Targeted disruption of transcription bodies causes widespread activation of transcription

Martino Ugolini[1,2], Ksenia Kuznetsova[2], Haruka Oda[3], Hiroshi Kimura[3], Nadine L. Vastenhouw[1,2*]

[1]Université de Lausanne (UNIL), Center for Integrative Genomics (CIG); Lausanne, Switzerland.
[2]Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG); Dresden, Germany.
[3]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology; Yokohama 226-8503, Japan.
[*]Corresponding author: nadine.vastenhouw@unil.ch

ABSTRACT

The localization of transcriptional activity in specialized transcription bodies is a hallmark of gene expression in eukaryotic cells, but it remains unclear if and how they affect gene expression. Here, we disrupted the formation of two prominent transcription bodies that mark the onset of zygotic transcription in zebrafish embryos and analyzed the effect on gene expression using enriched SLAM-Seq and live-cell imaging. We found that the disruption of transcription bodies results in the upregulation of genes that already have accessible chromatin and are poised to be transcribed when the two transcription bodies are present. Live-cell imaging suggests that the disruption of these transcription bodies enables poised genes to be transcribed in ectopic transcription bodies. Our results suggest that transcription bodies regulate transcription genome-wide by sequestering components of the transcriptional machinery.

RNA polymerase II (RNAPII) and transcription factors are often concentrated in specialized transcription bodies (*1–10*), and it has been suggested that these can bring together multiple genes (*11–14*). This has led to the hypothesis that transcription bodies increase the efficiency of transcription by promoting the biomolecular interactions underlying gene expression (*15–17*). Analysis of gene expression in artificially induced condensates supports this model (*18*). To understand how transcription bodies affect gene expression and development *in vivo*, it is necessary to analyze endogenous transcription bodies. This has been difficult because transcription bodies are often small, short-lived, and numerous. Here, we use the onset of transcription in zebrafish embryos to overcome this problem. In zebrafish embryos, development is initially driven by maternally loaded RNA and protein (Fig. 1A) (*19*). Transcription gradually begins via a process called zygotic genome activation (ZGA). Before transcriptional activity can be seen throughout the nucleus, it is confined to two micron-sized transcription bodies (Fig. 1B) (*9*, *20–23*), nucleated by the *mir430* locus, which contains multiple copies of the *mir430* gene (*21*, *24*). These bodies are enriched for RNAPII Serine 5 and Serine 2 phosphorylation (Ser5P, Ser2P), which mark the initiating and elongating form of RNAPII, respectively, as well as the transcription factors Nanog and Sox19b (*9*, *20–22*).

To investigate the role of transcription bodies in development, we generated a fish line in which the *mir430* locus is deleted (*9*), resulting in the specific disruption of the two *mir430* transcription bodies, as observed by the loss of Ser2P signal (Fig. 1C). *miR430* is required for downregulating maternally loaded transcripts, and its absence results in severe developmental phenotypes (*24*, *25*). Thus, effects on gene expression and development in the *mir430* mutant may be caused by the loss of the *mir430* transcription bodies, or the loss of the mature microRNA MiR430. To uncouple these two components and allow us to focus on the first, we supplemented mutant embryos with the mature form of microRNA MiR430 (Fig. 1D) (*24*), which restored degradation of mRNAs containing MiR430 binding sites (Fig. 1E, Fig. S1). This rescued the delay in epiboly (Fig. 1F,G) and the aberrant expression pattern of *goosecoid* (Fig. S2), which are characteristic for the loss of MiR430 (*24*, *25*). Embryos supplemented with mature microRNAs appeared normal, with the appearance of the Kupffer's vesicle (Fig. 1H), and overall developmental progression (Fig. 1I, Fig. S3) being indistinguishable from wild type (WT) and heterozygous siblings, and embryos

developed into fertile fish. Thus, the disruption of two large transcription bodies does not result in obvious developmental defects.

