The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation

Hadzhiev, Yavor; Wheatley, Lucy; Cooper, Ledean; Ansaloni, Federico; Whalley, Celina; Chen, Zhelin; Finaurini, Sara; Gustincich, Stefano; Sanges, Remo; Burgess, Shawn; Beggs, Andrew; Müller, Ferenc

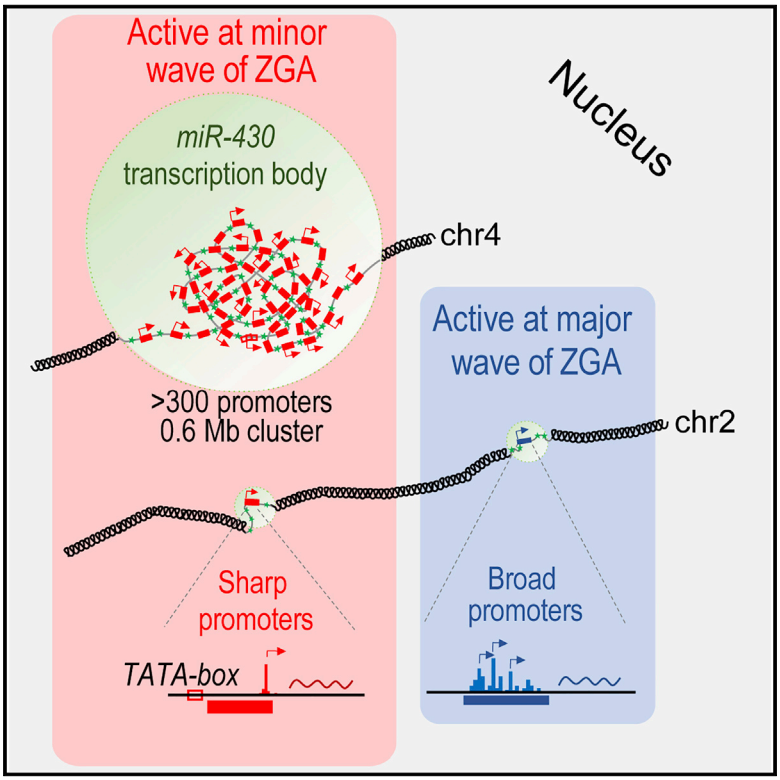
DOI:
10.1016/j.devcel.2022.12.007

Article

Developmental Cell

The *miR-430* locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation

Graphical abstract



Authors

Yavor Hadzhiev, Lucy Wheatley, Ledean Cooper, ..., Shawn Burgess, Andrew Beggs, Ferenc Müller

Correspondence

f.mueller@bham.ac.uk

In brief

Hadzhiev et al. study nuclear organization and promoter architectures of the first wave of genome activation in zebrafish. The first active genes are in a high promoter density gene cluster, forming a transcription body, and minor wave genes activated autonomously have distinct promoter features from the main wave.

Highlights

- A high promoter density gene cluster shapes a minor wave of ZGA in zebrafish
- Minor wave genes can be autonomously activated outside of the transcription body
- Minor wave promoter features are distinct from those of the major wave

Hadzhiev et al., 2023, Developmental Cell *58*, 155–170
January 23, 2023
https://doi.org/10.1016/j.devcel.2022.12.007

CellPress

Developmental Cell

CellPress
OPEN ACCESS

Article

The *miR-430* locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation

Yavor Hadzhiev,[1,6] Lucy Wheatley,[1,6] Ledean Cooper,[1] Federico Ansaloni,[1,2,3] Celina Whalley,[1] Zhelin Chen,[4,5] Sara Finaurini,[2] Stefano Gustincich,[3] Remo Sanges,[2,3] Shawn Burgess,[5] Andrew Beggs,[1] and Ferenc Müller[1,7,*]

[1]Institute of Cancer and Genomics Sciences, Birmingham Centre for Genome Biology, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK

[2]Area of Neuroscience, Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy

[3]Central RNA Laboratory, Istituto Italiano di Tecnologia (IIT), 16163 Genoa, Italy

[4]South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China

[5]Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-2152, USA

[6]These authors contributed equally

[7]Lead contact

*Correspondence: f.mueller@bham.ac.uk

https://doi.org/10.1016/j.devcel.2022.12.007

SUMMARY

In anamniote embryos, the major wave of zygotic genome activation starts during the mid-blastula transition. However, some genes escape global genome repression, are activated substantially earlier, and contribute to the minor wave of genome activation. The mechanisms underlying the minor wave of genome activation are little understood. We explored the genomic organization and *cis*-regulatory mechanisms of a transcription body, in which the minor wave of genome activation is first detected in zebrafish. We identified the *miR-430* cluster as having excessive copy number and the highest density of Pol-II-transcribed promoters in the genome, and this is required for forming the transcription body. However, this transcription body is not essential for, nor does it encompasse, minor wave transcription globally. Instead, distinct minor-wave-specific promoter architecture suggests that promoter-autonomous mechanisms regulate the minor wave of genome activation. The minor-wave-specific features also suggest distinct transcription initiation mechanisms between the minor and major waves of genome activation.

INTRODUCTION

The major wave of zygotic genome activation (ZGA) is characterized by activation of thousands of genes when dividing cells reach a threshold nucleo-cytoplasmic ratio (reviewed in Lee et al.,[1] Schulz et al.,[2] and Vastenhouw et al.[3]). At the major wave transcription machinery is suggested to first outcompete maternally deposited repressive factors, including an excess of histones, which get diluted by exponential cell divisions (reviewed in Lee et al.,[1] Vastenhouw et al.,[3] Palfy et al.,[4] Wragg and Müller,[5] Amedeo et al.,[6] and Joseph et al.[7]). In addition, several molecular mechanisms have been suggested to regulate the timing of genome activation: (1) availability of pioneering transcription factors,[8] (2) availability of transcription initiation machinery,[9–11] (3) sequential increase in chromatin accessibility,[12,13] (4) gradual formation of transcriptionally competent chromatin,[14,15] and (5) cell-cycle elongation.[15–18]

However, many genes escape global repression and become activated in advance of the rest of the genome, during what is called the minor wave of genome activation.[19–21] In zebrafish, the minor wave of genome activation starts at the 64-cell stage, reaching high levels of activity by the 512-cell stage interphase, and characterized by high production of *miR-430* primary transcripts.[15,20,22] Notably, in zebrafish *miR-430* is a key regulator of maternal mRNA clearance,[21] and during ZGA, *miR-430* nascent RNA (primary microRNA) accumulates in a transcription body. During the minor wave of genome activation this body carries most of the cell's detectable, transcriptionally active RNA Pol II (RNA Pol II Ser2P) and nascent RNAs,[15,22,23] and the high transcriptional activity of this body has been associated with local chromatin depletion.[23] Additionally, the genomic locus of *miR-430* genes is marked by CCCTC-binding factor (CTCF) and Rad21,[24] suggesting chromosome conformation mechanisms involved in its activation. *MiR-430* activation is dependent on BRD4 and histone H3K27ac marking of its chromatin[15,25] and on the pioneer transcription factors (also called stem cell factors): Nanog, Pou5f3, Sox19b.[8,15,25–27] These stem cell factors have been shown to activate the major wave of ZGA in

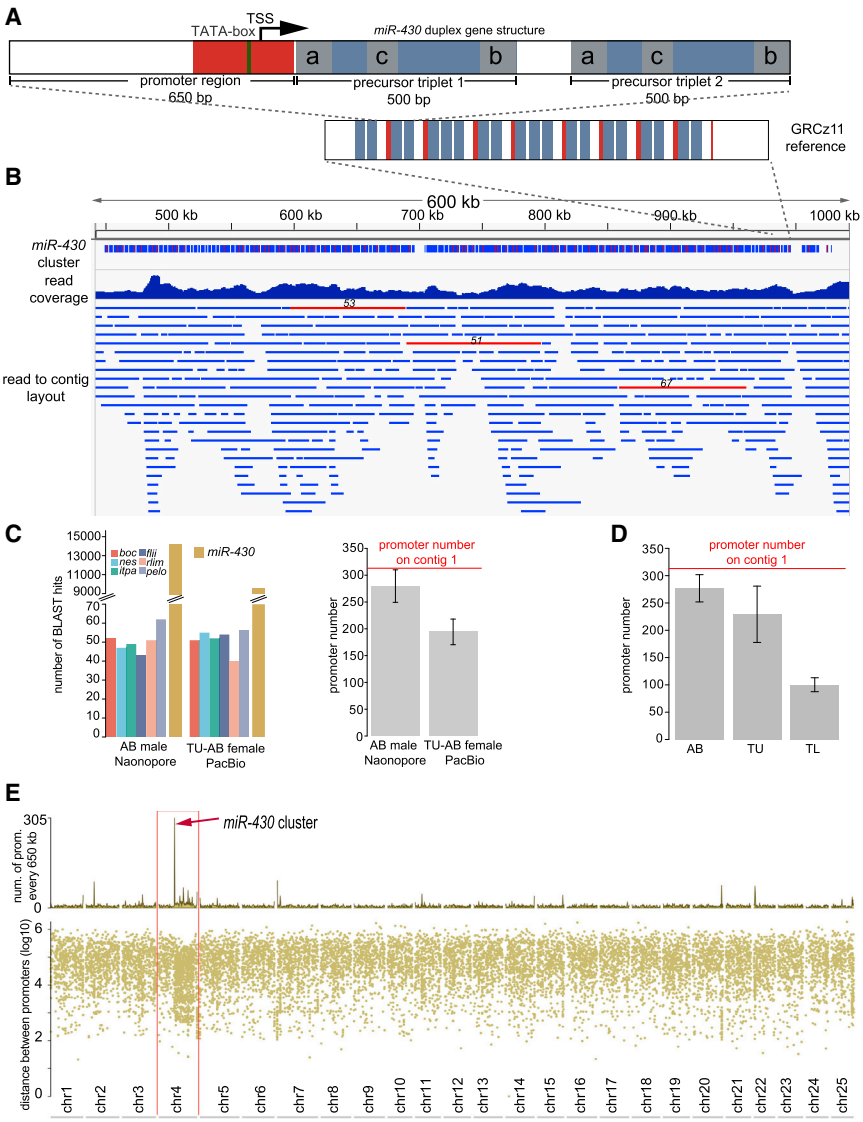


Figure 1. *De novo* assembly of the *miR-430* cluster from long-read sequencing data

(A) Schematic of the most common *miR-430* gene structure (top), the duplex consisting of promoter region (red) and two pre-miRNA triplets (blue) each containing *miR-430*a, *miR-430*b, and *miR-430*c (gray) pre-*miR-430* subtypes. Bottom: the *miR-430* gene cluster structure in current reference assembly (GRCz11).

(B) Genome browser view of the *miR-430* gene cluster assembled from long reads. The tracks show a low-resolution continuity structure of the gene cluster (top), long-read coverage histogram (middle), and the read to contig assembly layout (bottom). Raw reads with 50 or more promoters are shown as red horizontal bars with the promoter number indicated on top of them.

(C) Left: number of BLAST hits in raw long reads of single gene promoters and *miR-430*. Right: estimated *miR-430* promoter number normalized to that of single-copy genes.

(D) Estimated *miR-430* promoter number (red line) by digital droplet PCR (ddPCR) for three zebrafish strains and the assembled contig. Error bars are standard deviation from the mean of the *miR-430* promoter estimate normalized to each of the 6 single-copy genes and 6 technical ddPCR replicates respectively.

(E) Rainfall plot representing the promoter density across each chromosome of the newly assembled zebrafish genome (GRCz11 reference guided). Red arrow indicates the *mirR-430* gene cluster. Abbreviations: TSS, transcription start site; AB, TU-AB, TU, and TL are zebrafish strains used in (C) and (D).

See also Figure S1 and Tables S1 and S5.

fish[8,13,27,28] and homologs Pou5f3 and Sox3 in *Xenopus*,[29] while their role in minor wave genome activation is not yet understood.

Deposition of histone modifications (H3K4me1, H3K4me3, and histone variants [H2AZ]) associated with gene regulation, precede genome activation,[14] and are detected in gametes.[30] These epigenetic dynamics suggest potential inheritance from parents with a possibly instructive role in both minor and main waves of genome activation.[31–35]

Taken together, the genetic and epigenetic mechanisms underlying the activation of *miR-430* locus may explain the formation of this transcription body and provide insights into the mechanisms of the minor wave of genome activation.

In this study, we have examined the molecular mechanisms underlying the activation of minor wave genes using the *miR-430* locus. We describe a repetitive structure for the *miR-430* cluster and demonstrate that transcription of the *miR-430* locus is required for transcription body formation. In addition, we show that minor-wave-specific core promoters carry distinct architecture from that used by the major wave of genome activation. While promoter density enables the formation of a transcription body, single-copy genes with promoters, which share features with those in the transcription body, can escape global transcriptional repression and become activated in the minor wave independently from the transcription body.

RESULTS

Long-read-sequencing-mediated genome assembly reveals the highest promoter density gene cluster in the zebrafish genome

In the current genome assembly (GRCz11) the *miR-430* locus is a multicopy gene cluster, which consist of 8 complete *miR-430* gene units, each containing a 650-bp promoter region and triplets of precursor miRNAs, which themselves are repeated in 2 (duplex) or 3 times (triplex) per promoter[22] (Figure 1A). Each precursor miRNA triplet consists of 4 different subtypes of precursor genes: a, b, c, or the "a" variant, called "i."[22] In the reference genome assembly (GRCz11), the 5′ end of the locus is truncated, missing the promoter sequence, which suggests that this section of the assembly is incomplete, probably due to the difficulty of assembling the highly repetitive structure (Figure 1A).

Therefore, we embarked to reassemble the *miR-430* locus, generating Oxford nanopore long-read sequencing data from AB male zebrafish at ~44× coverage and complemented this with Pacific Biosciences (PacBio) sequencing from a heat-shock diploid "double heterozygous"[36] Tuebingen-AB hybrid female zebrafish at ~78× coverage.

Initial analyses of the long-read data revealed 7 reads, which contained 50 or more *miR-430* promoters (Figures 1B and S1A), with 74 promoters on a single 111-kb read with 128 triplets. The longest read spanning the *miR-430* cluster was 122 kb with 67 promoters (149 triplets). This analyses confirmed the previously described sub-structure within precursor triplets and revealed variable number of triplets per promoter (Figures 1A, 1B, and S1B). Assembly of the Oxford nanopore reads by the Canu assembler[37] resulted in 2 contigs containing the *miR-430* gene cluster, likely representing allelic variants. The first contig was ~3.44 Mb long with 313 *miR-430* promoters (655 triplets) spanning 577 kb (Figure 1B; Data S1). The second contig was 811 kb long containing 306 *miR-430* promoters (587 triplets) spanning 534 kb (Figure S1A; Data S1). A comparison of the read to contig layouts (Figures 1B and S1A), showed that contig1 had overall better read support/coverage and was used in further analyses.

The striking increase of copy number from 8 to >300 repeated gene units prompted us to seek independent validation of the copy number of the *miR-430* primary transcript genes. To this end, we estimated read coverage and compared with known single-copy genes on the genome. The raw reads of our Nanopore and PacBio sequencing were queried with the *miR-430* promoter sequence (Figure 1B) and the promoter sequences of 6 single-copy genes (Table S1) using basic local alignment search tool (BLAST).[38,39] An estimate of promoter number of *miR-430* was generated by normalizing the BLAST hit number to that of the single-copy genes. This analysis led to a calculated copy number of 280 ± 30 and 194 ± 24 for each sequencing run, respectively (Figure 1C). In addition, we estimated the *miR-430* promoter number by digital droplet PCR (ddPCR) using *shha* as single-copy reference gene for 3 zebrafish stains (AB, TU, and TL). The copy number estimated by this approach was 277 ± 25, 229 ± 52, and 100 ± 13 for AB, TU, and TL strains, respectively (Figure 1D). The assembly of the PacBio reads did not result in continuous, end-to-end assembly of the *miR-430* locus but was split between 14 contigs (Figure S1B). The total number of *miR-430* promoters and triplet structures on all contigs from the PacBio data was estimated to 359 and 768, respectively.

