: H3K9me-associated; Ns: No signal.

b, Chromatin states transition among different developmental stages.

c, Dynamic of five major chromatin states during different embryonic stages.

d, Percentage of changed chromatin state coverage to the total genome size among different embryonic stages. Colors filled represent the five major chromatin state groups as shown in **a**.

e, Histone modification dynamic in promoter and genic regions during wheat embryogenesis.

f, Regions of H3K4me3, H3K27me3 and bivalency at different developmental stages. **g,** Scatter plots correlating H3K27me3 (x axis) and H3K4me3 (y axis) enrichments ($\log_2$CPM) at proximal region. Different colors in the left panel indicate various modification states as labeled and chromatin accessibility of correlated plots is shown in the right panel with color bar for relative intensity ($\log_2$CPM).

h, Fractions of all protein-coding and TF-coding genes marked by bivalent promoter at five major developmental stages.

i, Comparisons of histone modifications H3K27ac, H3K4me3, H3K27me3 and H3K9me3 of collinear regions between Chr. 5A and Chr. 5B. Down- (Dn), and Up-regulated (Up) loci were defined by the $\log_2$(fold change Chr. 5B/ Chr. 5A) values bigger and lower than 1 and -1, respectively**.** Nc indicates no change. Color bars and grey lines indicate the chromosome segments and centromere regions as those in Fig. 2d.

j, Schematic represents the chromatin state dynamic during wheat embryogenesis. The dark and light colors represent the strength and weakness of the signal, respectively.
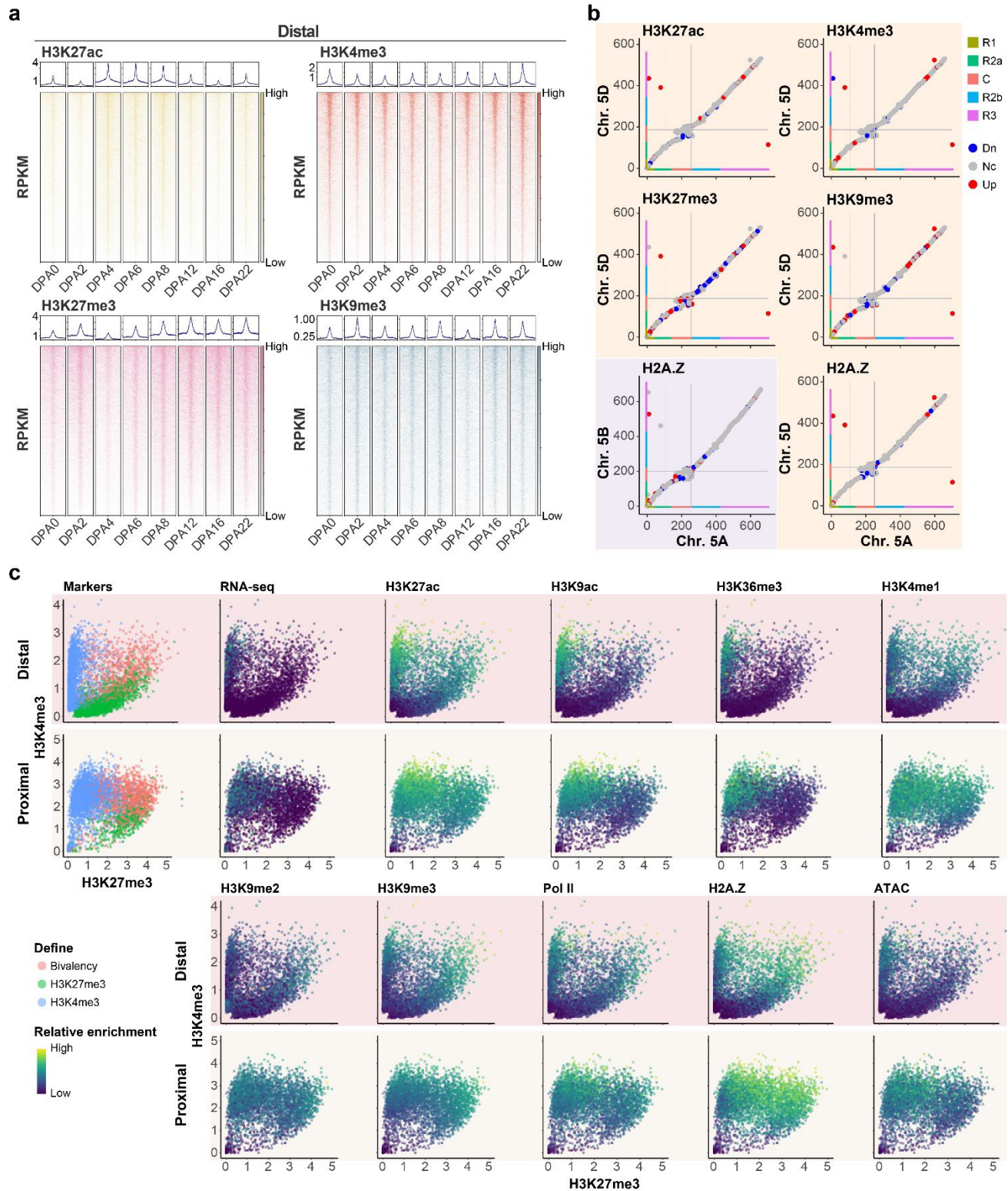
Fig. S5 | Histone modifications dynamic and subgenome comparison

a, Histone modifications dynamic in distal regions during wheat embryogenesis.

b, Comparisons of histone modifications H3K27ac, H3K4me3, H3K27me3, H3K9me3 and H2A.Z of collinear regions between three subgenomes of Chr. 5. Down- (Dn), and Up-regulated (Up) loci were defined by the $\log_2$(fold change Chr. 5B or Chr. 5D / Chr. 5A) values bigger and lower than 1 and -1, respectively. Nc indicates no change. Color bars and grey lines indicate the chromosome segments and centromere regions as those in Fig. 2d.