To assess the effect of the specific disruption of *mir430* transcription bodies on transcriptional activity, we developed eSLAM-Seq, which uses a combination of protocols to label, enrich, and detect nascent transcripts (Fig. S4; Materials and Methods), because the large amounts of maternally loaded RNA in early embryos (Fig. 1A) masks nascent RNAs in total RNA sequencing approaches. We used eSLAM-Seq to compare gene expression between WT and *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage, when transcriptional activity is largely restricted to the *mir430* transcription bodies in WT embryos (Fig. 1B). We observed a significant number of genes that were downregulated in the absence of the *mir430* transcription bodies (242) and a higher number of genes that were upregulated (716) (Fig. 2A,B; Table S1). Plotting the genomic distribution of genes that were up- and downregulated showed that both classes of genes are distributed across the genome (Fig. 2C). We conclude that the absence of two large transcription bodies has a widespread effect on transcription.

The identification of hundreds of up- and downregulated genes in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT embryos at the 256-cell stage contrasts with the observation that these embryos develop normally (Fig. 1). To reconcile these two observations, we generated eSLAM-Seq data for WT and *mir430*$^{-/-}$ + inj MiR430 embryos at the 1024-cell, Oblong, and Sphere stage and compared it with the data obtained at the 256-cell stage. The expression levels of genes that are misregulated in *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage (with a median of 3.5-fold difference with WT for the upregulated genes, and 2.9-fold for the downregulated genes; Fig. 2D) gradually recover during consecutive developmental stages, reaching WT expression levels at Sphere stage (with a median of 0.03-fold difference with WT for the upregulated genes, and 0.07-fold for the downregulated genes; Fig. 2D). Looking at all genes, we find that overall fewer genes are misregulated at later stages, from 716 up and 242 down at 256-cell to 51 up and 30 down at Sphere stage (Fig. S5). We conclude that in the absence of *mir430* transcription bodies, gene expression is misregulated, but that this effect is short-lived as expression levels return to WT levels in a timeframe of a couple of cell cycles, likely explaining the lack of developmental defects in *mir430*$^{-/-}$ + inj MiR430 embryos.

We then further analyzed the genes that are misregulated at the 256-cell stage. The downregulation of genes suggests that transcription bodies facilitate transcription of genes other than *mir430* itself, consistent with work in which transcribed genes were shown to colocalize (*12–14*, *26*), as well as the proposed role of transcription bodies in increasing transcriptional efficiency (*18*, *27*, *28*). We hypothesize that the genes that are downregulated in *mir430*$^{-/-}$ + inj MiR430 embryos localize to the *mir430* transcription bodies in WT embryos and are positively affected by this localization. The observation that downregulated genes return to WT expression levels over time suggests that transcription bodies that form during later stages in *mir430*$^{-/-}$ + inj MiR430 embryos (Fig. S6) may compensate for the loss of *mir430* transcription bodies.

We focus on the upregulated genes. Their identification and high number led us to hypothesize that the *mir430*-driven transcription bodies may sequester components of the transcriptional machinery, preventing genes elsewhere in the nucleus from being transcribed. To investigate this possibility, we visualized transcription elongation in WT and *mir430*$^{-/-}$ + inj MiR430 embryos. We injected 1-cell stage embryos with an antigen-binding fragment (Fab) recognizing RNAPII Ser2P (transcription elongation), as described before (*9*, *22*, *23*) and imaged them. We also visualized *miR430* transcripts using Morpholino Visualization of Expression (MOVIE, (*21*)), to be able to distinguish the transcription bodies that are nucleated by the *mir430* locus from other transcription bodies. At the 128-cell stage, we detected two prominent *mir430* transcription bodies in nuclei of WT embryos, whereas in nuclei of *mir430*$^{-/-}$ + inj MiR430 embryos we saw neither RNAPII-Ser2P nor *miR430* MOVIE signal (Fig. 3A and Fig. S7). In *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage, in contrast, we observed RNAPII-Ser2P bodies (Fig 3A, Fig. S8). Because they lacked *miR430* MOVIE signal and we do not detect them in WT embryos, we consider them to be ectopic transcription bodies. They are small and limited in number at the 256-cell stage, however, they are more prominent at the 512- and 1024-cell stage (Fig. 3A and Fig. S9,10). The quantification of ectopic transcription bodies at all stages confirmed these results (Fig. 3B). We conclude that the loss of *mir430* transcription bodies results in the appearance of ectopic transcription bodies.