While the copy-number figures from the above approaches differ, they are in a comparable range to that obtained by copy-number calculation on the assembled contigs (Figures 1C and 1D). Thus, the number of promoters in the *miR-430* locus is almost 2 orders of magnitude higher than what was annotated on the reference genome. We explored how this promoter density, repeated every 1.8 kb on average over a ~0.6 Mb region, compares with the rest of the genome. To analyze promoter density, we performed GRCz11 reference-guided chromosome scaffolding of a *de novo* assembly from the nanopore sequencing run (see STAR Methods). Cap analysis gene expression (CAGE-seq) data set, [40] multi-mapped to the newly scaffolded genome was used to identify promoters and calculate their density. As demonstrated on Figure 1E, the *miR-430* cluster indeed showed by far the highest promoter density in the assembled genome, more than 4 times higher than other promoter-dense regions.

Taken together, the much higher promoter copies and density of the *miR-430* gene locus than previously anticipated, provide a potential explanation for the timing and the scale of gene expression during the minor wave of genome activation, which appears in a large transcription body.

Distinct epigenetic regulation of the *miR-430* cluster and minor wave genes of genome activation

The occurrence of minor wave gene activation raises the question of what chromatin and DNA sequence features are contributing to its formation. We first aimed to characterize the epigenetic profiles and promoter architecture features of the *miR-430* cluster and other minor wave genes, in contrast with those activated at distinct phases of development. Publicly available epigenomic datasets representing chromatin opening (Assay for Transposase-Accessible Chromatin, ATAC-seq) and associated with *cis*-regulatory element regulation (H3K4me3-, H3K27me3-, and H3K27ac-ChIP-seq),[31,33,35,41] covering key early developmental stages were selected (see STAR Methods).

The datasets were mapped to a custom genome, created by adding the *miR-430* cluster containing contig1 sequence from the long-read assembly to GRCz11. To investigate the role of epigenomic features in early genome activation, their signal levels and dynamics during blastula stages were compared among 5 distinct gene sets (25 genes each) selected based on the timing of their onset of activation (Figure 2): (1) *miR-430* genes: the earliest and most highly transcribed gene cluster during the minor wave. (2) Minor wave genes: detected as activated during the first wave of the ZGA.[20] Genes were selected with no maternal contribution and >2.5-fold change between 128 and 512 cells. (3) Major wave genes: randomly selected genes activated at the major wave but not detected during the minor wave.[20] (4) Constitutive genes (CGs): housekeeping genes with maternal contribution and expressed throughout development. These genes were previously shown to use two different promoter codes present on the same core promoter sequence.[42] First, a TATA-like motif called W-box was utilized in the oocyte, then later replaced by a nucleosome positioning signal of AT/GC enrichment boundary downstream of the transcription initiation start site (TSS) at ZGA. (5) Post-gastrulation genes (PGs): activated after gastrulation during early segmentation stages. Expression dynamics for all gene sets were determined based on[40] and mapped to GRCz11/danRer11, ENSEMBL v95 gene annotations (Figures S2A–S2D; Table S2).

Next, the chromatin openness at the selected gene sets was examined by ATAC-seq (Figures 2A and 2B). The distribution profile of the signal was as expected for all datasets, peaking 100–150 bp upstream of the TSS in the nucleosome free region of all active promoters. Since *miR-430* signals were plotted from multi-mapped data, they were not quantitatively comparable to the rest of the gene sets and were used to explore and compare temporal trends. The signals on the *miR-430*, as well as on the early and late zygotic gene sets, correlated with their gene expression dynamics. *miR-430* was highly active early and shows the most open chromatin, whereas the minor wave genes become similarly open at earlier stage and were generally more open than the major wave genes. The latter genes were comparable

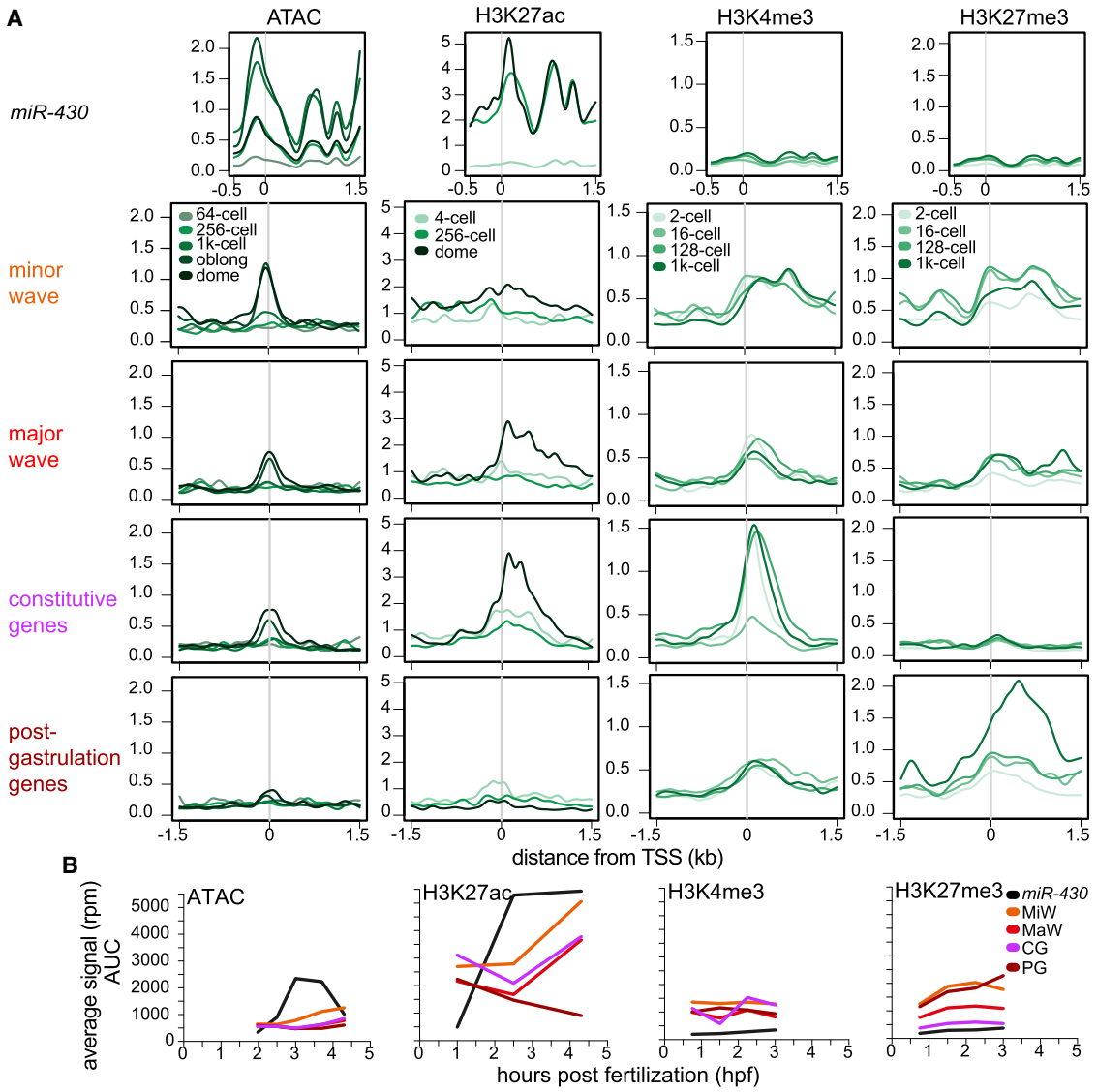


Figure 2. Epigenomic features of *miR-430* cluster compared with other gene sets activated at distinct phases of embryo development

(A) Aggregation plots showing signal distribution (mean) around the TSS for different gene sets (left side labels) and epigenetic features (top labels).

(B) Line graphs of total epigenetic signal. The *miR-430* data were multi-mapped (black) and not directly comparable to the other presented gene sets, which are unique mapped (color).

See also Figure S2 and Table S2.

to the CG set, in line with their zygotic activation at and post mid-blastula transition (MBT).

Next, we compared chromatin states with the appearances of promoter-associated histone modification marks. The levels of the gene activity-associated H3K27ac[43–45] correlated with that of the open chromatin states (Figures 2A and 2B). The *miR-430* gene clusters showed high signal already at 256-cell stage (Figures 2A and 2B). The minor wave gene set showed higher and broader H3K27ac signal distribution than the major wave genes during 256-cell stage, which in contrast, were more comparable to the CGs. Similar to the ATAC-seq profile, the lowest levels were observed in the post-gastrulation set.

Despite the high degree of open chromatin and enrichment in H3K27ac corresponding to the high expression levels at early blastula stages, the *miR-430* cluster showed low levels of the active or poised promoter mark H3K4me3 in comparison with other marks.[14,42,46,47] Minor wave genes showed presence of H3K4me3 extending to the gene body and correlating with the presence and extent of H3K27me3 on this gene set. The highest levels H3K4me3 were observed in the constitutive and minor wave gene sets, with the latter also having a broader distribution of signal (Figures 2A and 2B) in line with their activity states. The major wave genes showed lower signal at 256-cell stage, in line with their later expression initiation at sphere/dome stages. Although not expressed at early blastula and only becoming active at significantly later (early segmentation) stages, the PGs showed relatively high H3K4me3 signal, comparable to the major wave genes (Figures 2A and 2B).

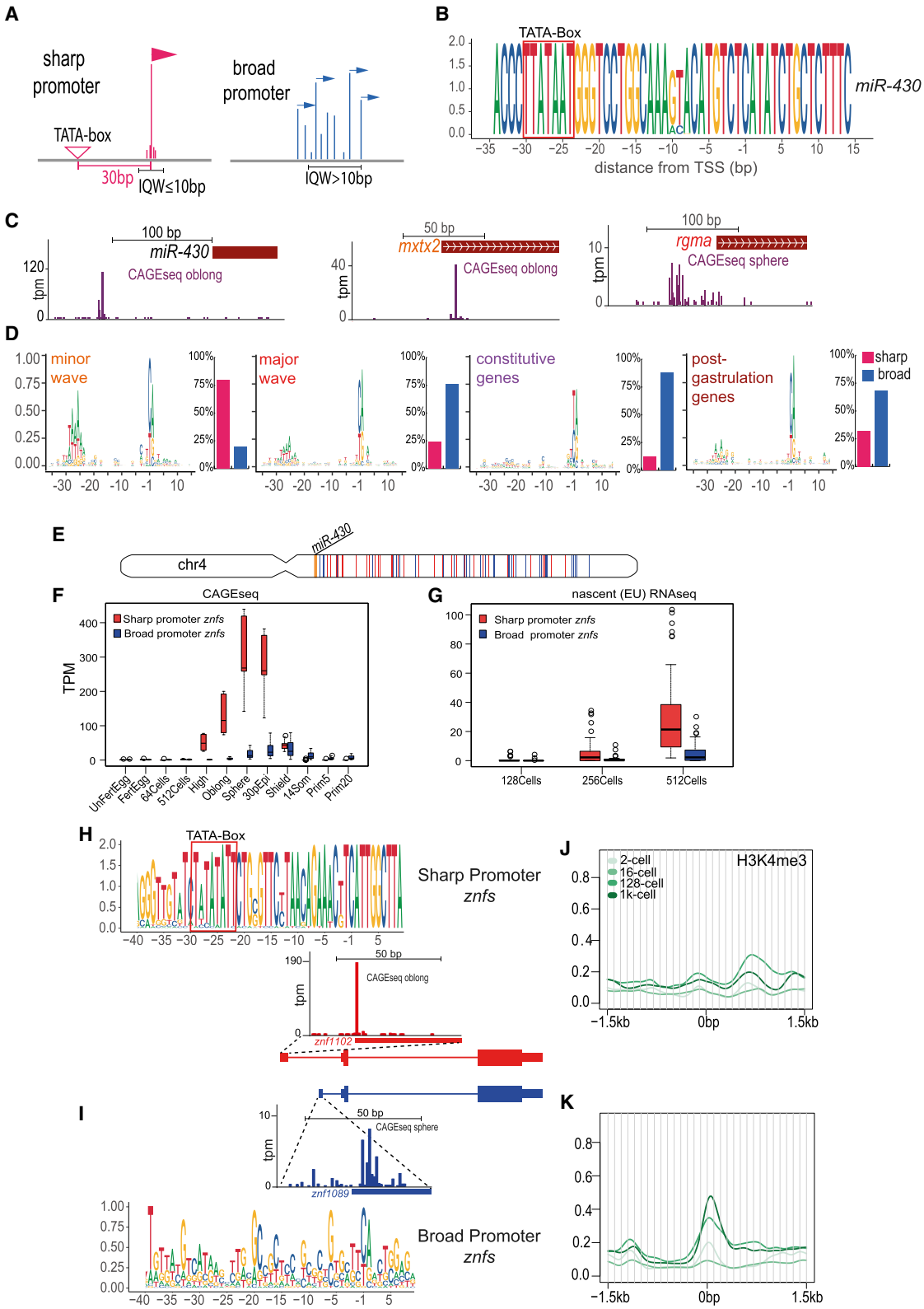


Figure 3. Promoter architecture features of minor and main wave ZGA genes

(A) Schematic comparing sharp and broad promoter architectures.

(B) Sequence logo of 80 *miR-430* genes showing the −35 to +15 region relative to the dominant TSS.

(legend continued on next page)

The contrasting dynamic of the H3K4me3 signal was also observed between stages, relatively high levels were observed at the two cell with a significant drop at 16 cell, most prominent in the constitutive and major wave gene sets. This dynamic indicates potential parental inheritance or early embryonic premarking, which is followed by subsequent erasure and re-deposition at later stages.[35] However, this ''reprogramming'' was not observed in the minor wave and PGs. In summary, minor wave genes show broadly similar epigenetic profiles to the major wave genes, but with higher H3K27me3 levels. However, there is a notable difference between *miR-430* and other minor wave genes in the apparent lack of H3K4me3 at this locus.

Core promoter features distinguish minor and major wave genome activation

Previously, it was shown that ZGA in *Drosophila* is characterized by enrichment for the TATA box,[48] elongated transcriptional activity states in early development[49] and often associated with sharp transcription initiation profiles (Figure 3A). Notably, *miR-430* promoters carry a canonical TATA box (Figure 3B) and sharp TSS profile (Figure 3C). In further pursuing features of minor wave genes, which may explain their distinct activation profiles, we analyzed core promoter architecture of minor wave genes in relation to the 3 other groups. Sequence analysis of the 4 gene groups indicated enrichment for TATA box at the canonical distance from the main TSS (Figure 3D) and corresponding sharp TSS in minor wave genes (Figure 3C). These promoter features related to but were not identical with the maternal transcription initiation code[42] (see discussion). However, the minor-wave-associated promoter profile was distinct from the major wave genes, characterized by lack of TATA box (Figure 3D) and presence of broad TSS profile (Figure 3C). This result together suggests that *miR-430* promoters carry shared and distinctive features with minor wave genes such as the presence of TSS-determining TATA box or TATA-like sequence signals (W-box) and sharp TSSs.

To further analyze the potential contribution of core promoter features to the timing of genome activation we have exploited a family of zinc-finger genes (*znfs*) with distinguishable promoter features and differential expression timing during ZGA. The *miR-430* cluster sits in a genomic environment, flanking a gene-poor region on the long arm of chr4, which shows distinct transcriptional dynamics from the rest of the genome.[50] The long arm of chr4, despite being relatively gene poor contains many C2H2-type *znfs*, some showing early zygotic transcription.[22,50] We investigated the transcriptional dynamics of these *znfs* using CAGE-seq and nascent RNA-seq revealing different temporal dynamic relating to distinct promoter architectures of these *znfs*. A subset of *znf*s with sharp, TATA-box promoters were found to be expressed at the minor wave, whereas broad promoter *znfs*, lacking a TATA box were active at the major wave (Figures 3E–3I; Table S3). This sharp promoter structure of the early expressed *znfs* was shared with *miR-430* promoters (Figure 3B). The broad promoter of the main wave *znf*s was in line with the predominantly broad promoter-containing main wave genes (Figures 3C and 3D). Notably, H3K4me3 signals were mostly lacking on the sharp promoter *znf*s expressed at the minor wave (Figure 3J), and thus resemble the *miR-430* promoters, in contrast to the broad promoter *znfs* expressed during main wave, which carry H3K4me3 (Figure 3K).