c, Scatter plots correlating H3K27me3 (x axis) and H3K4me3 (y axis) enrichments ($\log_2$CPM) at distal and proximal regions. Color represents chromatin state defined as indications, gene expression and multiple histone modifications level ($\log_2$CPM).
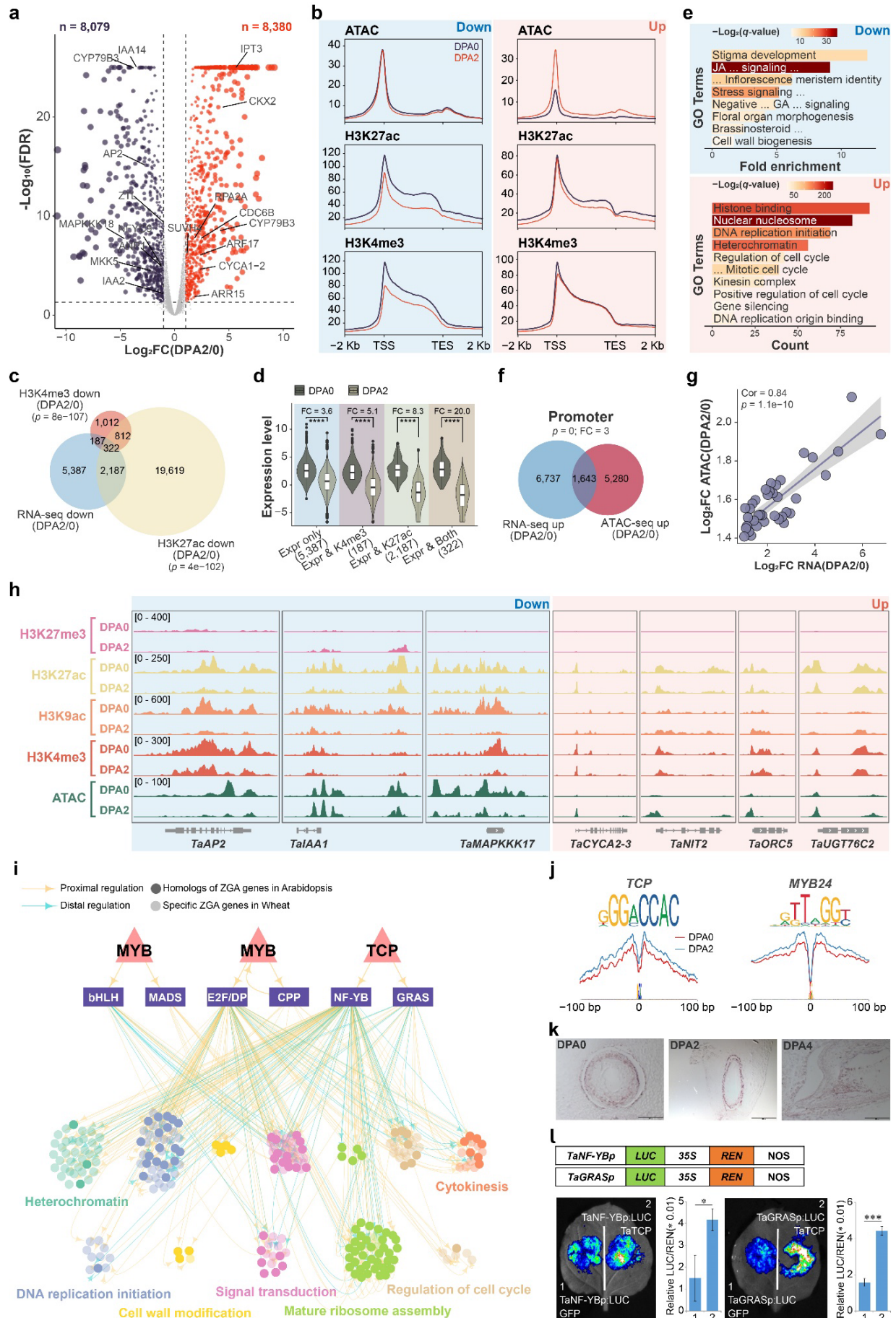
Fig. 4 | Transcription and chromatin reconfiguration during MZT

a, Volcano plot of differentially expressed genes during MZT. DEGs were defined by the threshold value of $\log_2$(Fold Change DPA2/DAP0) ≥ 1 and FDR ≤ 0.05 by edgeR.

b, Chromatin accessibility and histone modification profiles of down- (left) and up-regulated (right) genes.

c, Overlap between down-regulated genes and corresponding active histone modification marks decreasing.

d, Expression levels of different gene sets generated from **c**.

e, GO enrichment for 2,696 down-regulated genes in which active histone modification marks decreased (**c**) (top) and 1,643 up-regulated genes in which gained pACRs (**f**) (bottom).

f, Overlap between up-regulated genes and corresponding chromatin accessibility gains.

g, Correlation between gained pACRs and up-regulated genes based on 1643 genes overlapped. Genes were ranked by RNA-seq fold change and separated into 50 bins.

h, IGV showing histone modifications and chromatin accessibility changes on representative down-regulated genes *AP2*, *MAPKKK17*, as well as *IAA1* (light blue shade), and up-regulated genes *CYCA2-3*, *NIT2*, *ORC5*, as well as *UGT76C2* (light red shade).

i, GRNs for hierarchical regulation of zygotic active genes sets. Both proximal (orange) and distal (green) regulations were included. The Upper and lower layers of GRN represented the predicted regulation hierarchy. Each point represented a TF or gene. The target genes with orthologs of Arabidopsis ZGA genes were represented in light color, while the specific genes were represented in dark color.

j, ATAC-seq footprint of key TFs in **i** during MZT.

k, Expression patterns of *TCP* by *in situ* at DPA 0, 2 and 4.

l, Luciferase reporter assays validation of transcriptional activation capability of TCP on targets *TaNF-YB* and *TaGRAS*.

Hypergeometric distribution test was used for **c** and **f**. Mann-Whitney U test (two-sided) and Student's t test were used for **d** and **l**, respectively (* : $p <= 0.05$; ** : $p <= 0.01$; ***: $p <= 0.001$; ****: $p <= 0.0001$). Pearson correlation test was used in **g**.
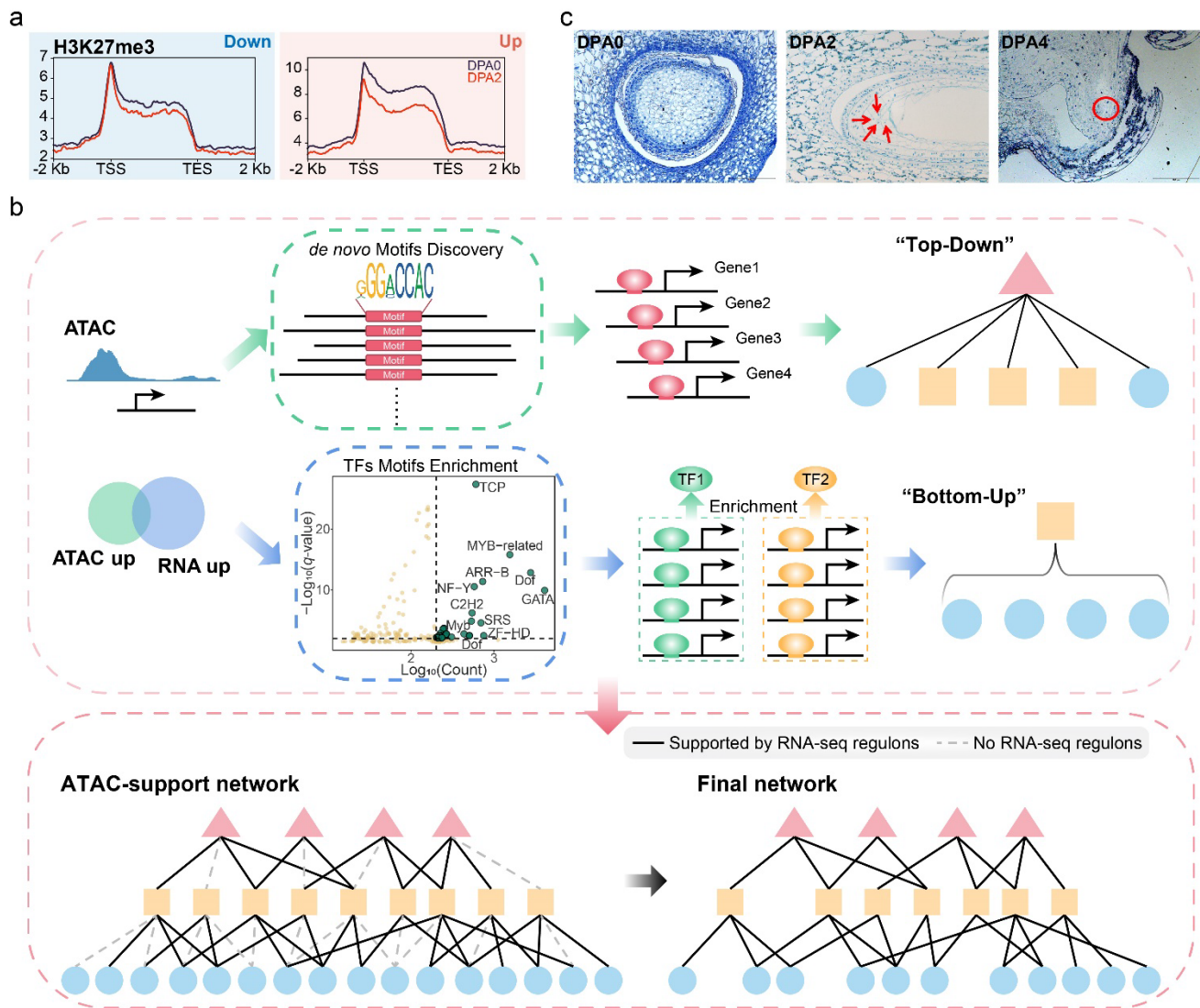
Fig. S6 | Histone modification change and build-up of transcription regulation network for ZGA

a, H3K27me3 change following fertilization at dn- and up-regulation gene sets as Fig. 4b.

b, Schematic of the strategy for GRNs building (see method for detail).