To identify component(s) of the transcriptional machinery potentially sequestered by the *mir430* transcription bodies, we characterized the genes that are upregulated in response to *mir430* transcription body removal. We asked when these genes, which are expressed at the 256-cell stage

in *mir430*$^{-/-}$ + inj MiR430 but not WT embryos, are induced in WT embryos, by analyzing their expression at the 1024-cell, Oblong, and Sphere stages, and comparing this to their expression at the 256-cell stage. Most of the genes (81%) are activated during genome activation in WT embryos (Fig. 4A), indicating that loss of the *mir430* transcription bodies prematurely activates the expression of genes that are poised to be activated, with 62% (446/716) being advanced by 30 minutes (at the 1024-cell stage in WT, Fig. 4A). There is a good correlation between the expression level of these genes in *mir430*$^{-/-}$ + inj MiR430 embryos at the 256-cell stage and WT embryos at the 1024-cell stage (Fig. 4B, Fig. S11).

To investigate the chromatin accessibility state of the genes that are upregulated in *mir430*$^{-/-}$ + inj MiR430 embryos, we analyzed an ATAC-Seq dataset we had previously generated in WT embryos (*29*). This revealed that the transcription start sites of genes that are upregulated in the *mir430*$^{-/-}$ + inj MiR430 are highly accessible in WT embryos (Fig. 4C), being as accessible as those of the most highly expressed genes (Fig. 4C,D), even though they are hardly expressed (Fig. 4E). This suggests that the sequestered factor(s) may function downstream of transcription-factor binding and chromatin opening. Consistent with this, motif analysis of the promoters of upregulated genes did not identify transcription factor motifs that were both sufficiently enriched compared to unexpressed genes and present in a sufficiently high fraction of the upregulated genes to explain our observations (Fig. S12). Even motifs of Nanog, Pou5f3 and Sox19b, transcription factors that have been shown to be required for the activation of transcription during ZGA in zebrafish (*9*, *29–32*) did not fulfill these two requirements. Thus, removal of the *mir430* transcription bodies advances expression onset of genes poised to be activated.

This reminded us of an observation that we had made previously: in addition to the *mir430* transcription bodies, where transcription elongation takes place, WT embryos from the 256-cell stage onward display transcription bodies that are positive for transcription initiation, but not elongation (*9*). We hypothesized that *mir430* transcription bodies might sequester a pause release factor, causing gene expression in these initiation bodies to be arrested at the initiation phase. If so, in nuclei of *mir430*$^{-/-}$ + inj MiR430 embryos, these initiation bodies would acquire elongation signals. To investigate this idea, we injected Fabs against RNAPII Ser5P and Ser2P into WT and *mir430*$^{-/-}$ + inj MiR430 embryos. Consistent with our hypothesis, we saw more Ser2P-positive

transcription bodies in *mir430*$^{-/-}$ + inj MiR430 than in WT embryos during the 512- and 1024-cell stages (Fig. 4F, and Fig. S13,14). Quantification of initiation-only, initiation/elongation, and elongation-only transcription bodies confirmed that more transcription bodies go into productive elongation in *mir430*$^{-/-}$ + inj MiR430 than in WT embryos (with a maximum of 15% elongation-positive transcription bodies in WT versus 33% in *mir430*$^{-/-}$ + inj MiR430 at the 256-cell stage, and a comparable difference at the 1024-cell stage; Fig. 4G). Thus, loss of the *mir430* transcription bodies causes genes that are arrested in the transcription initiation phase to transition into the transcription elongation phase, explaining both the appearance of ectopic elongation-positive transcription bodies (Fig. 3B) and premature gene expression onset in *mir430*$^{-/-}$ + inj MiR430 embryos (Fig. 2B).