The proximal promoter sequences of the *miR-430* gene cluster autonomously activates early transcription in a single copy on heterologous chromosomes

To identify the promoter sequence features that may inform on the regulation of *miR-430* expression we focused on regulators of the major wave ZGA.[8,25,27] Publicly available ChIP-seq datasets[27,51] were mapped on the *de novo*-assembled *miR-430* contig and signal distribution analyzed. The majority of the Nanog signal was distributed over the promoter region, upstream of the TSS (Figure 3A), while Pou5f3 and Sox19b showed broader distribution, still focused on the promoter region. The regions around the *miR-430* cluster showed significantly less signal suggesting that these factors mainly bind proximally, with no indication of distal enhancers (Figures 4A and S3A). *In silico* ClusterBuster transcription-factor-binding site prediction[52] resulted in the *miR-430* region being identified as one single cluster of binding sites (Figures 4A and S3A). The binding sites for Nanog and Pou5f1::Sox2 were predominantly located in vicinity of the TSSs, with a Nanog motif at the peak of Nanog ChIP-seq signal (Figure 4A) and a high scoring Pou5f1::Sox2 site ~300 bp upstream of the TSS (Figure 4A).

The multicopy nature of the *miR-430* locus may provide high density of *cis*-regulatory elements to counter the repressive effect of maternally deposited negative regulators of transcription, such as histones.[6,7] To test whether the multicopy state is required to drive activity at the minor wave of genome activation, we integrated a single copy of the *miR-430* promoter with a reporter gene into heterologous genomic locations. A 650-bp proximal promoter fragment, containing the TATA box plus Nanog and Pou5f1::Sox2-binding sites (Figure 4B; Data S1), was used to generate reporter transgenic lines using the PhiC31 integrase system[53,54] (Figure 4B; STAR Methods). Two lines were generated, integrated on chr17 and chr22, respectively. Expression driven by the *miR-430* promoter was detected in both lines at 256/512-cell and sphere stages by RT-PCR (Figure 4B), which together with the lack of detectable expression at earlier or later stages, was in line with the expression dynamics of the *miR-430*. In contrast, reporter expression from the *γ-crystalline* promoter active in the lens,[55] was detected only at 96 hpf as

(C) Genome browser views showing promoter shape as defined by CAGE-seq signal for *miR-430*, and example genes for the minor (*mxtx2*) and the major (*rgma*) wave gene sets.
(D) Promoter architecture comparison between the 4 gene sets. Left: sequence logos of the −35 to +15 region relative to the TSS, right: percentage of the genes with sharp (red) and broad (blue) promoter in each gene set, determined by the IQW of the corresponding CAGE consensus cluster as demonstrated (A).
(E) Chromosomal plot showing the location of sharp (red) and broad (blue) promoter *znfs* on chr4 and the location of the *miR-430* cluster (orange).
(F and G) Expression activity of sharp and broad promoter *znfs* determined by CAGE-seq (F) and EU nascent RNA-seq (G).
(H and I) Comparison of the promoter motifs/architecture of the sharp (H) and broad (I) *znfs*.
(J and K) Distribution of H3K4me3 signal around the promoter for the sharp (K) and broad (J) *znfs*. Abbreviations: TSS, transcription start site; IQW, interquantile width.
See also Tables S2 and S4.

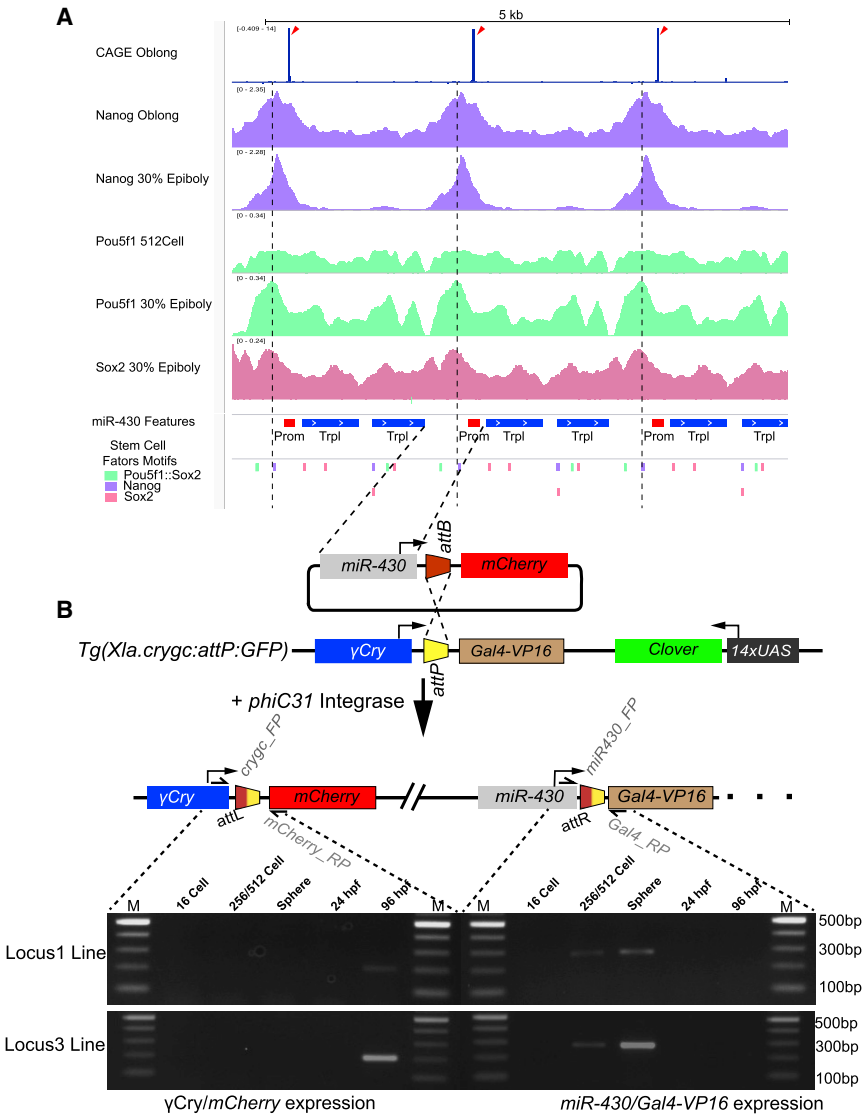


Figure 4. A single copy of the *miR-430* promoter including stem cell transcription-factor-binding sites is sufficient to activate reporter expression during the minor wave of genome activation

(A) Browser views showing CAGE-seq signal (top track, blue bar) marking the position of the TSS (red arrowheads). ChIP-seq signal of Nanog (purple), Pou5f1 (green), and Sox2 (pink) are shown below. Bottom annotation tracks show the *miR-430* gene features, with the core promoter (red), the precursor triplet (blue), and the position of predicted pluripotency-factor-binding sites. Purple dash line marks the Nanog binding site in the core promoter region close to the peak of the Nanog ChIP-seq signal.

(B) Top: schematic of the generation of *miR-430* promoter-containing stable transgenics reporter lines using the PhiC31 site specific integration system. Bottom: expression of the reporter *miR-430* promoter (left) and γ-crystalline promoter (right), detected by RT-PCR at the indicated stages. Abbreviations: TSS, transcription start site; γ-Cry, γ-crystalline C.

See also Figure S3.

expected. These results together demonstrate that the *miR-430* promoter sequence contained sufficient regulatory information to activate transcription at early pre-MBT/MBT stages.

The *miR-430* gene cluster and its activity are required for the formation of a transcription body at the minor wave of ZGA

To dissect the contribution of underlying genomic sequence and *miR-430* chr4 cluster transcription to transcription body formation, we used the CRISPR-Cas9 system to specifically target the *miR-430* loci and recruit Cas9-driven loss-of-function tools to the core of the transcription body. This system was optimized for both lesion-generating active Cas9 and catalytically dead Cas9 (dCas9), which blocks transcription by sterically hindering Pol II, in a process termed CRISPRi.[56] To address the impact of Cas9-mediated lesion and CRISPRi on *miR-430* activity, we used RT-PCR. A dramatic reduction in *miR-430* RNA was detected at the 512-cell stage, following targeting of the *miR-430* cluster by either active Cas9 or dCas9, comparable to total transcription loss achieved by triptolide or α-amanitin treatment (Figure 5A).

Using the MoVIE to image *miR-430* nascent RNA accumulation *in vivo*[22] following mosaic CRISPRi with dCas9-GFP, we monitored the effect of *miR-430* CRISPRi on the accumulation of *miR-430* RNA. A dual injection setup allowed for the labeling of *miR-430* RNA across the embryo, alongside mosaic and trackable CRISPRi reagents targeting either *miR-430* or a control gene (*gol*) in a mosaic fashion (Figure 5B). We observed efficient CRISPRi, with loss of *miR-430* transcription throughout the cell cycle and specifically in cells containing CRISPRi reagents targeting the *miR-430* cluster, suggesting a cell autonomous, direct effect (Figure 5B). In 512-cell stage embryos where Cas9 was targeted to the *miR-430* locus (Figures 5C and 5E), or dCas9 sterically blocked the miR-430 cluster (Figures 5D and 5F), we observed a loss of the transcription body; with nascent RNA accumulation and Pol II Ser2P signal reduced to background levels, and accompanied by a reduction in the size of *miR-430* territory (Figure 5E). As shown previously[22,23] the volume occupied by the *miR-430* locus chromatin correlated with the transcriptional activity of the *miR-430* cluster (Figure S4A). Additionally, under wild-type conditions the *miR-430* territory occupied a small nuclear volume (<0.07 μm^3) during early interphase of the 512-cell stage, which grew ($\leq$4 μm^3) as the cell progressed through prophase (Figure S4B), before re-compaction ready for the next cell division. A similar degree of transcription body loss (Figures S4C–S4E) and *miR-430* locus compaction (Figures S4F–S4H) was observed following global

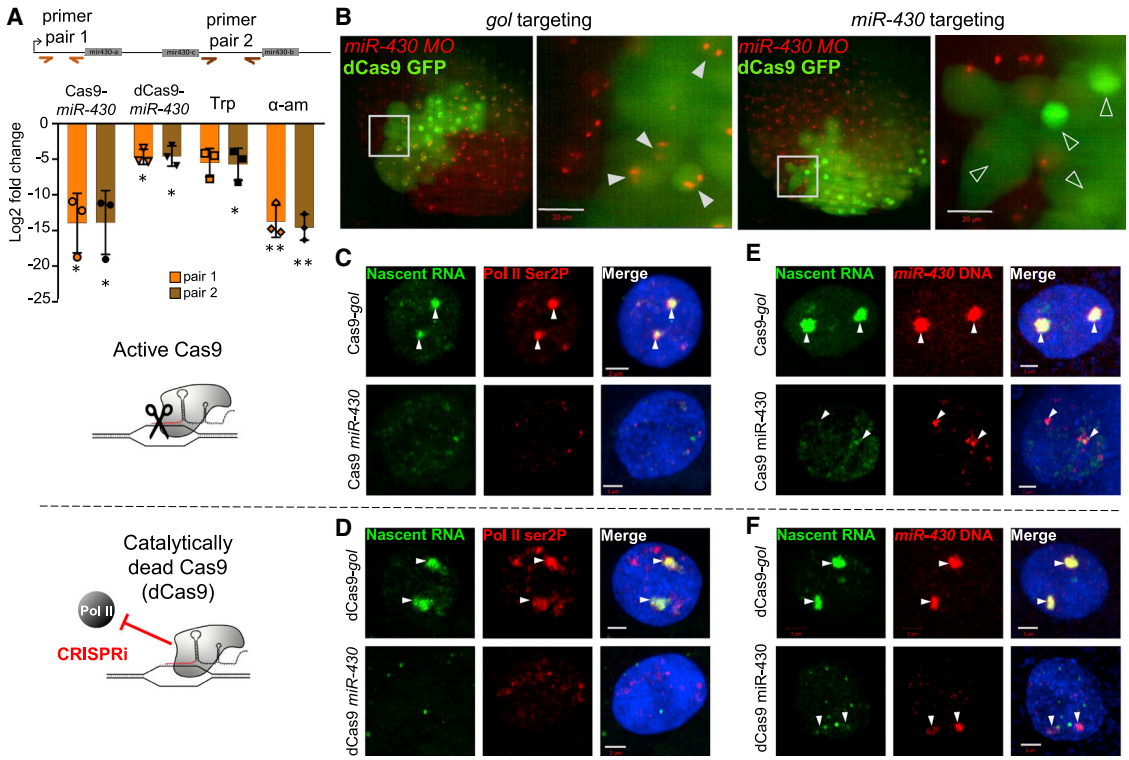


Figure 5. Manipulations that limit transcription of the *miR-430* cluster also disrupt the transcription body

(A) Schematic of a single *miR-430* triplet showing the position of two qRT-PCR primer sets (orange and brown arrows). Chart shows relative levels of *miR-430* RNA at the 512-cell stage, normalized to *5S rRNA*, following *miR-430* promoter targeting by Cas9 or dCas9, and triptolide (Trp) or α-amanitin (α-am) treatment, compared with control manipulations of active Cas9 targeting *gol* splice junction, dCas9 targeting *gol* splice junction and injection control, respectively. Data are based on three biological repeats. Error bars represent standard deviation and asterisks indicate significance, *$p < 0.05$ and **$p < 0.005$ (unpaired t test).

(B) *miR-430* MO (red) labeled transcription bodies in live 512-cell stage embryos following either *miR-430* promoter or *gol* splice junction mosaic targeting with dCas9-GFP (green). Gray boxes indicate enlarged region shown on the right; *gol* targeted: 10 embryos, *miR-430* targeted: 8 embryos. white arrowheads: nuclei with two transcription bodies, open arrowheads: nuclei without detectable transcription bodies, scale bars, 20 μm.

(C and D) EU labeling of nascent RNA (green), Pol II ser2P (red) immunostaining, and DAPI (blue) in 512-cell-stage embryos: active Cas9; *gol* targeted: 6 embryos, 31 nuclei; *miR-430* targeted: 6 embryos, 28 nuclei (C). Catalytically dead Cas9; *gol* targeted: 8 embryos, 52 nuclei; *miR-430* targeted: 5 embryos, 29 nuclei (D).

(E and F) EU labeling of nascent RNA (green) combined with FISH for *miR-430* DNA (red) and DAPI (blue) at 512-cell stage: active Cas9; *gol* targeted: 4 embryos, 21 nuclei. *miR-430* targeted: 4 embryos, 41 nuclei (E). Catalytically dead Cas9; *gol* targeted: 5 embryos, 41 nuclei. *miR-430* targeted: 8 embryos, 66 nuclei (F). White arrowheads (C–F) remnants of transcription body, scale bars, 2 μm.

See also Figure S4.

transcription inhibition. Expression analysis at dome stage for other early genes on chr4, confirmed the dynamic and transient nature of this locus compaction with no permanent reduction in expression following *miR-430* targeted manipulation distal to the *miR-430* locus (Figures S4I–S4L) suggesting that *miR-430* DNA volume changes following manipulations was not a result of chromosomal segregation or loss of large chr4 segments.

Together these observations suggest that transcription of the *miR-430* cluster is required for the formation and maintenance of the transcription body. We show that 3D expansion of the transcribed territory is dependent on *miR-430* transcription specifically, in line with previous observations of the repeats co-transcriptionally forming a growing transcription body which expands the volume occupied by the loci.[23] Our experiments identify high promoter density and subsequent transcription of the 0.6 Mb miR-430 repetitive locus as the primary reason for transcription body formation.