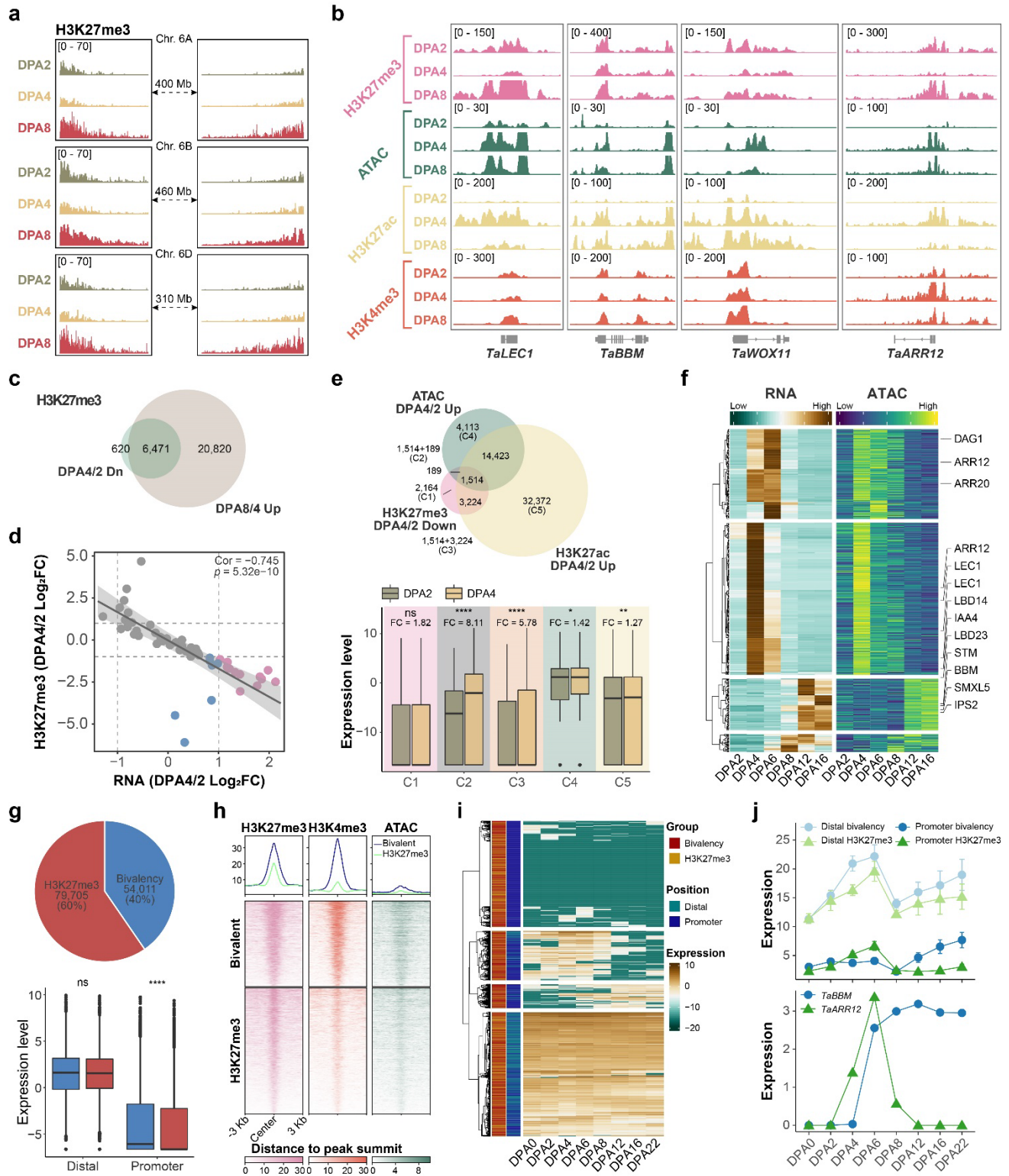**c**, Morphological change of embryo-sac during early embryogenesis.

Fig. 5 | H3K27me3 reprogramming ensures pre-embryo development

a, Dynamic of H3K27me3 modification in Chr. 6 during embryogenesis, with resetting at DPA4 and regaining at DAP8.

b, IGV showing histone modifications and chromatin accessibility changes on embryo developmental essential genes.

c, Overlap between genes with H3K27me3 decreasing at DPA4 and increasing at DPA8.

d, Correlation between H3K27me3 decreasing and target genes' activation from DAP4 to DPA2. Genes were ranked by RNA-seq fold change and separated into 50 bins. H3K27me3 loss with mRNA level increased genes are shown in red color, while loss of H3K27me3 without mRNA changed is shown in blue color, while gray color indicates no change of H3K27me3.

e, Overlap among genes with down-regulated H3K27me3 and up-regulated ATAC-seq and H3K27ac modification (top), and altered expression levels of different gene sets (bottom).

f, Synchronous pattern between the gain of chromatin accessibility and elevation of gene expression on the basis of H3K27me3 decreasing at DPA4. Several developmental essential genes were highlighted.

g, Proportion of H3K27me3 lonely and bivalent regions following H3K27me3 increasing at DPA8 (top), and the expression level of corresponding target genes. Distal targets assignment was based on reg-dACR-target links from Fig. 2h.

h, Histone modifications and ATAC-seq enrichment at H3K27me3 lonely or bivalent regions.