We show that removal of the *mir430* transcription bodies results in many genes being downregulated, consistent with a model where the bodies facilitate transcription by promoting the biomolecular interactions underlying gene expression (*3*, *16*, *18*). We also find that a significant number of genes are upregulated upon disruption of the *mir430* transcription bodies, with the transcription bodies likely sequestering a limiting component(s) of the transcriptional machinery, preventing genes that are not present in the transcription bodies from being transcribed (Fig. S15). Factor sequestration regulating biological processes has been described before (*33*, *34*) and a recent study showed that the loss of the transcription factor Zelda in *Drosophila* embryos caused RNAPII to relocate from Zelda-containing transcription bodies to the histone locus body (*35*). The factor(s) sequestered in *mir430* transcription bodies likely promotes pause release. In *mir430*$^{-/-}$ + inj MiR430 nuclei, disruption of the *mir430* transcription bodies would result in the release of this factor(s), allowing poised RNAPII to transition to elongation (Fig. S15). Loss of the *mir430* transcription bodies does not significantly impact development, likely because gene expression levels in the *mir430*$^{-/-}$ + inj MiR430 are restored to WT levels within a couple of cell cycles, implying embryos are tolerant to a temporal misregulation of gene expression during pre-gastrulation development. The observation that downregulated genes return to WT expression levels over time suggests that transcription bodies formed during later stages compensate for the loss of *mir430* transcription bodies, implying that transcription bodies form in a non-deterministic manner.

References

1. D. A. Jackson, A. B. Hassan, R. J. Errington, P. R. Cook, *EMBO J.* **12**, 1059–65 (1993).
2. A. Ghamari *et al.*, *Genes Dev.* **27**, 767–777 (2013).
3. A. Tsai *et al.*, *Elife*. **6**, 1–18 (2017).
4. W. K. Cho *et al.*, *Science*. **361**, 412–415 (2018).
5. B. R. Sabari *et al.*, *Science*. **361**, eaar3958 (2018).
6. S. Chong *et al.*, *Science*. **361**, eaar2555 (2018).
7. A. Boija *et al.*, *Cell*. **175**, 1842–1855 (2018).
8. A. V. Zamudio *et al.*, *Mol. Cell*. **76**, 753–766 (2019).
9. K. Kuznetsova *et al.*, *bioRxiv*, in press, doi:10.1101/2022.06.13.495463.
10. I. I. Cisse *et al.*, *Science*. **341**, 664–667 (2013).
11. F. J. Iborra, A. Pombo, D. A. Jackson, P. R. Cook, *J. Cell Sci.* **109**, 1427–1436 (1996).
12. C. S. Osborne *et al.*, *Nat. Genet.* **36**, 1065–1071 (2004).
13. S. Schoenfelder *et al.*, *Nat. Genet.* **42**, 53–61 (2010).
14. S. Fanucchi, Y. Shibayama, S. Burd, M. S. Weinberg, *Cell*. **155**, 606–620 (2013).
15. H. Sutherland, W. A. Bickmore, *Nat. Rev. Genet.* **10**, 457–466 (2009).
16. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, *Cell*. **169**, 13–23 (2017).
17. A. Tsai, R. Galupa, J. Crocker, *Development*. **147**, dev161430 (2020).
18. M. T. Wei *et al.*, *Nat. Cell Biol.* **22**, 1187–1196 (2020).
19. N. L. Vastenhouw, W. X. Cao, H. D. Lipshitz, *Dev.* **146**, dev161471 (2019).
20. S. H. Chan *et al.*, *Dev. Cell*. **49**, 867–881 (2019).
21. Y. Hadzhiev *et al.*, *Nat. Commun.* **10**, 691 (2019).
22. Y. Sato *et al.*, *Dev.* **146**, dev179127 (2019).
23. L. Hilbert *et al.*, *Nat. Commun.* **12**, 1360 (2021).
24. A. J. Giraldez *et al.*, *Science*. **308**, 833–838 (2005).
25. Y. Liu *et al.*, *Front. Genet.* **11**, 853 (2020).
26. K. Monahan, A. Horta, S. Lomvardas, *Nature*. **565**, 448–453 (2019).
27. L. Ma *et al.*, *Mol. Cell*. **81**, 1682–1697 (2021).
28. B. R. Sabari, *Dev. Cell*. **55**, 84–96 (2020).
29. M. Pálfy, G. Schulze, E. Valen, N. L. Vastenhouw, *PLoS Genet.* **16**, e1008546 (2020).