Activity from the chr4 *miR-430* locus is not required for minor wave ZGA globally

Nascent RNA and Pol II Ser2P imaging of pre-ZGA embryos showed that most minor wave transcription is concentrated within the *miR-430-driven* transcription body. This led us to query whether other early transcribed loci contributed or interacted with the transcription body, to enable their early escape from global transcription repression. To test whether the transcription body had a role in nuclear organization, we investigated whether manipulation of *miR-430* or its transcriptional activity would impact on other minor wave genes. Nascent RNA detection by RNA-seq upon *miR-430* CRISPRi did not lead to repression of most minor wave genes apart from a family of chr4 *znf* genes (Figure S5). This loss of *znf* gene activity could be a result of pleiotropic effects including potential interference with replication,[57] as suggested by activation of stress and DNA damage response genes (Figure S5B). Therefore, we focused on minor wave genes on

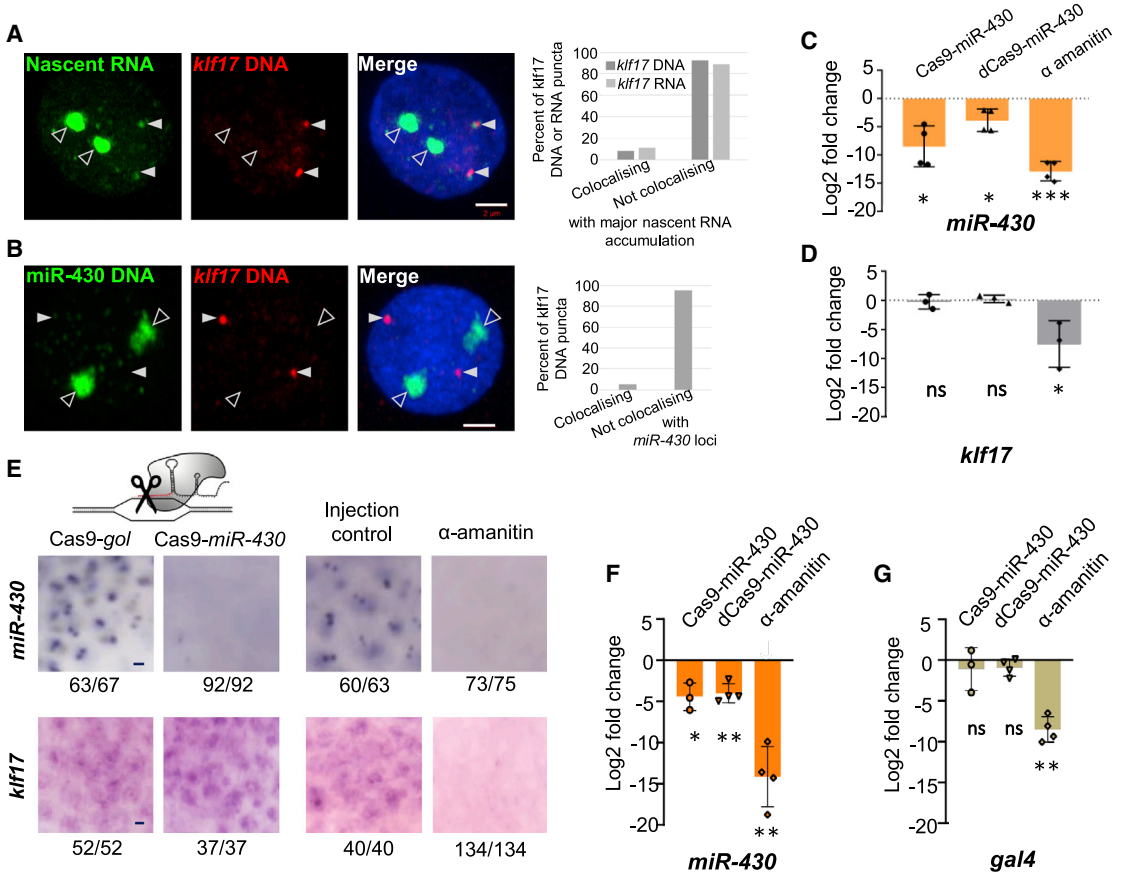


Figure 6. *miR-430* manipulations do not influence early zygotic transcription of minor wave genes on other chromosomes

(A) Nascent RNA (green) combined with FISH for *klf17* DNA (red) and DAPI (blue) in WT 512-cell stage embryos counter-stained with DAPI (blue). 59 nuclei from 16 embryos. Right, chart shows frequency of colocalization frequency between *klf17* DNA or RNA and major nascent RNA accumulation (transcription body) in 3D space (59 and 54 nuclei, respectively).

(B) Double DNA-FISH labeling the *miR-430* cluster (green) and the *klf17* locus on chr2 (red) in WT 512-cell embryos counter-stained with DAPI (blue). 51 nuclei from 8 embryos. Chart shows frequency of colocalization between *klf17* DNA and *miR-430* DNA in 3D space. White arrowheads indicate *klf17* locus, open arrowheads (A and B) indicate the *miR-430* transcription body in (A) and (B).

(C and D) Change in expression in WT 512-cell stage embryos, following manipulations, for *miR-430* normalized to 5S rRNA (C), and *klf17* normalized to TBP mRNA (D). Charts show mean $\log_2$(fold change) between Cas9/dCas9 targeting a control vs. targeting *miR430*, or α-amanitin treatment compared with injection control, from 4 biological repeats.

(E) Nascent *miR-430* (top) or *klf17* RNA (bottom) staining at 512-cell following manipulations. Numbers indicate frequency of staining pattern, scale bars, 5 μm.

(F and G) Relative change in expression, in Tg(Xla.crygc:attP-Gal4vp16,14UAS:Clover)UoBL1 embryos, at the 512-cell stage, following manipulations, for *miR-430* RNA, normalized to *5S rRNA* (F), or *gal4* RNA normalized to *TBP* mRNA (G). Charts show mean $\log_2$(fold change) between Cas9 (n = 3 biological repeats) or dCas9 (n = 4) targeting, a control vs. targeting *miR-430*, or α-amanitin treatment compared with injection control.

(C, D, F, and G) Error bars represent standard deviation and asterisks indicate significance: ns, $p > 0.05$, *$p < 0.05$, **$p < 0.005$ and ***$p < 0.0005$ (unpaired t test). See also Figures S5 and S6.

other chromosomes where replication interference was not expected.

We focused on *klf17*, a single-copy gene residing on chr2, which showed transcriptional activity at the 512-cell stage and a sharp promoter structure with TATA box. Using nascent RNA labeling combined with DNA-FISH, we observed that *klf17* loci reside in a distinct chromosomal territory to the transcription body, marked by nascent RNA (92.4% of *klf17* loci observed distal to a major nascent RNA accumulation marking the transcription body; Figure 6A). The distal (95.1%) nature of *klf17* locus from the *miR-430* locus was confirmed by dual labeling in DNA-FISH at the 512-cell stage (Figure 6B). Nascent RNA was accumulating proximal to the *klf17* loci (white arrowheads Figure 6A), suggesting that the speckles are sometimes visible both as nascent RNA and active Pol II (Figures 5C–5F) may be from other minor wave loci. These nascent RNA/RNA Pol II speckles were small with low signal intensity, emphasizing the small number of transcriptionally active loci across the whole genome and their relative expression level to the *miR-430* loci, at this early stage. Together, these data show that although the transcription body is highly enriched for nascent RNA, it is not a unique site of transcription within the early nucleus and single-copy genes can escape global transcriptional repression despite occupying distinct topological domains from the transcription body.

To investigate whether any transient interaction not captured by these fixed imaging approaches, enabled *miR-430* activity

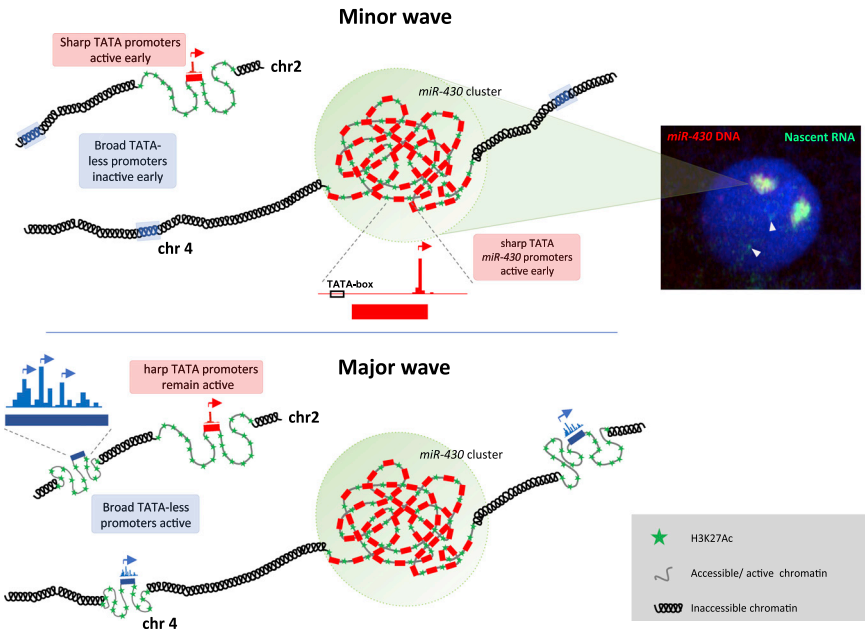


Figure 7. Models of minor wave gene activation in and outside of the *miR-430* transcription body

(A) The *miR-430* cluster on chr4 with high promoter density forms a microenvironment during the minor wave of ZGA. Minor wave gene activation occurs distal to the transcription body, such as at the *gata3* and *klf17* loci, which share promoter architecture features with *miR-430* gene but are activated independently from the transcription body (top). Genes that have distinct promoter features from minor wave genes are not activated until the major wave of ZGA (bottom), while minor wave genes continue to function during the major wave of ZGA.

to influence early *klf17* transcription we followed *klf17* RNA levels upon *miR-430* targeted disruption. The manipulation of *miR-430* by either Cas9 or dCas9 did not have significant effect on *klf17* RNA levels, while global transcription inhibition induced a significant loss in *klf17* RNA (Figures 6C–6E and S6A–S6C). Other early zygotically expressed genes, *wnt11f2* and *hspb1* (both chr5), also failed to show a response to *miR-430* targeted manipulations (Figures S6D and S6E). This lack of response to *miR-430* manipulations suggests that the transcription body did not exert a global effect influencing minor wave activation of genes on other chromosomes, or on the short arm of chr4 (*gata3*).

To further investigate dependence of minor wave of gene activation on the *miR-430* locus, we asked whether transgene activity from a single copy of the *miR-430* promoter, would be affected by loss of *miR-430* cluster activity. To answer this *miR-430* promoter manipulation was performed in the single *miR-430* promoter-driven *gal4* transgenic line (chr22 transgene; Figure 3B), and the transcriptional outputs of the *miR-430* cluster and the single *miR-430* promoter-driven *gal4* transgene were measured. The targeting guide RNA did not recognize the single transgenic *miR-430* promoter but specifically targeted the endogenous *miR-430* cluster. Figures 6F and 6G shows that single *miR-430* promoter activity was not significantly affected by either Cas9 or dCas9 targeting the chr4 *miR-430* cluster while global transcription block confirmed transgenic *miR-430* promoter activity before global genome activation. This suggests that the *miR-430* single promoter sequence contains the necessary determinants to efficiently engage in transcription before the main wave of global genome activation.

DISCUSSION

Much is already known about the mechanisms and regulatory networks underlying the main wave of ZGA, but the mechanisms governing the very first steps of global genome activation are still unclear.[1–3] Here, we explored the minor wave gene activation by studying the miR-430 cluster, the first known expressed genes in zebrafish.[20] We show that this cluster is composed of unexpectedly high promoter/gene density, which define and is required for the formation of a transcription body, enriched for nascent RNA and Pol II Ser2P during pre-ZGA cell cycles.[15,22,23] The promoter-dense locus structure provides explanation to the above listed nuclear topology-defining biochemical features and presents a useful experimental model for future transcription imaging and molecular dissection of native transcription machinery. We also demonstrated that not all minor wave transcription occurs in or is dependent on the activity of the *miR-430* transcription body. Moreover, we have demonstrated that minor wave genes including the *miR-430* cluster share epigenetic and promoter architecture features characteristically distinct form the major wave (Figure 7).

Extreme density of promoter/gene cluster underlies the formation of a minor wave-associated transcription body

We demonstrated by several independent lines of evidence that the copy number of *miR-430* transcribed units is almost two orders of magnitude higher than described in the reference genome GRCz11. This highlights the need for end-to-end genome assemblies to resolve highly repeated loci.[58] Our locus assemblies, from multiple zebrafish strains suggest that the hundreds of copies of promoters is a robust feature of common zebrafish strains. *MiR-430* homolog loci are also multicopy in medaka (manual analysis of the medaka[59] and goldfish genomes[60]) likely indicating a crucial role for abundance of these miRNAs in clearance of maternal mRNAs.[21] Nevertheless, the exact composition of the gene sets within the cluster appears to be variable and merits further analysis, given the significance of zebrafish chr4, with unique gene organization.[50]

Notably, genome activation is also concentrated in discrete, Pol II Ser2P-containing nuclear compartments in *Drosophila*.[16,48,61] Similarly, to zebrafish, genome activation in human and mouse embryos involves multicopy transcribed repeats.[62,63] Recent *in vivo* transcription imaging has revealed distinct Pol II Ser2P accumulation foci in several mammalian cells,[64,65] reminiscent to that seen in an exaggerated form in the minor wave in zebrafish embryos. Overall, these observations from a variety of models

suggest gene clusters are often organized into distinct nuclear topologies.[66] Thus, understanding the mechanisms underlying the formation of the zebrafish transcription body may reveal genome activation-associated transcription control as well as fundamental regulatory principles of transcription-associated nuclear topology.

The 0.6 Mb *miR-430* region packed with >300 promoters suggested that the mechanisms for the early escape from global transcriptional repression may arise from bulk properties and macro-nuclear dynamics. The *miR-430* promoter-dense cluster may form a transcriptionally competent body in which the nucleo-cytoplasmic ratio can be passed early in the local microenvironment. However, this model is not supported by two observations. First, *miR-430* early activation does not respond to change of the nucleo-cytoplasmic ratio by ploidy manipulation.[15,22] Second, in this study, minor wave activity of a single *miR-430* promoter inserted in a heterologous genomic location was demonstrated. This indicates that the large copy number of *miR-430* was not necessary for reaching a threshold of transcriptional competence, and the promoter sequence contained the necessary regulatory information for early activation. We propose that the high promoter density facilitates transcription body formation and boosts the promoter-autonomous responsiveness of the *miR-430* promoters for high levels of activation (Figure 7).

Distinct promoter features shared by minor wave ZGA genes

Previous studies have shown that *miR-430* transcription requires H3K27ac deposition, binding of stem cell factors Sox19b/SoxB1 Nanog and Pou5f1/Pou5f3,[25] and it is characterized by CTCF and Rad21 deployment[15,24,25] at a stage when topology associated domains have not yet formed.[32,67] However, the question remains: what mechanism triggers *miR-430* transcription and that of other minor wave genes? While the stem cell factors Pou5f3/Pou5f1, SoxB1/Sox19b, and Nanog were shown to be required for *miR-430* gene activation,[8,25,26] they are unlikely to be sufficient for minor wave genome activation, as many major wave genes also depend on these factors.[8,13,25,28,68]

We demonstrated that minor wave genes are enriched in sharp promoters, carry TATA box, and similarly to *Drosophila*[48,49] both are features that commonly lack H3K4me3 (reviewed in Lenhard et al.[69]). The TATA box and associated positionally constrained TSS usage resembles the maternal transcription initiation code we previously described.[42] Oocyte-active genes deposited as mRNA into the embryo are characterized by TATA-like sequences (W-box) and associated distance constraints on the TSS choice, similarly to that seen with the canonical TATA box. These TSS constraint similarities may indicate shared transcription initiation machineries regulating maternal and minor wave TSS determination despite the differences in their TSS profiles.