i.j, Expression pattern of genes associated with H3K27me3 lonely or bivalent modification across wheat embryogenesis. Heatmap showing clustering of gene expression (i), Line chart indicates the mean expression of various types of genes as indicated (top), as well as representative genes *TaBBM* (promoter bivalency) and *TaARR12* (promoter H3K27me3 alone) were shown (bottom) (**j**).
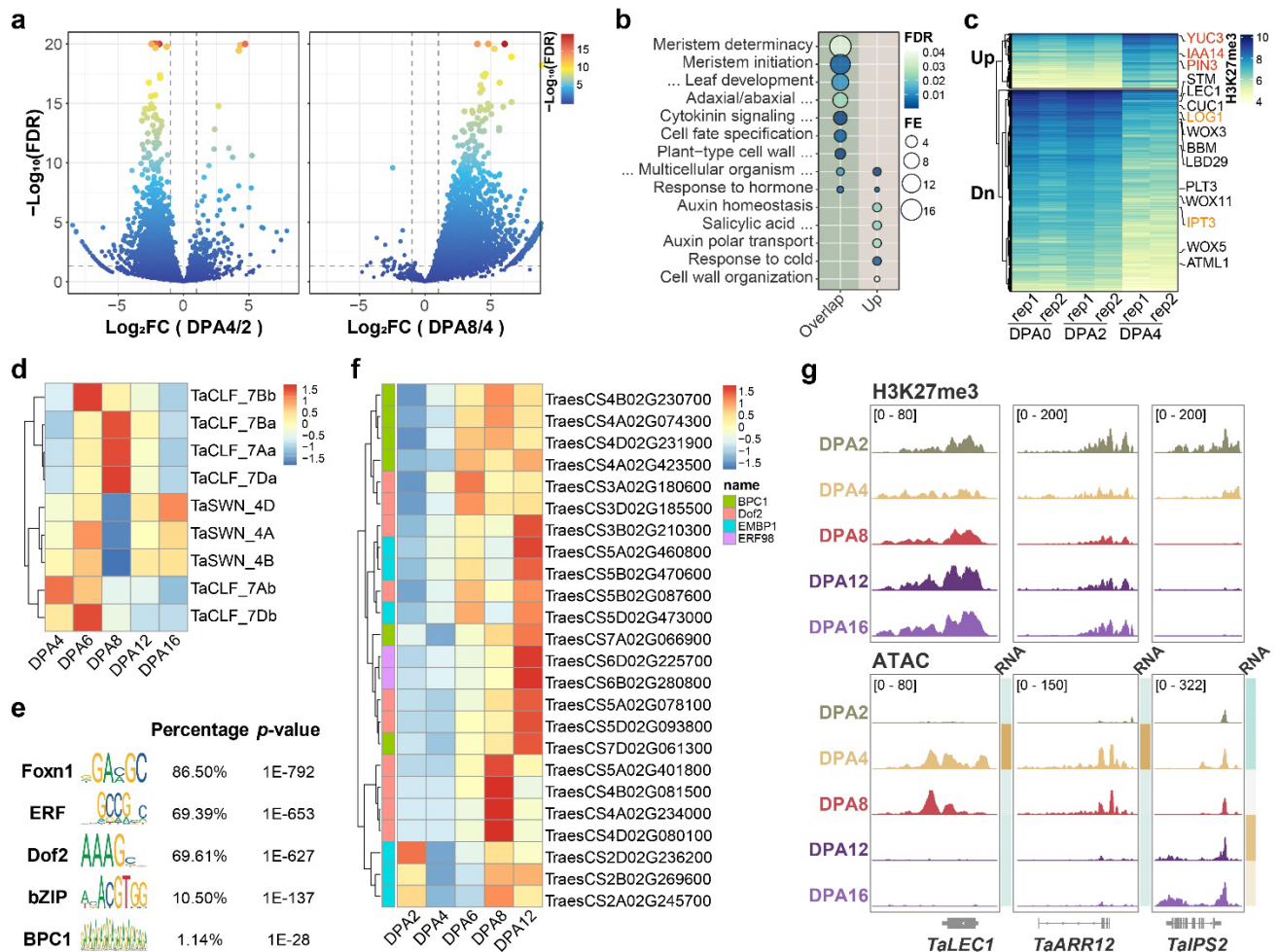
Fig. S7 | H3K27me3 reprogramming at pre-embryo

a, Volcano plot of differential H3K27me3 enriched regions between DPA2 and DPA4 (left), and between DPA4 and DPA8 (right). Differential regions were defined by the threshold absolute value of $\log_2$(Fold Change) ≥ 1 and FDR ≤ 0.05 by DESeq2.

b, GO enrichment of gene sets which lost H3K27me3 at DPA4 and re-gain at DPA8, and gene sets which *de novo* gain of H3K27me3 at DPA8.

c, Heatmap showing differential H3K27me3 modification genes between DAP4 and DAP2. Embryonic development essential genes were highlighted. Auxin-related genes were labeled in red and CK-related genes were labeled in orange.

d, The expression level of H3K27me3 methyltransferase coding genes in selected embryonic development stages.

e, f, Motif enrichment of H3K27me3 deposition loci at DAP8 (**e**) and the expression levels of cognate binding TFs (**f**). Homer was used for motif enrichment analysis.

g, H3K27me3 and chromatin accessibility dynamics at embryo essential genes *TaLEC1, TaARR12* and *TaIPS2*.
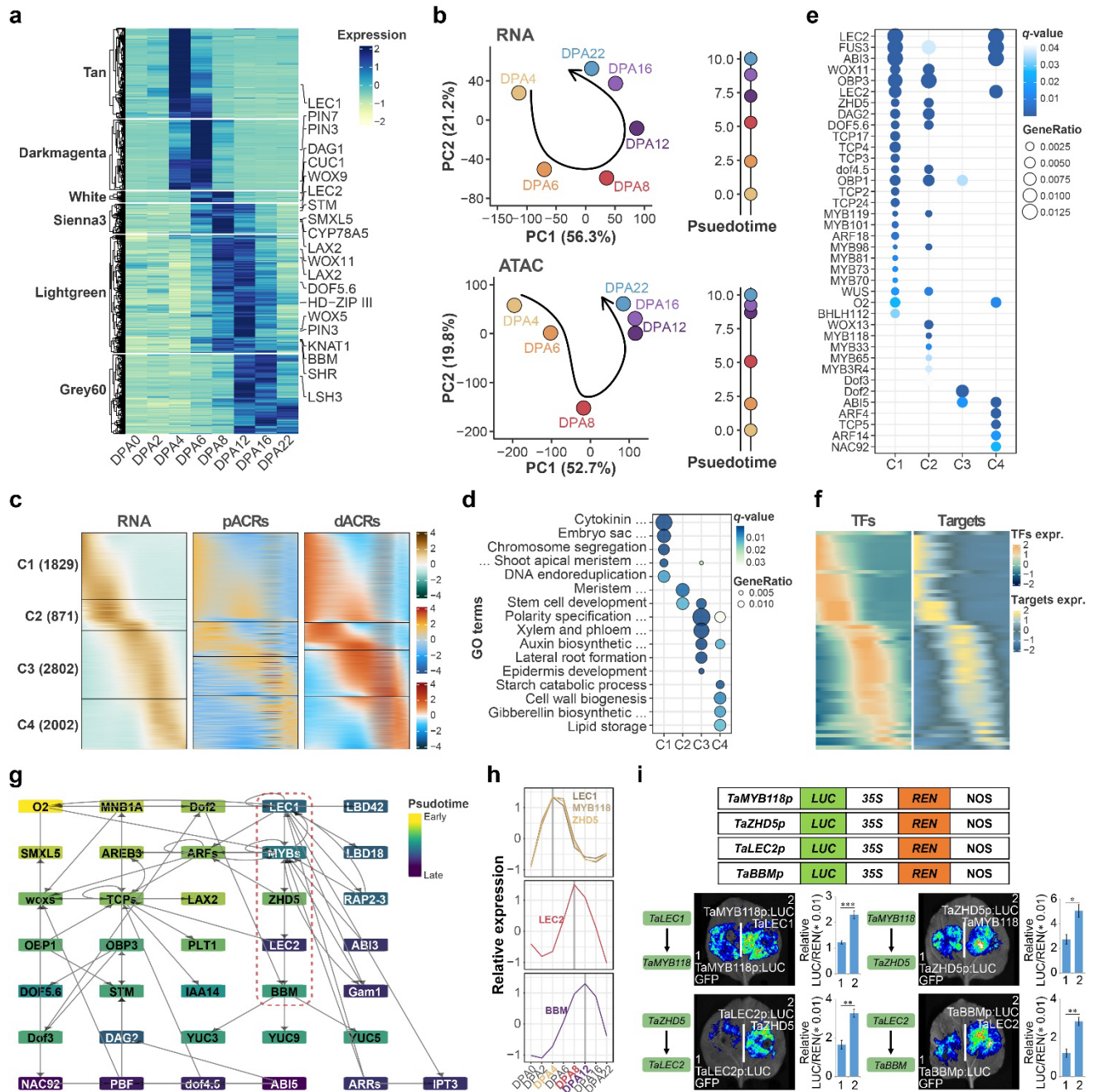
Fig. 6 | Transcription regulatory network governing embryonic patterning

a, Expression heatmap of selected key modules genes from Fig.1c. Genes with ortholog identified to function during embryogenesis in other species were indicated.