30. M. T. Lee *et al.*, *Nature*. **503**, 360–364 (2013).
31. M. Leichsenring, J. Maes, R. Mo̧ssner, W. Driever, D. Onichtchouk, *Science*. **341**, 1005–1009 (2013).
32. L. Miao *et al.*, *Mol. Cell*. **82**, 986–1002 (2022).
33. A. Klosin *et al.*, *Science*. **367**, 464–468 (2020).
34. Y. Harigaya *et al.*, *Nat. 2006 4427098*. **442**, 45–50 (2006).
35. S. K. Huang, P. H. Whitney, S. Dutta, S. Y. Shvartsman, C. A. Rushlow, *Curr. Biol.* **31**, 5102–5110 (2021).
36. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, *Dev. Dyn.* **203**, 253–310 (1995).
37. G. E. Truett *et al.*, *Biotechniques*. **29**, 52–54 (2000).
38. Y. Hayashi-Takanaka *et al.*, *Nucleic Acids Res.* **39**, 6475–6488 (2011).
39. T. J. Stasevich *et al.*, *Nature*. **516**, 272–275 (2014).
40. H. Kimura, K. Yamagata, *Methods Mol. Biol.* **1222**, 127–147 (2015).
41. C. Thisse, B. Thisse, *Nat. Protoc.* **3**, 59–69 (2008).
42. P. Heyn *et al.*, *Cell Rep.* **6**, 285–292 (2014).
43. V. A. Herzog *et al.*, *Nat. Methods*. **14**, 1198–1204 (2017).
44. K. Holler *et al.*, *Nat. Commun.* **12**, 3358 (2021).
45. P. Bhat *et al.*, *bioRxiv* (2022), doi:10.1101/2022.06.01.494399.
46. J. A. Schofield, E. E. Duffy, L. Kiefer, M. C. Sullivan, M. D. Simon, *Nat. Methods*. **15**, 221–225 (2018).
47. B. Reichholf *et al.*, *Mol. Cell*. **75**, 756–768 (2019).
48. E. E. Duffy *et al.*, *Mol. Cell*. **59**, 858–866 (2015).
49. E. E. Duffy, D. Canzio, T. Maniatis, M. D. Simon, *Nucleic Acids Res.* **46**, 6996–7005 (2018).
50. W. Sun, W. Chen, in *Methods in molecular biology* (Humana Press, New York, NY, 2018), vol. 1720, pp. 25–34.
51. A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics*. **30**, 2114–2120 (2014).
52. Y. Zhang, C. Park, C. Bennett, M. Thornton, D. Kim, *Genome Res.* **31**, 1290–1295 (2021).
53. H. Li *et al.*, *Bioinformatics*. **25**, 2078–2079 (2009).

54. Broad Institute, Picard (2022), (available at http://broadinstitute.github.io/picard/).
55. E. Boileau, J. Altmüller, I. S. Naarmann-de Vries, C. Dieterich, *Brief. Bioinform.* **22**, bbab219 (2021).
56. Y. Liao, G. K. Smyth, W. Shi, *Bioinformatics*. **30**, 923–930 (2014).
57. M. I. Love, W. Huber, S. Anders, *Genome Biol.* **15**, 1–21 (2014).
58. T. L. Bailey, C. E. Grant, *bioRxiv*, in press, doi:10.1101/2021.08.23.457422.
59. F. Ramírez *et al.*, *Nucleic Acids Res.* **44**, W160–W165 (2016).
60. R Core Team, R: A Language and Environment for Statistical Computing (2014), (available at http://www.r-project.org/).
61. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer International Publishing, New York, 2016).
62. T. Yin, D. Cook, M. Lawrence, *Genome Biol.* **13**, 1–14 (2012).
63. L. Herrgen, C. Schröter, L. Bajard, A. C. Oates, *Methods Mol. Biol.* **546**, 243–254 (2009).
64. T. Boothe *et al.*, *Elife*. **6**, e27240 (2017).
65. J. Schindelin *et al.*, *Nat. Methods*. **9**, 676–682 (2012).
66. E. Katrukha, ComDet (2022), (available at https://github.com/ekatrukha/ComDet).