In contrast to minor wave genes, TATA-dependent sharp promoters are distinctly lacking at major wave-specific genes[42] suggesting functional distinction in the transcription initiation machineries acting on them.[70–72] However, TATA-dependent sharp promoters are not restricted to the minor wave and yet unidentified factors likely play crucial roles in specifying early zygotic gene activity (candidates include homologs of GAF[73] or Dux[74,75]). The correlation between TATA box/sharp promoters with minor wave activation is further supported by our analysis of minor wave *znf* genes interspersed with major wave *znf*s on chr4. These major wave genes either lack binding sites for minor-wave-active pioneering factors or their core promoter sequence drives promoter architecture differences in activation-specificity[72,76,77] resulting in the lack of responsiveness of broad promoters to minor-wave-active *trans*-activators. Future work may address how minor wave genes bind pioneering factors, which may selectively recognize TATA-box-dependent, sharp TSS core promoter codes or exploit a favorable chromatin environment at the minor wave genes.

When searching for the promoter features of minor wave genes, we observed a distinct lack of H3K4me3 at *miR-430* and at minor wave-active *znf* genes. This mark is generally associated with gene regulation but decoupled from transcriptional activity: with H3K4me3 detected on promoters well before ZGA, upon inhibition of ZGA, and also in transcriptionally inactive sperm.[14,30,35,42] While this histone mark may be paternally inherited, as seen in mouse,[78] we do not see indication of such histone-modification-dependent inheritance in activation of minor ZGA genes in zebrafish.

We also observed that sharp TSS, TATA promoters are less likely to carry the H3K4me3 mark than broad promoters (reviewed in Lenhard et al.[69]) and further suggesting that minor wave gene activation is not globally dependent on H3K4me3-associated mechanisms, as observed in *Drosophila*.[79] However, minor wave gene promoters are not always deficient of the H3K4me3 mark. It is conceivable that H3K4me3 is not functional during minor wave activation but is a pre-marking mechanism utilized later in development. H3K4me3 marks may be interpreted by distinct transcription machineries in different stages of development, similarly to how distinct promoter sequence determinants within the same promoters are selectively used during maternal-to-zygotic transition.[42]

Notably, minor wave genes, with H3K4me3 enriched promoters also show H3K4me3 enriched at the gene body and coupled to H3K27me3. This observation may indicate enrichment for chromatin bivalency on developmental regulator genes as described in embryonic stem cells (reviewed in Voigt et al.[80]).

MiR-430 promoter cluster forms the transcription body but is not required for all minor wave gene activation

Previously, it was demonstrated that *miR-430* expression is a component of the transcription body.[15,22,23] Here, we demonstrate that transcription of the *miR-430* cluster is specifically required for the formation of this transcription body. Similarly, we observe a reduction in nuclear volume occupied by the *miR-430* cluster of chr4 following CRISPRi targeting. This correlates with the proposal that high levels of transcription expand chromatin[23] and was suggested to form a favorable microenvironment, which may further promote *miR-430* transcription. Our results suggest that during the interphase of these early cell cycles, the *miR-430* cluster becomes de-compacted. When *miR-430* transcription is terminated at a later stage of development in wild type, or during the minor wave in manipulated embryos, the locus becomes compact and occupies a small nuclear volume, throughout the cell cycle.

The high density of binding sites for stem cell factors contained within the *miR-430* promoter cluster may function as "enhancers" for these genes. In this context the 0.6 Mb *miR-430*

locus, enriched in H3K27ac and stem cell factor binding sites, is not dissimilar from previously described ''enhancer clusters'' or ''super-enhancers.''[81] The large number of promoters observed here fits with a recent computational model of enhancer-promoter interactions, suggesting generality of equivalent topological and functional impact of enhancers and promoter clusters.[82]

We have shown that despite most of the nascent RNA and Pol II Ser2P accumulating at the *miR-430*-dependent transcription body, minor wave genes do not necessarily colocalize or depend on the *miR-430* transcription body, for their activity. This indicates that the specificity of minor wave activation is not based on an exclusive and localized mechanism for transcriptional competence within the transcription body. As exemplified by our single-copy transgenic *miR-430* promoters and that of endogenous single-copy genes, they can autonomously escape the global repressive environment prior to the major wave of ZGA. Similarly, topologically independent activation of minor wave ZGA genes has also been detected in *Drosophila*.[83] Our promoter architecture analyses suggest that the autonomy of minor wave gene activation likely resides in the distinct promoter structure characterized by the TATA box and sharp peak TSS choice and does not ubiquitously require H3K4me3 histone marks at the promoter.

The gradual nature of genome activation which manifests as a continuity between the minor and major wave observed in several animal models led to suggestions that the two waves are not distinct (reviewed in Vastenhouw et al.[3]). However, the distinct promoter structure of minor wave genes, which discriminate them from the major wave of ZGA, suggest discontinuity in the mechanisms of minor and major wave of genome activation.

Limitations of the study

Due to limitations of the currently available genome assembly software to handle large tandem repeats (such as the *miR-430* locus), the exact copy number/structure of the *de novo*-assembled *miR-430* locus may not be entirely accurate. Strain and individual variations are also possible.

Short bursts of transcription in the very-short cell cycles (15–20 min) during cleavage stages may limit the quantification accuracy of lowly expressed genes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Zebrafish maintenance
- METHOD DETAILS
  - Isolation of high molecular weight zebrafish genomic DNA
  - Oxford Nanopore sequencing
  - Pacific Biosciences (PacBio) sequencing
  - *De novo* genome assembly of long read data
  - Estimation of *miR-430* promoter number from raw long reads
  - Estimation of *miR-430* promoter number by digital droplet PCR (ddPCR)
  - Promoter density analysis
  - Processing of CAGE-seq data
  - Processing of ChIP-seq and ATAC-seq data
  - Promoter architecture analyses
  - Transcription factor binding motif analyses
  - Generation of stable transgenic zebrafish reporter lines
  - Identification of transgene integration sites
  - Reporter expression analyses of *miR-430* promoter stable transgenic lines by RT-PCR
  - Injections at single cell stage for ubiquitous delivery vs late-stage injection for mosaic treatments
  - Nascent RNA capture and sequencing upon dCas9 inhibition of miR-430 transcription
  - Analysis of the nascent (EU) RNA-seq data
  - Global transcription block with α-amanitin and triptolide treatment
  - PCNA staging
  - *miR-430* MO labelling
  - EU labelling of nascent RNA
  - 3D DNA-FISH (*miR-430* or klf17)
  - Expression analysis following miR-430 manipulations
  - Guide RNA production
  - In situ probe production
  - Live embryo imaging
  - Whole mount antibody staining
  - Whole mount in situ hybridization
  - Single-molecule fluorescent double in situ hybridization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2022.12.007.

ACKNOWLEDGMENTS

This work was funded by a Wellcome Trust Investigator Award (106955/Z/15/Z) and funding from the MDS Research Development Fund to F.M. and was supported in part (S.B.) by the Intramural Research Program of the NHGRI (ZIAHG200386-06). We thank Genomics Birmingham for sequencing, the imaging facility of the Technology Hub core, and the BMSU facility at the University of Birmingham. We thank S. Branford and J.-B. Cazier for support from the University of Birmingham BlueBEAR/CaStLeS high performance computing cluster. We also thank A. Jimenez-Gonzalez for embryo injections and P. Balwierz for advice on gene copy-number calculations. We thank M. Lagha for their smiFISH protocol. We thank Darius Balciunas for providing the pDB783:Xla.crygc-attP-Gal4vp16-14UAS:eGFP plasmid.

AUTHOR CONTRIBUTIONS

Y.H., L.W., and F.M. conceived the study. Y.H., C.W., and A.B. carried out long-read-based assembly. L.C. and Y.H. did transgenic work. L.W. did locus manipulations and locus-specific DNA/RNA imaging. Y.H., F.A., and S.G. carried out copy-number analyses. S.F. and R.S. did copy-number validations. Z.C. and S.B. did PacBio sequencing of zebrafish strains. Y.H., L.W., and F.M. wrote the manuscript with contributions by all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 11, 2021
Revised: August 10, 2022
Accepted: December 16, 2022
Published: January 23, 2023

REFERENCES

1. Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annu. Rev. Cell Dev. Biol. *30*, 581–613. https://doi.org/10.1146/annurev-cellbio-100913-013027.
2. Schulz, K.N., and Harrison, M.M. (2019). Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. *20*, 221–234. https://doi.org/10.1038/s41576-018-0087-x.
3. Vastenhouw, N.L., Cao, W.X., and Lipshitz, H.D. (2019). The maternal-to-zygotic transition revisited. Development *146*, dev161471. https://doi.org/10.1242/dev.161471.
4. Pálfy, M., Joseph, S.R., and Vastenhouw, N.L. (2017). The timing of zygotic genome activation. Curr. Opin. Genet. Dev. *43*, 53–60. https://doi.org/10.1016/j.gde.2016.12.001.
5. Wragg, J., and Müller, F. (2016). Transcriptional regulation during zygotic genome activation in zebrafish and other anamniote embryos. Adv. Genet. *95*, 161–194. https://doi.org/10.1016/bs.adgen.2016.05.001.
6. Amodeo, A.A., Jukam, D., Straight, A.F., and Skotheim, J.M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proc. Natl. Acad. Sci. USA *112*, E1086–E1095. https://doi.org/10.1073/pnas.1413990112.
7. Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., and Vastenhouw, N.L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. eLife *6*, e23326. https://doi.org/10.7554/eLife.23326.
8. Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360–364. https://doi.org/10.1038/nature12632.
9. Bártfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orbán, L., and Müller, F. (2004). TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. Curr. Biol. *14*, 593–598. https://doi.org/10.1016/j.cub.2004.03.034.
10. Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M.J., et al. (2007). The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. EMBO J. *26*, 3945–3956. https://doi.org/10.1038/sj.emboj.7601821.
11. Veenstra, G.J., Destrée, O.H., and Wolffe, A.P. (1999). Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition. Mol. Cell. Biol. *19*, 7972–7982. https://doi.org/10.1128/MCB.19.12.7972.
12. Blythe, S.A., and Wieschaus, E.F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. eLife *5*, e20148. https://doi.org/10.7554/eLife.20148.
13. Pálfy, M., Schulze, G., Valen, E., and Vastenhouw, N.L. (2020). Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. PLoS Genet. *16*, e1008546. https://doi.org/10.1371/journal.pgen.1008546.
14. Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Østrup, O., Winata, C., Mathavan, S., Müller, F., Aleström, P., and Collas, P. (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Dev. Cell *21*, 993–1004. https://doi.org/10.1016/j.devcel.2011.10.008.
15. Chan, S.H., Tang, Y., Miao, L., Darwich-Codore, H., Vejnar, C.E., Beaudoin, J.D., Musaev, D., Fernandez, J.P., Benitez, M.D.J., Bazzini, A.A., et al. (2019). Brd4 and P300 confer transcriptional competency during zygotic genome activation. Dev. Cell *49*, 867–881.e8. e868. https://doi.org/10.1016/j.devcel.2019.05.037.
16. Blythe, S.A., and Wieschaus, E.F. (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell *160*, 1169–1181. https://doi.org/10.1016/j.cell.2015.01.050.
17. Collart, C., Smith, J.C., and Zegerman, P. (2017). Chk1 inhibition of the replication factor Drf1 guarantees cell-cycle elongation at the Xenopus laevis mid-blastula transition. Dev. Cell *42*, 82–96.e3. https://doi.org/10.1016/j.devcel.2017.06.010.
18. Strong, I.J.T., Lei, X., Chen, F., Yuan, K., and O'Farrell, P.H. (2020). Interphase-arrested Drosophila embryos activate zygotic gene expression and initiate mid-blastula transition events at a low nuclear-cytoplasmic ratio. PLOS Biol. *18*, e3000891. https://doi.org/10.1371/journal.pbio.3000891.
19. Abe, K.I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M.G., Schultz, R.M., and Aoki, F. (2018). Minor zygotic gene activation is essential for mouse preimplantation development. Proc. Natl. Acad. Sci. USA *115*, E6780–E6788. https://doi.org/10.1073/pnas.1804309115.
20. Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. *6*, 285–292. https://doi.org/10.1016/j.celrep.2013.12.030.
21. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79. https://doi.org/10.1126/science.1122689.
22. Hadzhiev, Y., Qureshi, H.K., Wheatley, L., Cooper, L., Jasiulewicz, A., Van Nguyen, H., Wragg, J.W., Poovathumkadavil, D., Conic, S., Bajan, S., et al. (2019). A cell cycle-coordinated polymerase II transcription compartment encompasses gene expression before global genome activation. Nat. Commun. *10*, 691. https://doi.org/10.1038/s41467-019-08487-5.
23. Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Jülicher, F., Honigmann, A., Zaburdaev, V., and Vastenhouw, N.L. (2021). Transcription organizes euchromatin via microphase separation. Nat. Commun. *12*, 1360. https://doi.org/10.1038/s41467-021-21589-3.
24. Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., O'Sullivan, J.M., and Horsfield, J.A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. Development *145*, dev156521. https://doi.org/10.1242/dev.156521.
25. Miao, L., Tang, Y., Bonneau, A.R., Chan, S.H., Kojima, M.L., Pownall, M.E., Vejnar, C.E., Gao, F., Krishnaswamy, S., Hendry, C.E., and Giraldez, A.J. (2022). The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. Mol. Cell *82*, 986–1002.e9. https://doi.org/10.1016/j.molcel.2022.01.024.
26. Kuznetsova, K., Chabot, N.M., Ugolini, M., Wu, E., Lalit, M., Oda, H., et al. (2022). Nanog organizes transcription bodies. Curr Biol. https://doi.org/10.1016/j.cub.2022.11.015.
27. Leichsenring, M., Maes, J., Mössner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science *341*, 1005–1009. https://doi.org/10.1126/science.1242527.
28. Veil, M., Yampolsky, L.Y., Grüning, B., and Onichtchouk, D. (2019). Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. Genome Res. *29*, 383–395. https://doi.org/10.1101/gr.240572.118.
29. Gentsch, G.E., Spruce, T., Owens, N.D.L., and Smith, J.C. (2019). Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. Nat. Commun. *10*, 4269. https://doi.org/10.1038/s41467-019-12263-w.

30. Wu, S.F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res. *21*, 578–589. https://doi.org/10.1101/gr.113167.110.

31. Liu, G., Wang, W., Hu, S., Wang, X., and Zhang, Y. (2018). Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. Genome Res. *28*, 998–1007. https://doi.org/10.1101/gr.228833.117.

32. Wike, C.L., Guo, Y., Tan, M., Nakamura, R., Shaw, D.K., Díaz, N., Whittaker-Tademy, A.F., Durand, N.C., Aiden, E.L., Vaquerizas, J.M., et al. (2021). Chromatin architecture transitions from zebrafish sperm through early embryogenesis. Genome Res. *31*, 981–994. https://doi.org/10.1101/gr.269860.120.

33. Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., Zhang, Y., Wang, Q., Li, Y., Meng, A., et al. (2018). Widespread enhancer dememorization and promoter priming during parental-to-zygotic transition. Mol. Cell *72*, 673–686.e6. https://doi.org/10.1016/j.molcel.2018.10.017.

34. Murphy, P.J., Wu, S.F., James, C.R., Wike, C.L., and Cairns, B.R. (2018). Placeholder nucleosomes underlie germline-to-embryo DNA methylation reprogramming. Cell *172*, 993–1006.e13. https://doi.org/10.1016/j.cell.2018.01.022.

35. Zhu, W., Xu, X., Wang, X., and Liu, J. (2019). Reprogramming histone modification patterns to coordinate gene expression in early zebrafish embryos. BMC Genomics *20*, 248. https://doi.org/10.1186/s12864-019-5611-7.