b, PCA trajectories of genes expression (top) from (a) and corresponding chromatin accessibility (bottom). Developmental time units (DTU) values on the right panel were calculated based on the straight distance between two adjacent points and then scaled to 0 to 10.

c, Sorted standardized temporal genes, pACRs, dACRs and expression profiles. Four clusters were generated based on expression order.

d,e, GO and TFs enrichment for four clusters of genes from (**c**).

f, Heatmap showing the synchronous patterns between TFs and target genes. Each row in heatmap represents links between one TF and corresponding genes.

g, GRNs for key TFs participated in embryonic patterning. Genes were roughly ranked by the expression order from top to bottom and from left to right. Colors indicate the expression timing wave during embryo patterning.

h, Expression pattern of representative key genes in the GRN network.

i, Luciferase reporter assays validation of transcriptional regulation among representative TF-target pairs. Student's t test was used for the statistic significance (* : $p <= 0.05$; ** : $p <= 0.01$; ***: $p <= 0.001$).
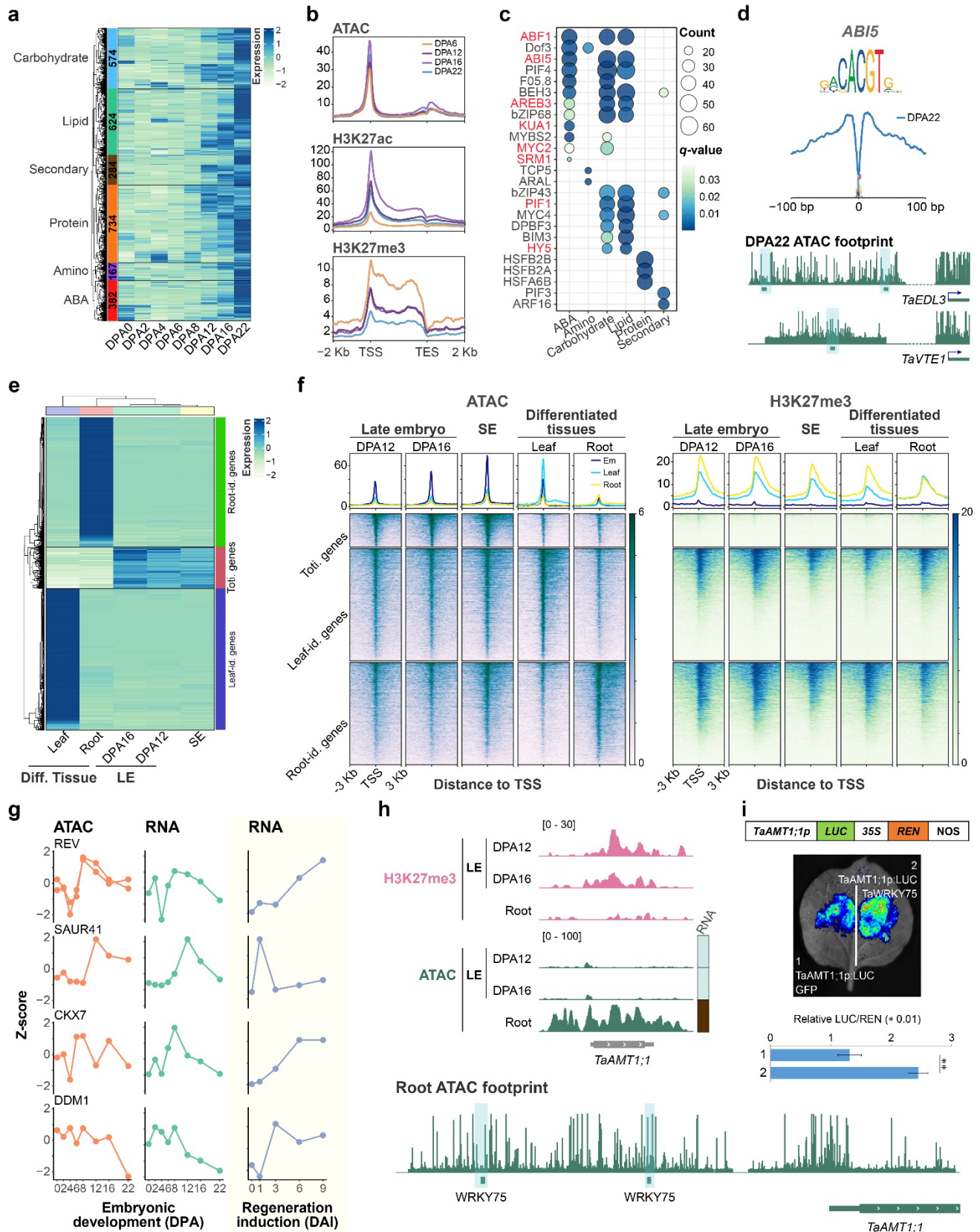
Fig. 7 | Maturation of embryogenesis in the context of chromatin regulation

a, **b**, Six pathway genes highly expressed during the late embryonic development (**a**), and corresponding epigenetic modifications (**b**).

c, **d**, Regulation of six pathway genes by specific cis motif and trans-factors. Binding TF enriched for each pathway genes (**c**). ABI5 footprint was highlighted (**d** top), and corresponding binding site at the promoters of carbohydrate-related gene *TaEDL3* and ABA-related gene *TaVTE1* (**d** bottom).

e, Identification of organ identity and totipotent genes based on expression specification across different tissues.

f, ATAC and H3K27me3 modification on the promoters of different genes sets in **e** cross five tissues, SE: somatic embryo.

g, Chromatin accessibility (left) and expression (mid) of regeneration genes during wheat embryogenesis, and their expression during callus induction (right).

h, H3K27me3 and chromatin accessibility regulated root-identity gene *TaAMT1;1* specific expression pattern in root and late embryo (LE). The ATAC footprint tracks showing the binding sites of TaWRKY75, a key TF predicted to bind *TaAMT1;1* as indicated in supplementary Fig. 8e.

i, Luciferase reporter assays show the transcriptional activation capability of TaWRKY75 to *TaAMT1;1* in *N. benthamiana* leaves. Student's t test was used for the statistic significance (* : $p <= 0.05$; ** : $p <= 0.01$; ***: $p <= 0.001$).

Fig. S8 | Inhibition of deep organogenesis in late embryo at chromatin level

a, Callus induction rate drops as embryo developing after DPA 12.

b, GO enrichment of tissue identity genes found in Fig. 7d.

c, Synchronous patterns between root- and leaf-identity genes expression and corresponding pACRs in individual tissue.

d, H3K27me3 modification levels on root- and leaf-specific genes in individual tissue.

e, Motifs enrichment within regulatory regions of root- (left) and leaf-specific genes (right). Percentage showed the presence frequency, while P-value indicates the enrichment.