Acknowledgements

We thank the members of the Vastenhouw lab for technical support, helpful feedback, and stimulating discussions. We thank the Pauli and Ameres labs for technical advice on SLAM-Seq. We thank Maciej Kerlin, Aleksandar Vještica, and Edlyn Wu for comments on the manuscript, and the following facilities and services for their support: MPI-CBG - fish facility, light microscopy facility, and scientific computing, UNIL - genomic technologies facility, cellular imaging facility and the bioinformatics competence center. **Funding:** M.U. was supported by the Boehringer Ingelheim Fonds. Research in N.L.V's laboratory was supported by the Max Planck Society, the University of Lausanne, a European Research Council Consolidator Grant (101003023), the Volkswagen Foundation (94773), and the German Research Foundation (VA 1209/2-1). Research in H.K.'s laboratory was supported by the Japan Society for the Promotion of Science KAKENHI (JP18H05527, JP21H04764 and JP20K06484). **Author contribution:** Conceptualization: M.U., N.V. Methodology: M.U., K.K., H.U. Investigation: M.U., K.K. Data analysis: M.U. Writing – original draft: M.U. and N.V. Writing – reviewing & editing: all authors.

Funding acquisition: M.U., N.V., H.K. Supervision: N.V. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** Raw sequencing data will be uploaded on Gene Expression Omnibus (GEO) prior to publication. Raw imaging data are available upon request. All remaining data are available in the main text or the supplementary materials.