36. Westerfield, M.; ZFIN (2000). The Zebrafish Book: a Guide for the Laboratory Use of Zebrafish Danio (Brachydanio) rerio, Fourth Edition (ZFIN).

37. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.

38. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

39. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). Blast+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

40. Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S., et al. (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res. *23*, 1938–1950. https://doi.org/10.1101/gr.153692.112.

41. Baranasic, D., Hörtenhuber, M., Balwierz, P.J., Zehnder, T., Mukarram, A.K., Nepal, C., Várnai, C., Hadzhiev, Y., Jimenez-Gonzalez, A., Li, N., et al. (2022). Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. Nat. Genet. *54*, 1037–1050. https://doi.org/10.1038/s41588-022-01089-w.

42. Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381–385. https://doi.org/10.1038/nature12974.

43. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107.

44. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49. https://doi.org/10.1038/nature09906.

45. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283. https://doi.org/10.1038/nature09692.

46. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318. https://doi.org/10.1038/ng1966.

47. Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922–926. https://doi.org/10.1038/nature08866.

48. Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., and Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. eLife *2*, e00861. https://doi.org/10.7554/eLife.00861.

49. Pimmett, V.L., Dejean, M., Fernandez, C., Trullo, A., Bertrand, E., Radulescu, O., and Lagha, M. (2021). Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. Nat. Commun. *12*, 4504. https://doi.org/10.1038/s41467-021-24461-6.

50. White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. eLife *6*, e30860. https://doi.org/10.7554/eLife.30860.

51. Xu, C., Fan, Z.P., Müller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T., et al. (2012). Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. Dev. Cell *22*, 625–638. https://doi.org/10.1016/j.devcel.2012.01.003.

52. Frith, M.C., Li, M.C., and Weng, Z. (2003). Cluster-Buster: finding dense clusters of motifs in DNA sequences. Nucleic Acids Res. *31*, 3666–3668. https://doi.org/10.1093/nar/gkg540.

53. Roberts, J.A., Miguel-Escalada, I., Slovik, K.J., Walsh, K.T., Hadzhiev, Y., Sanges, R., Stupka, E., Marsh, E.K., Balciuniene, J., Balciunas, D., et al. (2014). Targeted transgene integration overcomes variability of position effects in zebrafish. Development *141*, 715–724. https://doi.org/10.1242/dev.100347.

54. Hadzhiev, Y., Miguel-Escalada, I., Balciunas, D., and Müller, F. (2016). Testing of cis-regulatory elements by targeted transgene integration in zebrafish using PhiC31 integrase. In Zebrafish: Methods and Protocols, K. Kawakami, E.E. Patton, and M. Orger, eds. (Springer), pp. 81–91. https://doi.org/10.1007/978-1-4939-3771-4_6.

55. Davidson, A.E., Balciunas, D., Mohn, D., Shaffer, J., Hermanson, S., Sivasubbu, S., Cliff, M.P., Hackett, P.B., and Ekker, S.C. (2003). Efficient gene delivery and gene expression in zebrafish using the Sleeping Beauty transposon. Dev. Biol. *263*, 191–202. https://doi.org/10.1016/j.ydbio.2003.07.013.

56. Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. *8*, 2180–2196. https://doi.org/10.1038/nprot.2013.132.

57. Matozel, E.K., Parziale, S., and Price, A.C. (2022). A programmable DNA roadblock system using dCas9 and multivalent target sites. PLoS One *17*, e0268099. https://doi.org/10.1371/journal.pone.0268099.

58. Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. *13*, 36–46. https://doi.org/10.1038/nrg3117.

59. Ichikawa, K., Tomioka, S., Suzuki, Y., Nakamura, R., Doi, K., Yoshimura, J., Kumagai, M., Inoue, Y., Uchida, Y., Irie, N., et al. (2017). Centromere evolution and CpG methylation during vertebrate speciation. Nat. Commun. *8*, 1833. https://doi.org/10.1038/s41467-017-01982-7.

60. Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., Wada, H., Fujiyama, A., Toyoda, A., Zhang, S., et al. (2019). De novo assembly of the goldfish (Carassius auratus) genome and the evolution of genes after whole-genome duplication. Sci. Adv. *5*, eaav0547. https://doi.org/10.1126/sciadv.aav0547.

61. Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin architecture emerges during zygotic genome activation independent of transcription. Cell *169*, 216–228.e19. e219. https://doi.org/10.1016/j.cell.2017.03.024.

62. Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221–225. https://doi.org/10.1038/nature14308.

63. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63. https://doi.org/10.1038/nature11244.

64. Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. Science *341*, 664–667. https://doi.org/10.1126/science.1239053.

65. Conic, S., Desplancq, D., Ferrand, A., Fischer, V., Heyer, V., Reina San Martin, B., Pontabry, J., Oulad-Abdelghani, M., Babu, N.K., Wright, G.D., et al. (2018). Imaging of native transcription factors and histone phosphorylation at high resolution in live cells. J. Cell Biol. *217*, 1537–1552. https://doi.org/10.1083/jcb.201709153.

66. Cook, P.R., and Marenduzzo, D. (2018). Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. Nucleic Acids Res. *46*, 9895–9906. https://doi.org/10.1093/nar/gky763.

67. Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic loss and gain of chromatin architecture throughout zebrafish development. Cell Rep. *24*, 1–10.e4. https://doi.org/10.1016/j.celrep.2018.06.003.

68. Gao, M., Veil, M., Rosenblatt, M., Riesle, A.J., Gebhard, A., Hass, H., et al. (2022). Pluripotency factors determine gene expression repertoire at zygotic genome activation. Nat Commun *13*, 788. https://doi.org/10.1038/s41467-022-28434-1.

69. Lenhard, B., Sandelin, A., and Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. *13*, 233–245. https://doi.org/10.1038/nrg3163.

70. Müller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. Curr. Opin. Genet. Dev. *20*, 533–540. https://doi.org/10.1016/j.gde.2010.06.004.

71. Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13–25. https://doi.org/10.1016/j.cell.2014.02.009.

72. Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. *19*, 621–637. https://doi.org/10.1038/s41580-018-0028-8.

73. Gaskill, M.M., Gibson, T.J., Larson, E.D., and Harrison, M.M. (2021). GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. eLife *10*, e66668. https://doi.org/10.7554/eLife.66668.

74. Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *49*, 925–934. https://doi.org/10.1038/ng.3844.

75. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. *49*, 941–945. https://doi.org/10.1038/ng.3858.

76. Gehrig, J., Reischl, M., Kalmár, E., Ferg, M., Hadzhiev, Y., Zaucker, A., Song, C., Schindler, S., Liebel, U., and Müller, F. (2009). Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. Nat. Methods *6*, 911–916. https://doi.org/10.1038/nmeth.1396.

77. Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559. https://doi.org/10.1038/nature13994.

78. Lismer, A., Dumeaux, V., Lafleur, C., Lambrot, R., Brind'Amour, J., Lorincz, M.C., and Kimmins, S. (2021). Histone H3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring. Dev. Cell *56*, 671–686.e6. https://doi.org/10.1016/j.devcel.2021.01.014.

79. Li, X.Y., Harrison, M.M., Villalta, J.E., Kaplan, T., and Eisen, M.B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. eLife *3*, e03737. https://doi.org/10.7554/eLife.03737.

80. Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338. https://doi.org/10.1101/gad.219626.113.

81. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A phase separation model for transcriptional control. Cell *169*, 13–23. https://doi.org/10.1016/j.cell.2017.02.007.

82. Zhu, I., Song, W., Ovcharenko, I., and Landsman, D. (2021). A model of active transcription hubs that unifies the roles of active promoters and enhancers. Nucleic Acids Res. *49*, 4493–4505. https://doi.org/10.1093/nar/gkab235.

83. Huang, S.K., Whitney, P.H., Dutta, S., Shvartsman, S.Y., and Rushlow, C.A. (2021). Spatial organization of transcribing loci during early genome activation in Drosophila. Curr. Biol. *31*, 5102–5110.e5. https://doi.org/10.1016/j.cub.2021.09.027.

84. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics *36*, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

85. Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., and Schatz, M.C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. *20*, 224. https://doi.org/10.1186/s13059-019-1829-6.

86. Otto, C., Stadler, P.F., and Hoffmann, S. (2014). Lacking alignments? The next-generation sequencing mapper segemehl revisited. Bioinformatics *30*, 1837–1843. https://doi.org/10.1093/bioinformatics/btu146.

87. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

88. Haberle, V., Forrest, A.R., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015). CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. *43*, e51. https://doi.org/10.1093/nar/gkv054.

89. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

90. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

91. Volpe, M., Miralto, M., Gustincich, S., and Sanges, R. (2018). ClusterScan: simple and generalistic identification of genomic clusters. Bioinformatics *34*, 3921–3923. https://doi.org/10.1093/bioinformatics/bty486.

92. Gel, B., and Serra, E. (2017). karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics *33*, 3088–3090. https://doi.org/10.1093/bioinformatics/btx346.

93. Stempor, P., and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res. *1*, 14. https://doi.org/10.12688/wellcomeopenres.10004.1.

94. Nettling, M., Treutler, H., Grau, J., Keilwagen, J., Posch, S., and Grosse, I. (2015). DiffLogo: a comparative visualization of sequence motifs. BMC Bioinformatics *16*, 387. https://doi.org/10.1186/s12859-015-0767-x.

95. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

96. Green, M.R., and Sambrook, J. (2017). Isolation of high-molecular-weight DNA from mammalian tissues using proteinase K and phenol. Cold Spring Harb. Protoc. *2017*. https://doi.org/10.1101/pdb.prot093484.

97. Manni, M., Berkeley, M.R., Seppey, M., and Zdobnov, E.M. (2021). BUSCO: assessing genomic data quality and beyond. Curr. Protoc. *1*, e323. https://doi.org/10.1002/cpz1.323.

98. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

99. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. *48*, D87–D92. https://doi.org/10.1093/nar/gkz1001.

100. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94. https://doi.org/10.1093/nar/gkh012.

101. Kawakami, K. (2004). Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol. *77*, 201–222. https://doi.org/10.1016/s0091-679x(04)77011-9.

102. Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev. Dyn. *236*, 3088–3099. https://doi.org/10.1002/dvdy.21343.

103. Green, M.R., and Sambrook, J. (2019). Inverse polymerase chain reaction (PCR). Cold Spring Harb. Protoc. *2019*. https://doi.org/10.1101/pdb.prot095166.

104. Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., et al. (2016). smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. *44*, e165. https://doi.org/10.1093/nar/gkw784.

105. Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr. Biol. *23*, 2140–2145. https://doi.org/10.1016/j.cub.2013.08.054.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Pol II S2P mouse monoclonal antibody	Diagenode	Cat# C15200005; RRID: AB_2713925
Chemicals, peptides, and recombinant proteins		
mApple:PCNA fusion protein	Hadzhiev et al.[22]	NA
Fluorescently labeled morpholino oligonucleotides	Hadzhiev et al.[22]	NA
dCas9 protein	NEB	M0652
dCas9-GFP protein	Novateinbio	PR-137213G
Cas9 protein	NEB	M0386M
Critical commercial assays		
Genomic DNA Ligation Sequencing Kit	Oxford Nanopore	SQK-LSK110
FISH Tag™ DNA Multicolor Kit	Thermo Fisher	F32951
Click-iT™ RNA Alexa Fluor™ 488 Imaging Kit	Thermo Fisher	C10329
Click-iT™ Nascent RNA Capture Kit	Thermo Fisher	C10365
HiScribe RNA synthesis T7 kit	NEB	E2040
NEBNext® Ultra™ II Directional RNA Library Prep Kit	NEB	E7760S
Deposited data		
Oxford Nanopore and PacBio long read sequencing and Nascent RNA capture sequencing	This paper	BioProject: PRJNA900028
4-thio-UTP nascent RNA-seq	Heyn et al.[20]	BioProject: PRJNA207343
CAGE-seq data	Nepal et al.[40]	SRA: SRA055273
H3K4me3 and H3K27me3 ChIP-seq	Zhu et al.[35]	BioProject: PRJNA434216
H3K27ac ChIP-seq	Zhang et al.[33]	BioProject: PRJNA473799
ATAC-seq	Liu et al.[31]	BioProject: PRJNA395463
Nanog ChIP-seq	Xu et al.[51]	BioProject: PRJNA156233
Pou5f1 and Sox2 ChIP-seq	Leichsenring et al.[27]	BioProject: PRJNA171706
Experimental models: Organisms/strains		
Zebrafish transgenic line: *Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL1*	This paper	N/A
Zebrafish transgenic line: *Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL3*	This paper	N/A
Zebrafish transgenic line: *Tg(Xla.crygc:attL-mCherry-miR430-attR-Gal4vp16, 14UAS:Clover)UoBL1*	This paper	N/A
Zebrafish transgenic line: *Tg(Xla.crygc:attL-mCherry-miR430-attR-Gal4vp16, 14UAS:Clover)UoBL3*	This paper	N/A
Recombinant DNA		
pDB783:Xla.crygc-attP-Gal4vp16-14UAS:Clover	This paper	N/A
pJET:miR430-attB-mCherry	This paper	N/A
Software and algorithms		
Canu assembler (v.2.1.1)	Koren et al.[37]	https://github.com/marbl/canu/releases
ncbi-blast+(v2.11.0)	Altschul et al.[38] and Camacho et al.[39]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
purge_dups (v1.2.5)	Guan et al.[84]	https://github.com/dfguan/purge_dups
RaGOO (v1.1)	Alonge et al.[85]	https://github.com/malonge/RaGOO
segemehl aligner (v0.3.4)	Otto et al.[86]	https://www.bioinf.uni-leipzig.de/Software/segemehl/
bowtie (v1.2.3)	Langmead et al.[87]	https://github.com/BenLangmead/bowtie/releases
CAGEr (v1.18)	Haberle et al.[88]	https://doi.org/10.18129/B9.bioc.CAGEr

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STAR (v2.7.3a)	Dobin et al.[89]	https://github.com/alexdobin/STAR/
cluster-buster	Frith et al.[52]	https://github.com/weng-lab/cluster-buster/
DESeq2 (v1.36.0)	Love et al.[90]	https://doi.org/10.18129/B9.bioc.DESeq2
ClusterScan(v0.2.2)	Volpe et al.[91]	https://github.com/pyrevo/ClusterScan
karyoploteR (v1.18.0)	Gel et al.[92]	https://doi.org/10.18129/B9.bioc.karyoploteR
SeqPlots(v1.27)	Stempor et al.[93]	https://doi.org/10.18129/B9.bioc.seqplots http://przemol.github.io/seqplots/
DiffLogo	Nettling et al.[94]	https://doi.org/10.18129/B9.bioc.DiffLogo
ggseqlogo	Wagih.[95]	https://github.com/omarwagih/ggseqlogo

RESOURCE AVAILABILITY

Lead contact

Ferenc Mueller (f.mueller@bham.ac.uk)

Materials availability

Materials and reagents generated in this study are available upon request from the lead contact.

Data and code availability

- Long-read sequencing and nascent (EU) RNA-seq data are deposited at NCBI Sequence Read Archive (SRA) under BioProject: PRJNA900028. The sequences of the *de novo* assembled contigs, containing the *miR-430* locus, i.e. contig 1 and contig 2 has been deposited in the GenBank database under accession numbers GenBank: OP854626 and GenBank: OP854627 respectively.
- This paper does not report original code.
- Raw microscopy images and additional information required for data reanalysis are available upon request from the lead contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Zebrafish maintenance

All zebrafish strains were maintained in a designated facility (according to UK Home Office regulations) in a recirculating system (ZebTEC, Tecniplast) at 26°C in a 10-h dark, 14-h light photoperiod and fed 3 times daily. Fish were kept in 3 or 6 litre containers at recommended density of 5 adult fish per litre.

Zebrafish embryos were obtained by natural breeding and maintained at 28°C in E3 medium until desired developmental stages were reached.