f, Expression patterns of candidate TFs that can bind motifs found in **e**, cross different tissues and during embryogenesis.
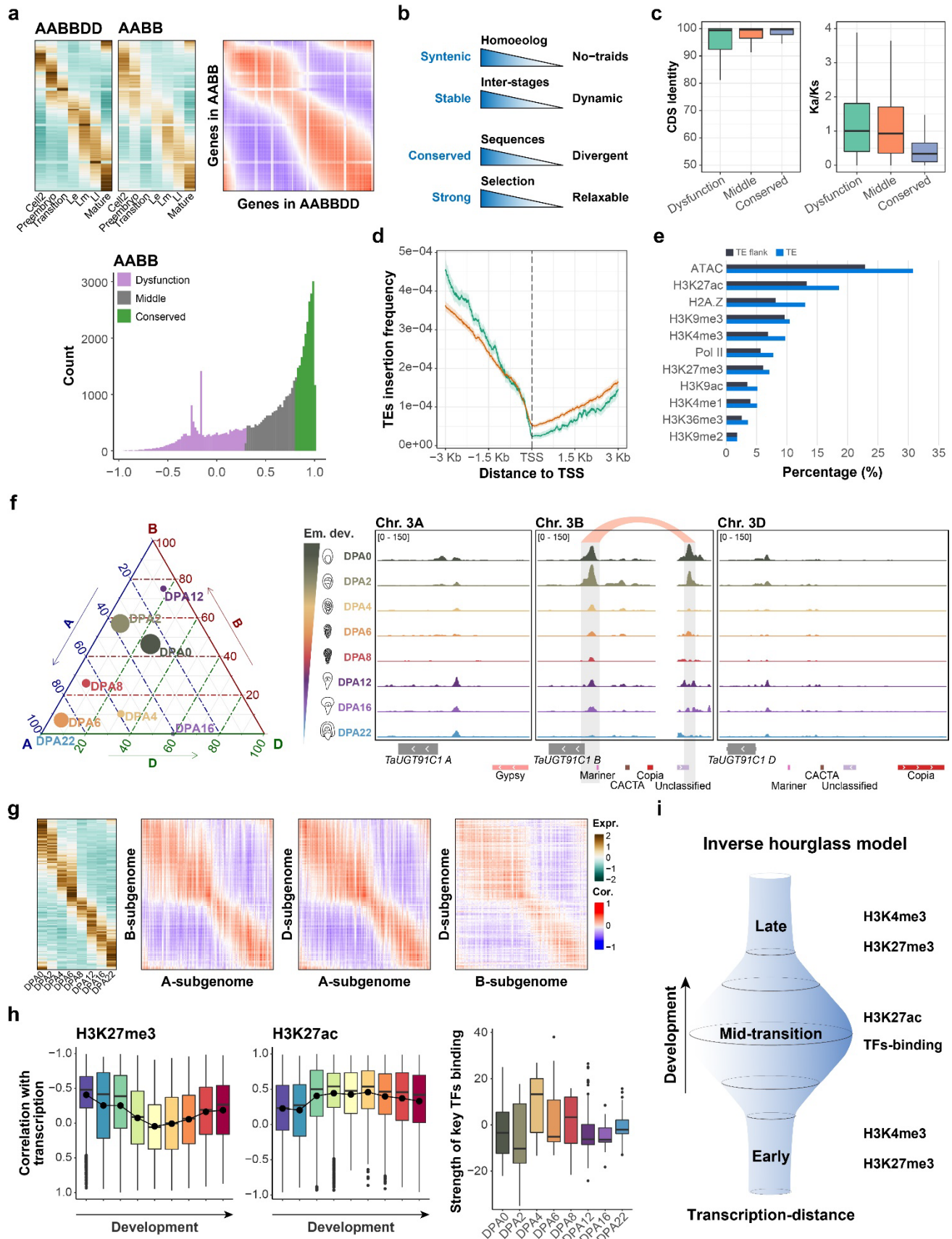
Fig. 8 | Transcriptional divergence between different ploidy wheat and epigenetic governed inverse hourglass model

a, Comparisons of gene expressions between hexaploid wheat (AABBDD) and ancestor (AABB). Each row represents an individual gene in hexaploid wheat and its counterpart ancestor (top left). Pearson correlation index was calculated between genes expression in hexaploid wheat and counterpart in ancestor (top right). Genes were clustered into three categories based on Pearson index (bottom).

b, Genomic features differences between dysfunction and conserved cluster genes generated in **a**. Statistics were generated from supplementary Fig. 9b-f.

c, Coding sequence similarity and Ka/Ks comparisons among different gene sets in **a**.

d, TEs insertion frequency in the promoters of different gene sets in **a**.

e, epigenetic modification enrichment at TEs insertion loci.

f, Homoeologs bias expression of representative gene TaUGT91C1 (left) and epigenetic regulations (right).

g, Comparisons of homoeologs expressions among subgenomes of hexaploid wheat. Each row represents an individual gene A-subgenome of hexaploid wheat, and the genes were ranked by activated order (left). Pearson correlation index was calculated between homoeologs expression in different subgenomes of hexaploid wheat as those in **a**.

h, H3K27me3 and H3K27ac contribution to genes expression across wheat embryogenesis (left) and key TFs binding strength at different developmental stages (right). The top 100 variable motifs calculated by chrVAR were used as key TFs.

i, The 'inverse hourglass model'. The transcription distance was large in the mid-stage. Histone modification H3K4me3 and H3K27me3 contributed to the early- and late-stage, and H3K27ac and TFs-binding contributed to the mid-stage.
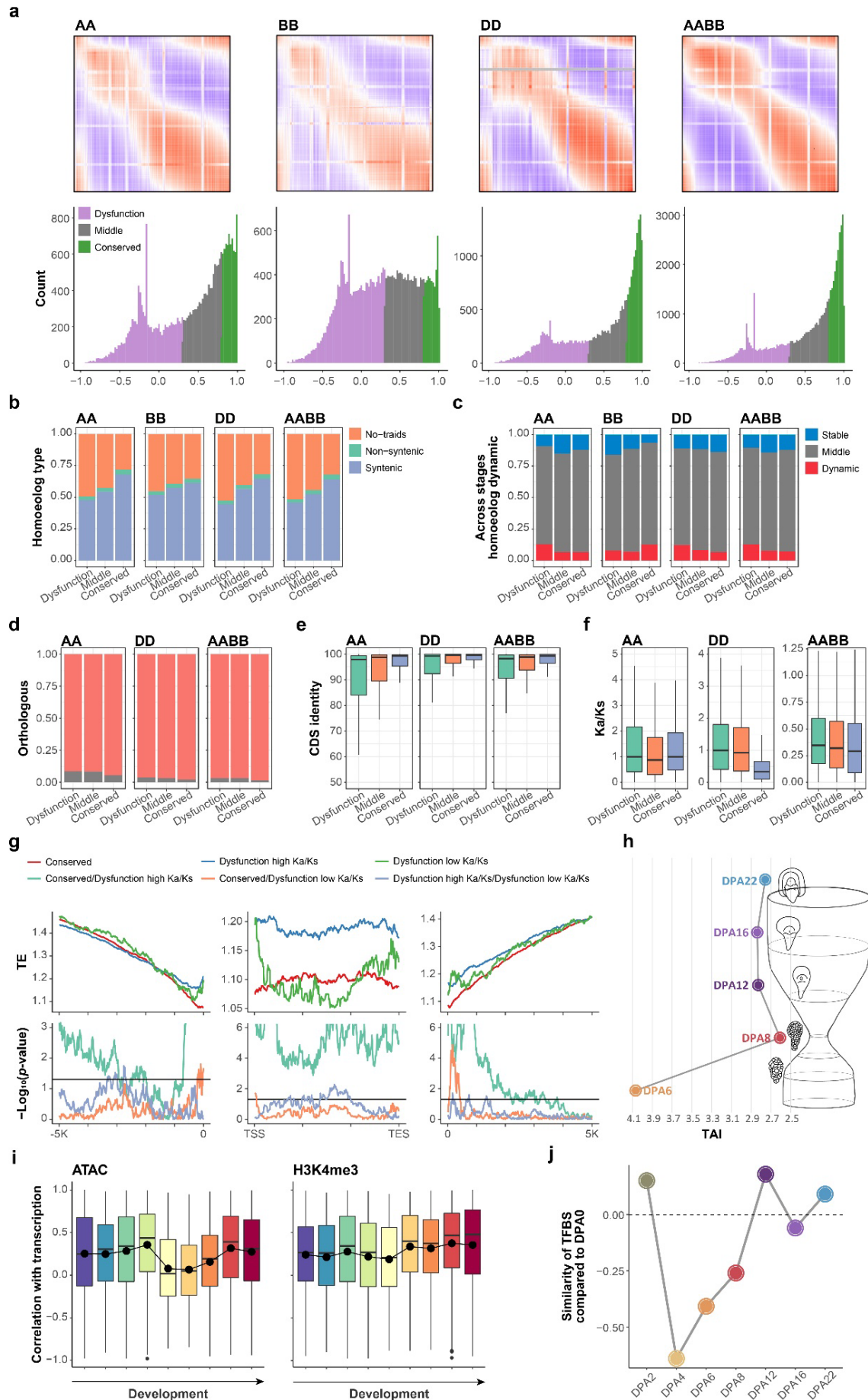
Fig. S9 | Comparisons of embryogenesis among different ploidy wheat and hourglass model

a, Comparisons of gene expressions between hexaploid wheat (AABBDD) and ancestors (AA, BB, DD and AABBDD) and corresponding Pearson correlation index. Each row represents an individual gene in hexaploid wheat and each line represents an individual gene in ancestor. Pearson correlation index was from -1 to 1, and indicated by colors from blue to red. Genes were clustered into three categories based on Pearson index from the comparisons between hexaploid wheat and ancestors, which are dysfunction, middle and conserved (bottom).