Supplementary Materials

Materials and Methods

Figs. S1 to S16

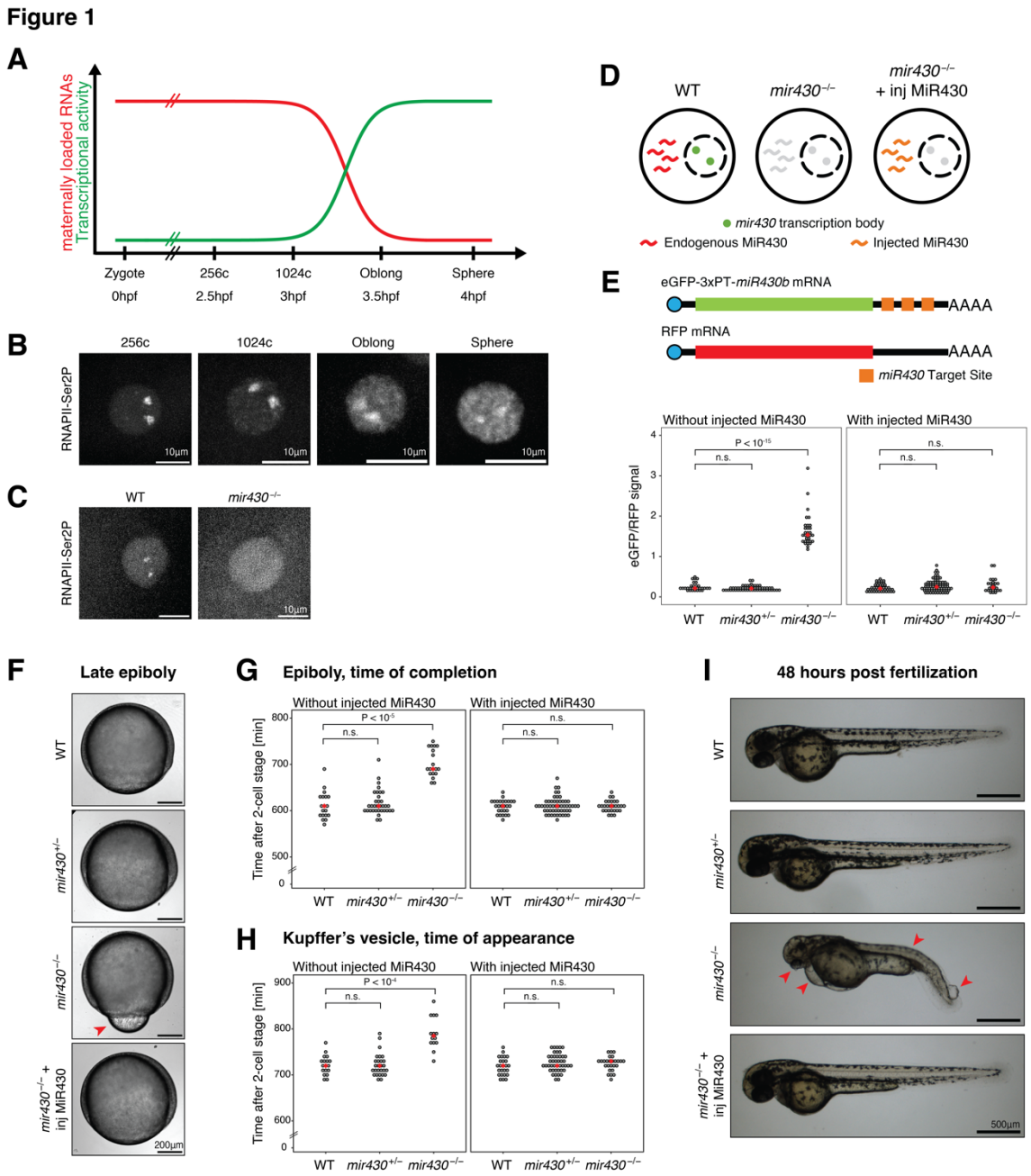

Fig. 1. Disruption of *mir430* transcription bodies does not substantially impact development. **A.** Schematic representation of the maternal-to-zygotic transition in zebrafish embryos. Hpf, hours post fertilization. **B.** Visualization of elongating RNAPII (Ser2P) with antigen-binding fragments (Fabs) in WT embryos. Shown are representative micrographs of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. **C.** Visualization of elongating RNAPII (Ser2P) with antigen-binding fragments (Fabs) in WT and *mir430*$^{-/-}$ embryos at the 128-cell stage. Shown are representative micrographs of individual nuclei, extracted from a spinning disk confocal

microscopy time lapse. **D.** Schematic representation of a nucleus in WT, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. **E.** Rescue of *miR430* activity as assessed by comparing the expression of eGFP encoded by an mRNA with three perfect target sites for *miR430* and RFP encoded by an mRNA without such sites, in different genotypes at 24hpf. Normalized eGFP signal in WT, *mir430*$^{+/-}$ and *mir430*$^{-/-}$ embryos without (left) and with (right) injected MiR430. N=3, 27≤n≤74. See Fig. S1 for images. **F.** Rescue of *miR430* activity as assessed by epiboly progression. Shown are representative micrographs of embryos at late epiboly stage in different genotypes. The misregulation of yolk internalization in *mir430*$^{-/-}$ embryos is indicated. **G.** Time at which epiboly is completed in different genotypes without (left), and with (right) injected MiR430 RNA. N=3, 19≤n≤53. **H.** Time at which Kupffer's vesicle appears in different genotypes without (left), and with (right) injected MiR430. N=3, 14≤n≤44. **I.** Larvae at 48 hours post fertilization for different genotypes. Representative micrographs are shown. The malformation of trunk morphology and eye, the development of heart edema, and the appearance of blisters at the tail tip in *mir430*$^{-/-}$ embryos are indicated.