Animal work presented in this study was carried out under UK Home Office project licence P51AB7F76 assigned to the University of Birmingham, UK, except the generation of heat shock diploid ‘‘double heterozygous’’ Tuebingen-AB hybrid zebrafish, which was carried out in Shawn Burgess’s laboratory at NIH/NHGRI.

METHOD DETAILS

Isolation of high molecular weight zebrafish genomic DNA

High molecular weight genomic DNA was isolated from one AB male by proteinase K and phenol extraction following the protocol by Green and Sambrook.[96] Adult fish was euthanized by overdose of anaesthetic (Tricaine) and snap frozen in liquid nitrogen. The frozen tissue was pulverised with liquid nitrogen precooled ceramic pestle and mortar. The powder was slowly spread over the surface of 20 ml lysis buffer in 100 ml beaker. After complete submerging of the tissue powder into the lysis buffer the lysate was transferred to 50 ml falcon tube and treated with pancreatic RNase for 1 h at 37°C and proteinase K overnight at 50°C, followed by three phenol extractions. The DNA was precipitated from the purified lysate by addition of 0.2 volume of 10 M ammonium acetate and 2 volumes of ethanol, washed twice with 70% ethanol air-dried and reconstituted in TE buffer.

Oxford Nanopore sequencing

Nanopore sequencing library was prepared with Genomic DNA Ligation Sequencing Kit (Oxford Nanopore, SQK-LSK110) according to the manufacture instructions, using 1 μg genomic DNA as input and sequenced on Oxford Nanopore Promethion device. The sequencing run resulted in 1.7133×10^{7} reads and 1.3273×10^{11} bp in total (app. 44x coverage of the ~1.7Gb zebrafish genome).

Pacific Biosciences (PacBio) sequencing

For the PacBio assembly, Gynogenic offspring of Tuebingen-AB hybrid zebrafish were generated according to the protocol for Production Of Homozygous Diploid Embryos by heat shock as described in The Zebrafish Book.[36] In more details zebrafish eggs were in-vitro fertilised with ultraviolet light (UV) inactivated sperm. Thirteen min after fertilisation, embryos were heat shocked at 41°C in water bath for 2 min and immediately transferred to 28°C, causing the fertilised embryos to skip the first cell division. Homozygosity was confirmed by PCR amplifying 8 independent loci and confirming there were no single nucleotide variants (SNVs) within any sequence. Genomic DNA was purified using TissueLyser II (Qiagen) and Blood & Cell Culture DNA Maxi Kit (Qiagen). The molecular size of genomic DNA at the peak of 40- to 50-kb was confirmed using the Pippin pulse electroporation system (NIPPON genetics). Genomic DNA from the TU/AB double heterozygous fish described above was used to perform whole-genome shotgun sequencing on a PacBio RS II sequencer. ~16 million subreads with a peak length of ~8kb (100X coverage) were collected and used to de novo assemble the genome using the CANU assembler.

De novo genome assembly of long read data

The Oxford nanopore and PacBio sequencing reads were assembled into contigs using Canu software v.2.1.1[37] with the following parameters for nanopore reads:

- nanopore genomeSize=1.7g minReadLength=5000 minOverlapLength=1000 and PacPio reads:
- pacbio genomeSize=1.7g minReadLength=1000 minOverlapLength=500

To produce guided whole genome assembly from the nanopore data a haplotype purge was performed on the assembled contigs with purge_dups,[84] following the pipeline guide with default parameters as outlined on (https://github.com/dfguan/purge_dups).

Using the current zebrafish genome (GRCz11) as reference, a guided scaffolding into chromosomes of the haplotig purged assembly (~1.4 Mb in size) was performed with RaGOO v1.1[85] according to the default pipeline (https://github.com/malonge/RaGOO), enabling misassembly correction with the error corrected nanopore reads (generated by Canu during the assembly step):

ragoo.py hap_purged_ctigs.fa GRCz11chr_ref.fa -R CorrNPRds.fa -T corr

All assemblies were carried out on the University of Birmingham high performance computing cluster BlueBEAR.

Estimation of *miR-430* promoter number from raw long reads

To estimate the *miR-430* promoter copy number 250 bp promoter region sequences of *miR-430* and 6 single-copy genes six single copy genes were randomly selected from the Benchmarking Universal Single-Copy Orthologs (BUSCO) database[97,98] (Actinopterygii gene set) and confirmed by BLAST[38,39] search that the selected 250 bp sequence produced a single hit in the current (GRCz11) zebrafish reference genome. The *miR-430* and the 6 single copy genes promoter sequences were used to perform a BLAST search against raw reads BLAST database for each sequencing run. BLAST hits with bit score $\geq$ 150 were counted for each gene and *miR-430* counts were normalised to each of the 6 single copy genes and the average was used as *miR-430* promoter number estimate. The genomic coordinates (danRer11) proximal promoter coordinates of *miR-430* and the 6 single copy genes are provided in Table S1.

Estimation of *miR-430* promoter number by digital droplet PCR (ddPCR)

Genomic DNA from three stains AB, Tuebingen (TU) and Tuebingen Longfin (TL) was prepared from 50, 2-day old embryos collected from a single pair for each strain. The DNA extraction was carried out with the PureLink™ Genomic DNA Mini Kit (ThermoFisher Scientific, K182001) following the manufacture instructions. Genomic DNA was digested with EcoRI-HF (New England Biolabs, NEB), for 1 h. Initially dilutions down to 2 ng/μl were prepared from the digested DNA and quantified spectrophotometrically, using NanoDROP (Thermo Scientific, Thermo Fisher Inc, USA). Further dilutions in TE buffer (10mM Tris-HCl, 1 mM EDTA, pH 7.5) to a final concentration of 100 pg/μl were performed using (Fluro dsDNA High Sensitivity kit, DeNovix California, USA), following manufacturer's instructions. Standard curve ranging from 0 ng/ml to 25 ng/ml, DNA samples and blank (no template) were measured in duplicate using DS-11 spectrophotometer-fluorometer (DeNovix Inc. USA). By plotting the measured fluorescence versus DNA, concentrations were calculated through a standard curve plot.

For the ddPCR reaction mixture, DNA was diluted to 50 ng/μl and then added to 2 ×DX200™ ddPCR™ Evagreen supermix (Bio-Rad, USA) and primers (0.5 μM) in final volume of 22 μl. 20μl of mixture was loaded into a cartridge (Bio-Rad, USA) together with 70 μl of droplet generation oil (Bio-Rad, USA) and placed into the droplet generator (Bio-Rad, USA). After that, each sample was transferred to a 96-well PCR plate (Bio-Rad, USA) and PCR amplification was performed using C1000 thermal cycler (Bio-Rad, USA) with the following program: 95°C for 5 min, 40 cycles at 95°C for 30 sec and 60°C for 30 sec; 4°C for 5 min followed by 90°C for 5 min; 4° C for 30 min. After the amplification, the plate was loaded on the QX200 Droplet Digital Reader (Bio-Rad, USA) and acquired data were analysed with QuantaSoft Analysis Pro software (Bio-Rad, USA).

After a serial dilution assay, 50 ng was selected as DNA concentration giving 1 copy/μl of the *shha*, single copy gene, used as invariant gene. The final number of *miR-430* promoter copies is calculated by the ratio of *miR-430* primer pair (copy/μl) and the *shha* primer pair (copy/μl) (Tables S4 and S5).

Promoter density analysis

CAGE-seq data from Nepal et al.[40] were mapped to the zebrafish long-reads assembly genome with Bowtie v1.2.3[87] in default n-mode allowing 2 mismatches in the 27 bp seed region and reporting all the best alignments up to 1000 (-S -n 2 -l 27 -a -m 1000 –best –strata). The CAGEr package[88] was used for downstream processing of the data. First, CAGE Transcription Start Site (CTSS) were called in each of the analysed samples removing the additional G nucleotide (due to the CAGE protocol) not mapping to the genome. Next, reads were counted at each CTSS and subsequently normalized as tags per million (tpm). CTSS supported by at least 0.5 tpm in at least two samples and closer than 20 nucleotides were clustered thus defining the so-called transcriptional clusters (TCs), discarding singletons TCs (i.e., clusters containing only one CTSS). TCs were then trimmed on the edges to obtain a more robust estimation of the promoter width. Toward this end, first the cumulative distribution of CAGE signal along the promoter was defined and the promoter region comprised between the 10th and 90th percentiles of the CAGE signal distribution was considered, as suggested in.[88] Finally, TCs across all the analysed samples were aggregated, if supported by at least 5 tpm and closer than 100 bp, to form consensus clusters (CC) for downstream analyses. Having defined the final set of promoters (i.e., consensus clusters), the promoter density was calculated as follows. First, using the clusterdist algorithm of the ClusterScan tool (-d 10,000 parameter),[91] the zebrafish genome was scanned grouping together all the consecutive promoters closer than 10 kb. Then, for each group of promoters (clusters), the promoter density was calculated by dividing the number of promoters by the cluster length in kilobases. Finally, the rainfall plot representing the promoter density shown in Figure 1E was obtained using the karyoploteR R library.[92]

Processing of CAGE-seq data

CAGE-seq data set (accession SRA: SRA055273) from Nepal et al.,[40] were mapped with Bowtie v1.1.2[87] in default n-mode allowing 2 mismatches in the 27 bp seed region and reporting alignments of reads with up to 1 or 1000 valid alignments for unique ('-S -n 2 -l 27 -a –best -m 1') and multimapping ('-S -n 2 -l 27 -a –best -m 1000') respectively. The CAGEr package[88] v1.18 was used for downstream processing of the data, i.e., CAGE Transcription Start Site (CTSS) calling and clustering into tag and consensus clusters (TCs and CCs), interquantile width calculation.

Processing of ChIP-seq and ATAC-seq data

Prior to mapping ChIP-seq and ATAC-seq reads were trimmed from adapters and filtered for quality using TrimGalore with automatic adaptor detection option and quality threshold of 20. Only reads where both mates survived the filtering steps were kept in case of paired end data (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore).

The data were mapped to a reference genome (danRer11) using segemehl aligner (v0.3.4)[86] Allowing up to 2500 alignments per read and alignments accuracy of 90 % (command line parameters: ' segemehl.x -M 2500 -A 90'). Genome coverage was generated using STAR[89] with command line parameters: '–runMode inputAlignmentsFromBAM -outWigType bedGraph –outWigStrand Untranded –outWigNorm RPM'.

To compare the dynamics of the epigenetic features between the different gene sets (Figures 2A and 2B) and datasets, the distribution of average normalised signal relative to the transcription start site for each set and feature was visualised on aggregation plots with the Bioconductor SeqPlots package[93] providing the coverage file and genomic regions of features of interest as input with 20 bp bin size. For visualisation purposes a smoothing spline function was applied before plotting.

The transcription start sites (TSSs) were determined using the CAGE-seq dataset[40] at a developmental stage where there was sufficiently high signal to accurately determine the dominant TSSs for each gene in the set, or to minimize interference from the maternal contribution in the case of the constitutive gene set. For the *miR-430*, minor and major wave gene sets high, sphere and 30%-epiboly were used for dominant TSS detection, respectively. For both the constitutive and post-gastrulation sets prim-5 stage was used.

The publicly available ChIP-seq and ATAC-seq data used in this study were obtained from the Sequence Read Archive (SRA) deposited under following BioProject accessions: BioProject: PRJNA43421614 (H3K4me3 and H3K27me3 ChIP-seq); BioProject: PRJNA473799 (H3K27ac ChIP-seq); BioProject: PRJNA395463 (ATAC-seq); BioProject: PRJNA156233 (Nanog ChIP-seq); BioProject: PRJNA171706 (Pou5f1 and Sox2 ChIP-seq).

Promoter architecture analyses

To compare core promoter motif composition between gene set (minor wave, major wave, constitutive and postulation) promoter sequences were extracted and centre-aligned to the dominant TSS.

Position weight matrices (PWM) for each core promoter set was calculated from the sequences using the "DiffLogo" R package[94] and motif sequence logos were visualised with the "ggseqlogo" R package.[95]

Promoter architecture (sharp/broads), was determined for each gene in the above four datasets, using the 0.1–0.9 interquantile width (the length of the region in bp where 10%-90% of the CAGE signal is concentrated) of the promoter associated CAGE-tags consensus clusters. Promoters with interquantile width up to 10 bp were classified as sharp and with more than 10 bp as broad.

Transcription factor binding motif analyses

Enrichment for clusters of TF binding sites was done using Cluster-Buster[52] with the following parameters: 'cbust -r10000 -m 5 -c 10'. The optimal motif weights as abundances of the motifs (occurrences per kb) and the average distance between neighbouring motifs to use with Cluster-Buster was calculated using Cluster-Trainer, ran with the following parameters: 'ctrain -t 10 -r 10000'. As input, a 1.6Mb sequence region, extracted from chr4 of the guided genome assembly of the Nanopore sequencing, flanking the *miR-430*

cluster (500 kb up- and downstream) was used. Position frequency matrices for Nanog (UN0383.1), Pou5f1::Sox2 (MA0142.1) and Sox2 (MA0143.3) where obtained from JASPAR(2020) database[99,100] (http://jaspar.genereg.net)

Generation of stable transgenic zebrafish reporter lines

The construct *Xla.crygc:attP-Gal4vp16-14UAS:Clover* for generation of docking transgenic lines for usage with the PhiC31 integrase mediated targeted transgene integration system was generated by replacing the eGFP reporter form *pDB783:Xla.crygc:attP-Gal4vp16-14UAS:eGFP* (gift form Darius Balciunas, Temple University, PA) with the brighter green fluorescent protein Clover.

Docking lines were generated using the Tol2-mediated transgene integration system.[53,101,102] In detail 1-cell stage AB wildtype embryos with 1nl solution containing 20ng/μl plasmid and 20ng/μl Tol2 mRNA, supplemented with 0.1% phenol red as injection marker. Injected embryos were kept in E3 medium at 28°C and screen for lens expression (green fluorescence) between 72-96hpf. Positive embryos were grown to adulthood, outcrossed to wild type fish and germline transmitting founders (F0) were identified by screening the resulting F1 embryos for lens expression (green fluorescence) at 72-96hpf. Positive F1 embryos were grown to adulthood to establish stable transgenic lines. Three docking lines in total were established and two of them (*Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL1 and Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL3*) were used to generate *miR-430* promoter reporter lines (see below).

The *miR-430* promoter region fragment (~650 bp, Data S1) was PCR amplified from zebrafish genomic DNA using the following forward: actcacgtgtggtaccCTTAGTCACCTCTGCCCACCAAG and reverse: catcgataagcggccgcTCCGTTCATAGTCTTTAGCAGC AAG primers (lower case letters correspond to the vector homology arms, upper case letters are specific to the *miR-430* promoter regions). The fragment was then cloned into modified pJet:attB:mCherry vector[53] containing *phiC31 attB* site and *mCherry* reporter using In-Fusion HD Cloning Kit (TaKaRa/Clontech) according to the manufacture instructions. In brief, the pJet vector was digested with Acc65I and NotI restriction enzymes. The purified linearized vector and *miR-430* promoter fragments (in 1:3 molar ratio) were incubated with 1x In-Fusion HD cloning mix at 50°C for 15 min and 2 μl of the cloning reaction was transformed into the kit-provided Stellar competent cells. Positive colonies were identified by PCR and the isolated plasmid verified by restriction digestion and sequencing.

MiR-430 promoter reporter lines were generated by following the PhiC31 integrase mediated targeted transgene integration protocol as described in Hadzhiev et al.[54] One-cell stage docking line embryos were micro-injected with 1 nl solution containing the reporter plasmid (20 ng/μl) supplemented with PhiC31-nanos 3'UTR mRNA (30 ng/μl). The efficiency of integration events was determined by frequency of lens colour change (from green to red) in the injected embryos at 72-96 hpf. If lens colour change is observed in more than 5 % of the embryos in an injection batch all the embryos from the batch are grown to adulthood, including those without lens colour change, because integration events may have happened with higher frequency in the PGCs due to nanos 3'UTR driven localisation of the PhiC31 integrase mRNA occurring predominantly in the PGCs.