b, **c,** subgenome synteny (**b**) and triad dynamic expression across developmental stages (**c**) for gene sets defined in **a**.

d, Percentage of genes has paralogous in corresponding ancestors for gene sets defined in **a**..

e, Sequence similarity between hexaploid wheat and different ancestors for the corresponding gene sets defined in **a**.

f, Ka/Ks comparisons for different gene sets defined in **a**.

g, TEs insertion frequency at different positions relative to different gene sets (top) and comparisons between different gene sets (bottom). Mann-Whitney U test (two-sided) was used for the significant difference.

h, TAI index for expressed genes from DPA6 to DPA22, which represent an hourglass model.

i, ATAC but not H3K4me3 contributed to genes expression across wheat embryogenesis.

j, Similarity of TFs binding pattern compared with that in DPA0. All TFs binding patterns were calculated by chromVAR.
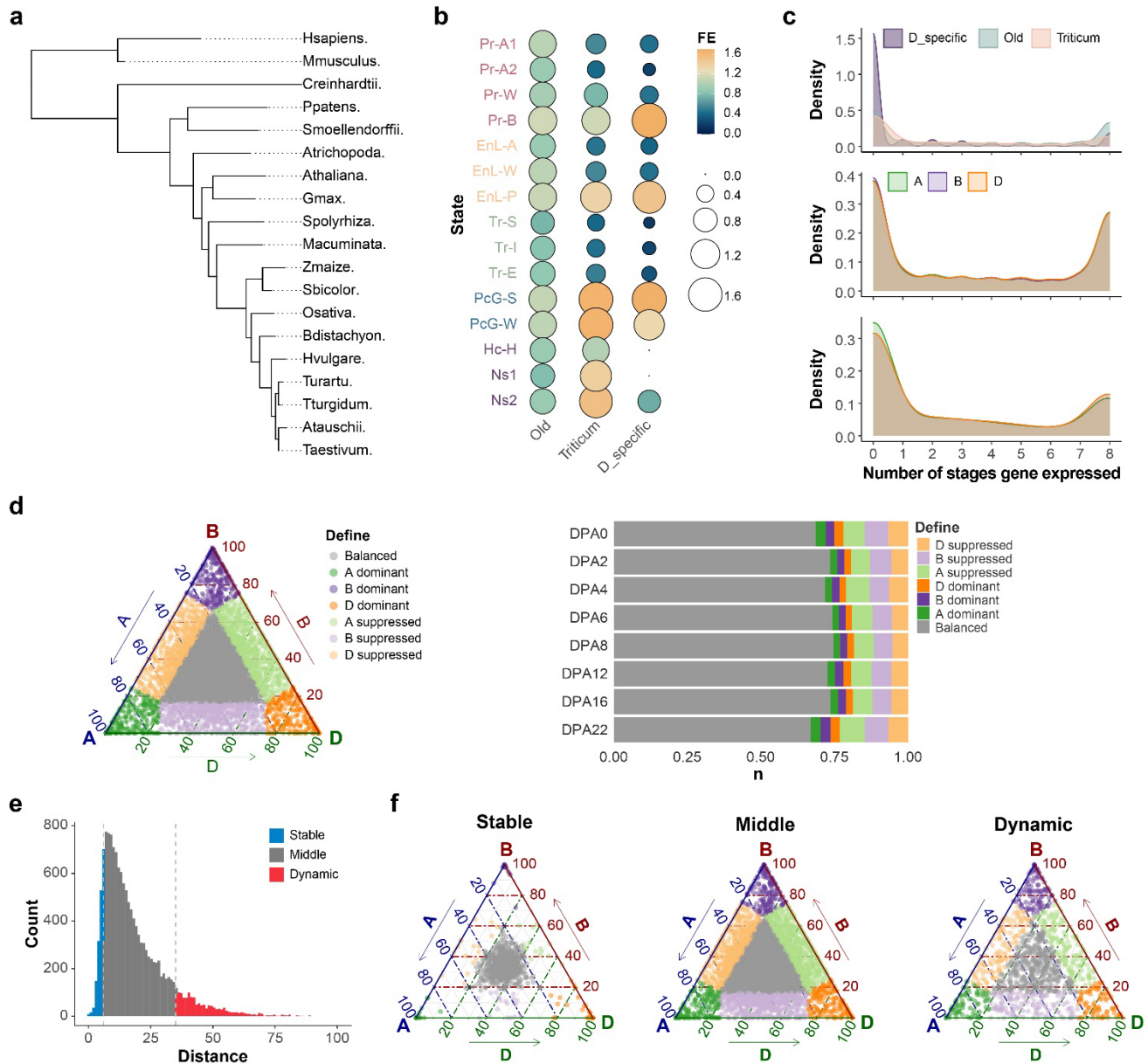
Fig. S10 | Chromatin landscape affected sub-genome bias expression

a, Phylostratigraphic map of various wheat species

b, Chromatin states enrichment for different evolutionary age genes in wheat embryo.

c, Expression spectrum of different evolutionary age genes in wheat embryo.

d, Define of homeologs bias expression (left) and percentage of different homeologs bias expression types across eight wheat embryonic developmental stages (right).

e, Homeologs dynamic expression across wheat embryogenesis. Euclidean distance method was used to analysis as (Ramírez-González et al., 2018).