Identification of transgene integration sites

The integration site of the *Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL3* line has been previously identified (by nanopore sequencing) to be on chr17 ~7 kb upstream of *slc16a9a* and ~6kb downstream of *npy4r* (chr17:20640317-20646440, danRer11).

The integration site of the *Tg(Xla.crygc:attP-Gal4vp16, 14UAS:Clover)UoBL1* was identified by inverse PCR following the protocol byGreen and Sambrook.[103] Genomic DNA was prepared from about 100 4 days old transgenic embryos using PureLink™ Genomic DNA Mini Kit (Themo Fisher, K182002). 5 μg of genomic DNA was digested in three separate reactions with TaqI, a mix of BamHI/BclI/BglII and a mix of XbaI/SpeI/AvrII/NheI restriction enzymes for 2 h. The restriction digest reactions were purified by phenol-chloroform extraction. The digest DNA was ligated overnight at 16°C at final concentration of 500pg/μl using 5U of T4 DNA ligase in a total volume of 100 μl. Amplification of the genomic flanking sequences was achieved by performing two nested PCR reactions in total volume of 50 μl. For all enzyme mixes, the flanking sequence of the 3′ end of the transgene was amplified with primer pair 3PinvP_F1/3PinvP_R1 in the first PCR reaction using 3 μl of the ligation mixes as template, followed by the second PCR reaction with 3PinvP_F2/3PinvP_R2 using 1 μl from the first PCR reaction as template.

Similarly, for TaqI and BamHI/BclI/BglII mix the flanking sequence of the 3′ end of the transgene was amplified using the primer pairs 5PinvP_F1/5PinvP_R1 and 5PinvP_F2/5PinvP_R2 in the first PCR and second PCR reaction respectively. The 5' end transgene flank from the XbaI/SpeI/AvrII/NheI digestion mix was amplified using the primer pairs 5PinvP_F3/5PinvP_R1 and 5PinvP_F4/5PinvP_R2 in the first PCR and second PCR reaction respectively.

The products of the second PCR reactions were purified using AMPure XP beads and sequenced. The sequencing result of all six PCR products (for both 5' and 3' ends and 3 different restriction enzyme combinations for each end) confirmed the integration site to be on chr22 into the 5'UTR of the *ccdc50* gene (chr22:24318507-24318514, danRer11). All primers sequences are provided in Table S4.

Reporter expression analyses of *miR-430* promoter stable transgenic lines by RT-PCR

Total RNA was extracted from ~50 embryos at the relevant developmental stage, using miRNeasy mini kit (QIAGEN, 217004) following kit instructions. On-column DNAse treatment was performed using RNAse-free DNAse Set (QIAGEN, 79254). Reverse transcription was carried out using 1 μg of total RNA, 0.5 μg of random hexamers (Promega, C1181) and 200U M-MLV reverse transcriptase (Promega, M170A) in 25 μl reaction volume. PCR reactions were performed with GoTaq® Hot Start Master Mix (Promega, M5122) using 1 μl of the reverse transcription reaction as template in 20 μl reaction volume. Thirty amplification cycles were carried

out in total and PCR reactions were analysed on agarose gel electrophoresis. The following primer pairs were used: miR430_Prom_FP: AGCAGACAACAAGATGCGTGTG with Gal4VP16_RP: ACTTCGGTTTTTCTTTGGAGCAC and γ-crystalline_Prom_FP GAAACTTCCACTCAGTCAGACTTGC with mCherry_RP: CACCTTGAAGCGCATGAACTC.

Injections at single cell stage for ubiquitous delivery vs late-stage injection for mosaic treatments

For dCas9 and Cas9 targeting, a 1:1 ratio mix of two guide RNAs targeting the *miR-430* repeats or one guide targeting the gol promoter was used. Injection mixes for CRISPR mediated manipulations consisted of 650 ng/ul dCas9 protein (NEB, M0652) or Cas9 protein (NEB, M0646), and 200 ng/ul of guide RNA. For visualisation of CRISPRi knockdown reagents 400 ng/ul dCas9-GFP protein (Novateinbio, PR-137213G) and 200 ng/ul of guide RNA was used. These injection mixes were preincubated together at 37°C for 5 min to facilitate complex formation, before supplementation with 0.1% phenol red and injection. Ubiquitous delivery of knock-down and labelling reagents was achieved by 1 nl injection of embryos at the one-cell stage. Mosaic knock-down was produced by 1 nl injection into a single cell at the 8-cell stage.

Nascent RNA capture and sequencing upon dCas9 inhibition of miR-430 transcription

To capture nascent RNAs, The Click-iT® Nascent RNA Capture Kit (ThermoFisher Scientific, C10365) was used. Embryos were injected with CRISPRi reagents (as described above) and ∼30-40 pmols of Click-iT® EU (5-ethynyl Uridine) at one-cell stage. At 512-cell, total RNA from 20 embryos of each treatment and control groups was extracted using TRIzol reagent (Invitrogen). The isolated RNA was biotinylated and captured on streptavidin coated magnetic beads, following the manufacturer's instructions of the Click-iT® Nascent RNA Capture Kit. The beads were resuspended in 5 μl Click-iT wash buffer2 and used directly as input for cDNA synthesis and stranded RNA-seq library preparation with the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (E7760S). The RNA-seq libraries (two biological replica for each treatment and control groups) were sequenced (2x100bp, paired-end) on Illumina NovaSeq 6000 platform.

Analysis of the nascent (EU) RNA-seq data

Reads were mapped to the zebrafish reference genome (danRer11) using segemehl (v0.3.4) in split(splice) mode, allowing up 2500 multimappers: 'segemehl.x -M 2500 -S'

Htseq-count (v 0.12.4) was used to count reads (including multimappers) associated with a gene feature (ENSEMBL 95 gene annotations):'htseq-count –a=0 –nonunique all –-stranded=reverse'

DESeq2 Bioconductor package[90] was used to perform differential expression analyses. Low expressed genes (<5 read counts) and very highly expressed 5S rRNA, and some mitochondrial genes (>50000 read counts) were excluded from the analysis. The linear model was adjusted to treat the two replicas pairwise, e.g., '∼replicate + condition'. The DESeq2 output of the differential expression analysis is provided Table S6.

Gene ontology enrichment analysis of overexpressed genes was performed using Panther gene ontology web interface (http://geneontology.org/). The gene ontology analysis output is provided in Table S7.

Global transcription block with α-amanitin and triptolide treatment

Global transcription block was performed either by treatment with 2 μM triptolide from the one-cell stage (Sigma T3652) in E3 media, or by microinjection of 200 pg α-amanitin (Sigma, A2263) at single-cell stage. As an injection control for α-amanitin, 10 % DMSO in water was used.

PCNA staging

mApple:PCNA fusion protein (400 pg) was injected at the one-cell stage to enable ubiquitous labelling of all daughter nuclei for the selection of synchronous embryos. At the cell cycle of interest embryos were selected for ubiquitous, bright, homogenous nuclear staining, indicative of nuclear-wide replication. These late interphase embryos were immediately snap frozen for RNA extraction or fixed with cold 4% PFA for in situ hybridization.

miR-430 MO labelling

Single-cell stage embryos were injected with 1 nl solutions containing Cy5-labelled *miR-430* targeting morpholinos (Gene Tools LLC)[22] at a concentration of 7 μM each in nuclease-free water and 0.1% phenol red.

EU labelling of nascent RNA

Single-cell stage zebrafish embryos were injected with 1 nl of 50 mM ethynyl-uridine (EU, Thermo Fisher, C10329) until fixation with 4% PFA at the stage of interest. Embryos were dehydrated and rehydrated using MeOH gradients then fluorescently labelled using the Click-iT™ RNA Alexa Fluor™ 488 Imaging Kit (Thermo Fisher, C10329), following the manufacturer's protocol. Labelled embryos were used for subsequent protein RNA, or DNA detection.

3D DNA-FISH (*miR-430* or klf17)

BACs DKEY-69C19 and CH1023-918E12 were used as templates for DNA-FISH probe production, corresponding to 214kb of the *miR-430* locus on chr4 and 38kb of the *klf17* locus on chr2 respectively. The FISH Tag™ DNA Multicolor Kit (ThermoFisher, F32951)

was used to label digested BAC fragments with Alexa dyes following kit instructions, with a 60 min digestion per μg of DKEY-69C19 and 10 min per μg of CH1023-918E12. 50–100 ng of probe was used for each sample in with hybridization buffer [50% formamide, 4x SSC, 100mM NaPO4 pH 7.0, 0.1% Tween-20].

Embryos were fixed at the developmental stage of interest using 4% PFA. Isolated animal caps were equilibrated with hybridization buffer, through a series of gradient washes. Embryonic DNA was then denatured at by 85°C for 10 min before immediate application of probe and incubation overnight at 37°C. A hybridization buffer: PBS-T gradient of washes was used to remove unbound probe from samples before counterstaining with DAPI. Samples were imaged on a Zeiss LSM 880 with FastAiryscan at maximum resolution, with a 63 ×1.40 numerical aperture objective lens. 20–50 optical sections (500 nm slice thickness) were acquired of each nuclei, and resulting images were processed using Zen Black software (Zeiss).

Expression analysis following miR-430 manipulations

Total RNA was extracted from >30 PCNA-synchronised 512-cell or dome stage embryos, and 500ng of total RNA used to produce equal amounts of total RNA, as detailed above. At least three independent experimental samples were collected for each manipulation condition.

For all manipulation experiments performed in WT embryos, 10 μl RT-PCR reactions were prepared with 10 ng of cDNA with 500 nM of each forward and reverse gene-specific primers and PowerUp SYBR Green MasterMix (Applied bio systems, A25742) on a QuantStudioV (Thermo Fisher Scientific) machine. For manipulation experiments performed in the single *miR-430* promoter transgenic embryos, 40ng of cDNA was used for *gal4* and *TBP* quantification, or 10ng cDNA for *miR-430* and 5S rRNA, both in 10 μl reactions. Thermo Fisher Connect software was used for data acquisition and PRISM for statistical analysis. -RT controls gave CT values either >7 cycles greater than +RT samples or equivalent to alpha-amanitin treated samples representing complete block of transcription.

Guide RNA production

Double stranded DNA templates for sgRNA guides were produced through PCR annealing and extension of 100 ng common guide RNA backbone oligo and 100 ng target specific sgRNAR oligo (Merck) using Phusion Hot-start II DNA Polymerase (ThermoScientific, F549) with 35 cycles and annealing at 60 °C. Resultant DNA was gel extracted then 400 ng template used for *in vitro* transcription with the HiScribe RNA synthesis T7 kit (NEB, E2040) for 2 h at 37 °C, then TurboDNase (Ambion, AM2238) treated for 15 min at 37 °C. Protein and non-incorporated nucleotides were removed using the Monarch RNA Clean up kit (NEB, T2040), before quantification by nanodrop.

In situ probe production

Templates for *in situ* probes were produced by amplification from 100 ng of 1k cell stage cDNA using gene-specific primers with promoter overhangs and Phusion Hot-start II DNA Polymerase (ThermoScientific, F549). Resultant templates were gel extracted then 500 ng transcribed into RNA using T7 RNA polymerase (ThermoScientific, EP0111) and either dig- or FITC-labelling mix (Roche, 3935420 and #32874620 respectively). RNA probes were purified using Microspin G25 (GE Healthcare, 27-5325-01) columns then final RNA concentration determined by nanodrop. 30 ng of pre-exhausted probe in Hybe+ was used per 50 embryo sample, for RNA-ISH and-FISH.

Live embryo imaging

The Zeiss Lightsheet Z1 was used to image live embryos mounted in 1% low melt agarose within size three glass capillaries, incubated at 28 °C in E3 media during imaging. Z-stacks of ~100–200 slices in 1 μm steps were acquired every 30–60 seconds for 1.5–2.5 h, with the ×20 objective.

Whole mount antibody staining

Embryos were fixed at the 512-cell stage in 4% PFA then washed with PBS-0.3% TritonX-100 to permeabilize. 10% goat serum in PBST was used to block embryos before addition of pre-blocked Pol II S2P monoclonal antibody (Diagenode, C15200005) at 1:1000 dilution. PBS-0.1% Tween-20 washes removed excess primary antibody before application of pre-blocked Alexa 633 goat anti-mouse IgG (Life technologies, A21052) secondary antibody at 1:2000 dilution. Embryos were mounted in antifade mounting media with DAPI (Vectashield, H-1200, Vector Laboratories) and imaged with a Zeiss 880 confocal microscope with Fast Airyscan Module with a 63 ×1.40 numerical aperture objective lens and the resulting images were processed using Zen Black software (Zeiss).

Whole mount in situ hybridization

Embryos were fixed in 4% PFA at the developmental stage of interest, then washed with PBST before manual dechorionation and isolation of animal caps. Samples were dehydrated then rehydrated using a MeOH gradient before being equilibrated with hybridization buffer [50% formamide, 1.3x SSC, 5mM EDTA pH 8.0, 0.2% Tween-20]. Embryos were incubated with probe overnight at 58-65°C then unincorporated probe removed by a series of Hybe:SSCT and SSCT:PBST washes. Samples were then blocked in 10% goat serum in PBST before application of pre-blocked anti-dig-AP antibodies at a 1:2000 dilution in block. Excess antibody was removed with dig wash buffer (Roche, 11585762001) before staining with NBT and BCIP in 100mM NaCl, 100mM Tris.HCl pH9.5, 50mM $MgCl_2$ & 0.1% Tween-20.

Samples were imaged on a Zeiss AxioZoom.V16, with an ApoZ 1.5x/0.37 numerical aperture objective lens and the resulting images were processed using Zen Black software (Zeiss) and ImageJ.

Single-molecule fluorescent double in situ hybridization

Target-specific-probes were designed using the Oligostan script developed byTsanov et al.[104] RNA labelling protocol was based on smiFISH method for Drosophila byGarcia et al.[105] Briefly, FLAP oligos (Integrated DNA technologies) were directly labelled at both their 3' and 5' termini, with either FITC for FLAP-Y, or Cy5 for FLAP-X. FLAP oligos were hybridised onto the target-specific-probes (FLAP-X for *miR-430*, FLAP-Y for klf17) at equimolar ratio in 1x NEB buffer 3, by incubating at 85°C for 3 min, then 65°C for 3 min. Embryos were fixed in 4% PFA, manually dechorionated to isolate animal caps, which were then dehydrated and rehydrated using a MeOH gradient. Samples were equilibrated with 15% formamide in 1xSSC before hybridization in FLAP-hybridized probe mix [0.83uM of each probe, 200ng/ul BSA, 2mM Ribonucleoside Vanadyl Complex (Sigma-Aldrich, 102207064), 15% formamide, 2x SSC, 10% dextran sulphate (Fisher, BP1585), 0.5ug/ul E.coli tRNA] overnight at 37°C. Unincorporated probes were removed by a series of PBST washes before counterstaining and mounting with DAPI/DABCO (Vectashield, H-1200, Vector Laboratories). Samples were imaged with a Zeiss 880 confocal microscope with Fast Airyscan Module with a 63 ×1.40 numerical aperture objective lens and the resulting images were processed using Zen Black software (Zeiss). All primers used in this study are provided in Table S4.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the expression analysis by qPCR, following miR-430 manipulations (Figures 5A, 6C, 6D, 6F, 6G, S4I–S4L, S4I–S4L, S6D, and S6E), the ΔΔCt values from independent biological repeats were evaluated by the Shapiro-Wilk test for normal distribution. The p-values were determined by unpaired t-test.

Quantification and differential expression analyses of nascent RNAseq data (Figure S5C) was performed with DESeq2 R package which uses Wald test for null hypothesis testing and Benjamini-Hochberg test for FDR/adjusted p-values calculation

GO enrichment of differentially expressed genes (Figure S5B) was carried out with Panther GO web tool using Fisher's Exact test.