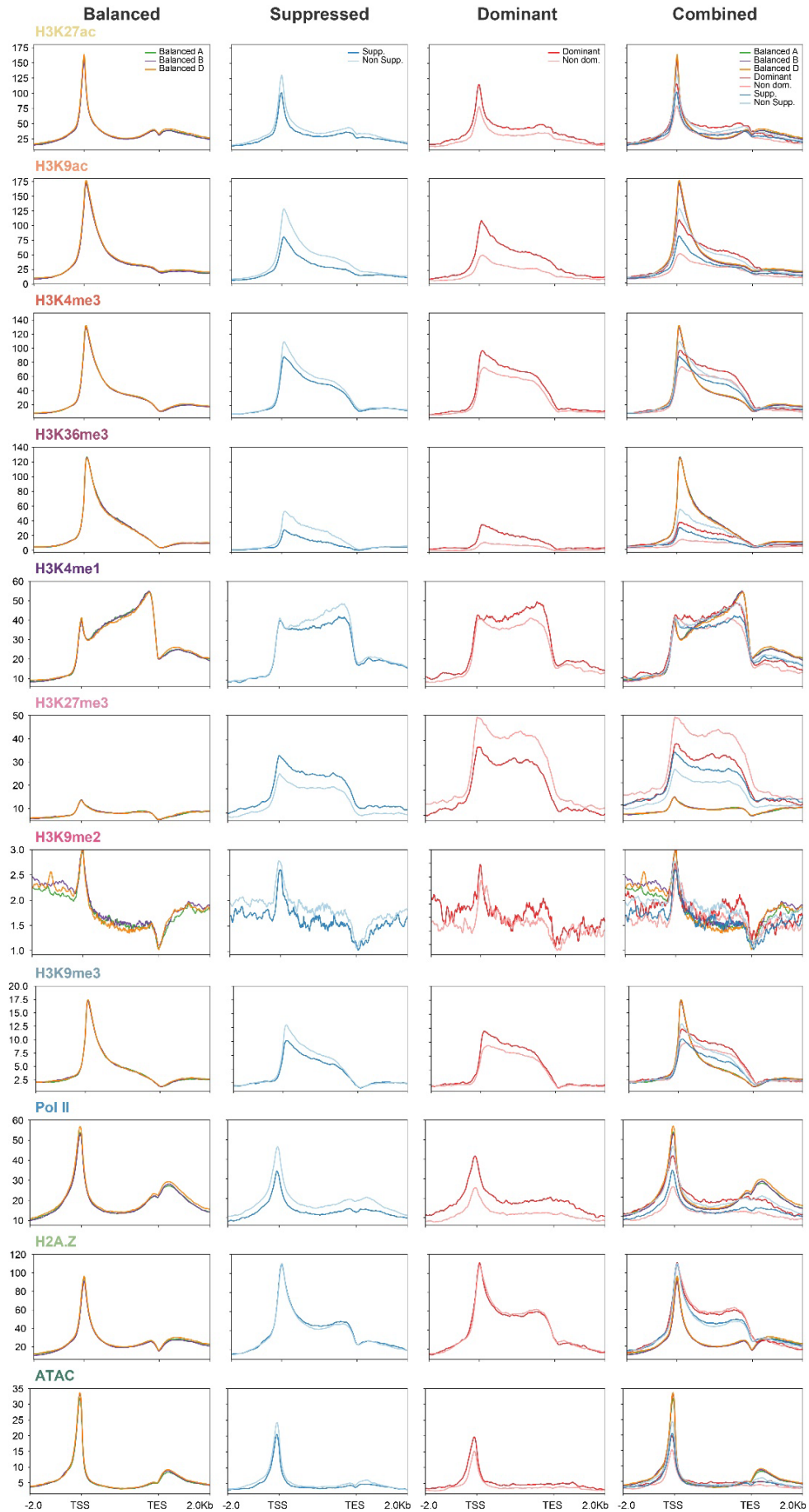**f**, Homeologs bias expression for gene sets defined in **e**.

Fig. S11 | Epigenetic modification on different types of homeologs bias expressed genes.
Balanced, suppressed, and dominant genes were shown in the first three columns, and a combination of different types of homeologs was shown in the fourth column. Epigenetic modification signals were normalized using RPKM, with 10 bp bin size.
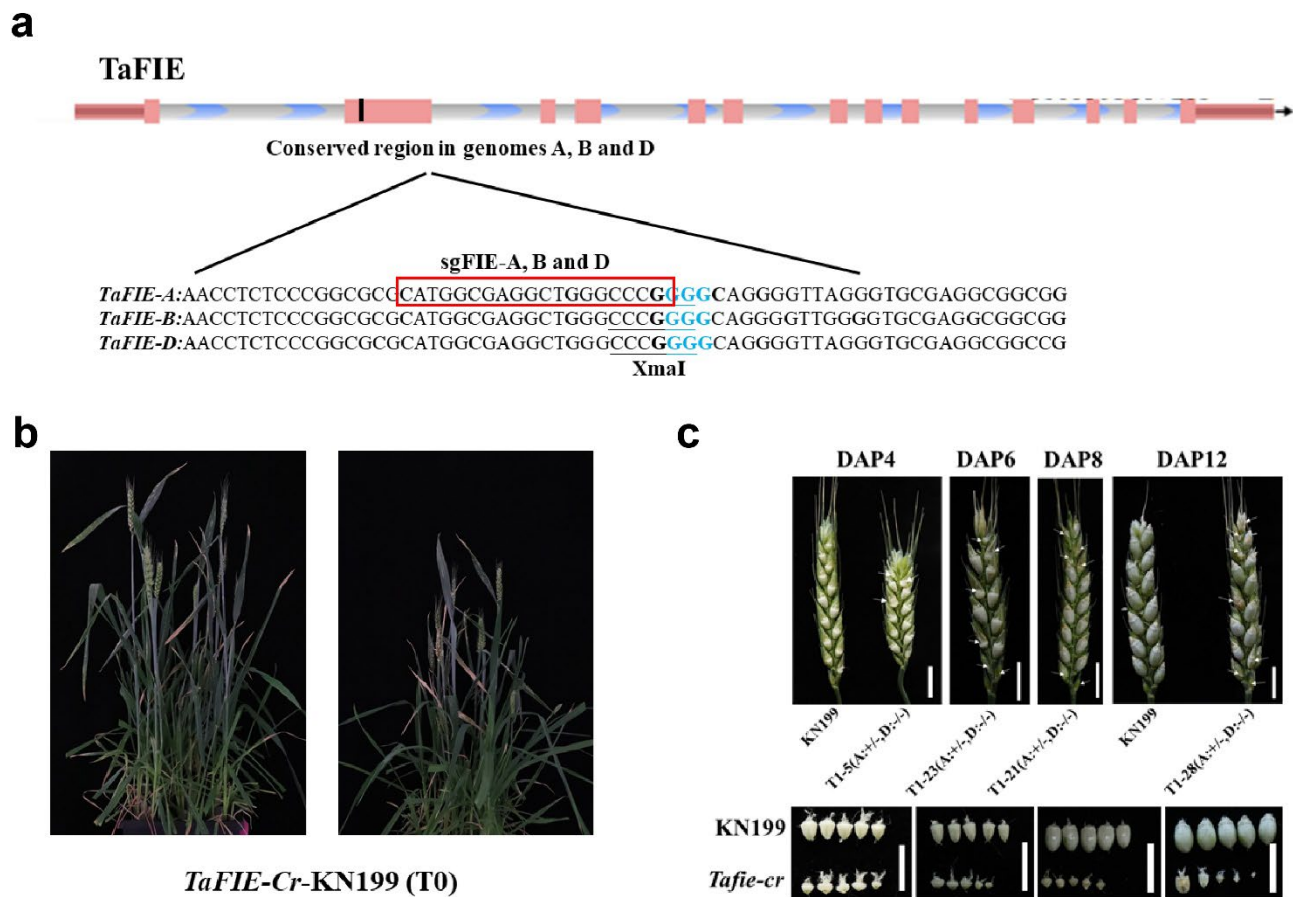
a

TaFIE

Conserved region in genomes A, B and D

sgFIE-A, B and D

TaFIE-A:AACCTCTCCCGGCGCC[CATGGCGAGGCTGGGCCCG]GGGCAGGGGTTAGGGTGCGAGGCGGCGG
TaFIE-B:AACCTCTCCCGGCGCGCATGGCGAGGCTGGGCCCGGGGCAGGGGTTGGGGTGCGAGGCGGCGG
TaFIE-D:AACCTCTCCCGGCGCGCATGGCGAGGCTGGGCCCGGGGCAGGGGTTAGGGTGCGAGGCGGCCG
XmaI

b

TaFIE-Cr-KN199 (T0)

c

DAP4 DAP6 DAP8 DAP12

KN199
Tafie-cr

Fig. S12 | Loss of core component of PRC2 led to aberrant seeds in wheat.
a, Genotyping of CRISPR plant at *TaFIE*.
b, Growth phenotype of *TaFIE-Cr-KN199* as compared to wildtype KN199.
c, Aberrant seed produced by *TaFIE-Cr-KN199*.