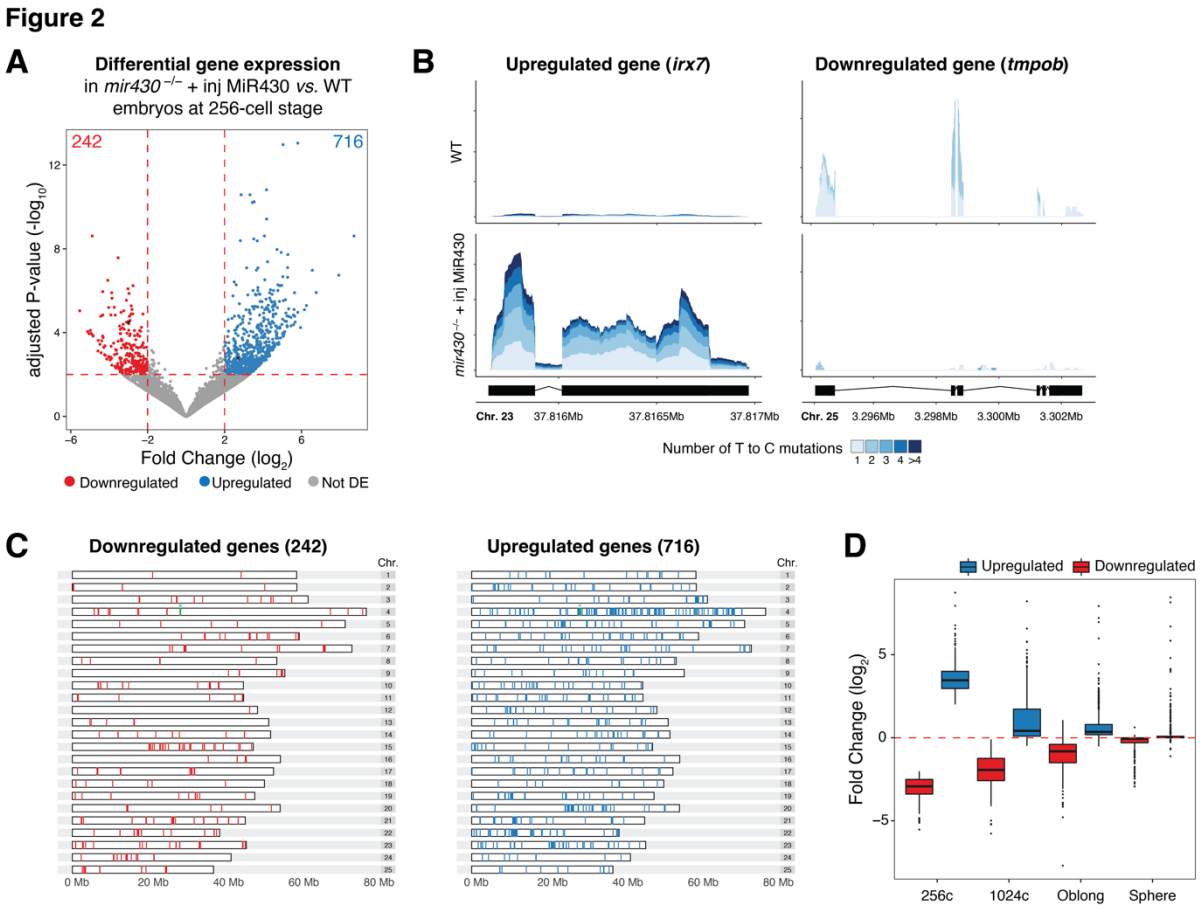

Fig. 2. Disruption of *mir430* transcription bodies causes widespread misregulation of gene expression. A. Volcano plot showing up- and downregulated genes in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT embryos at 256-cell stage. Genes whose coverage is shown in Fig. 2B are shown in black. **B.** Coverage plot of *irx7* (upregulated) and *tmpob* (downregulated) in WT and *mir430*$^{-/-}$ + inj MiR430. The single strata visualize the labeling degree of the reads. **C.** Distribution of up- and downregulated genes across the genome. The *mir430* locus on chromosome 4 is shown in green and labeled by an asterisk. **D.** Difference in average gene expression between *mir430*$^{-/-}$ + inj MiR430 and WT embryos across stages for those genes that were identified to be differentially expressed at 256-cell stage.

Figure 3

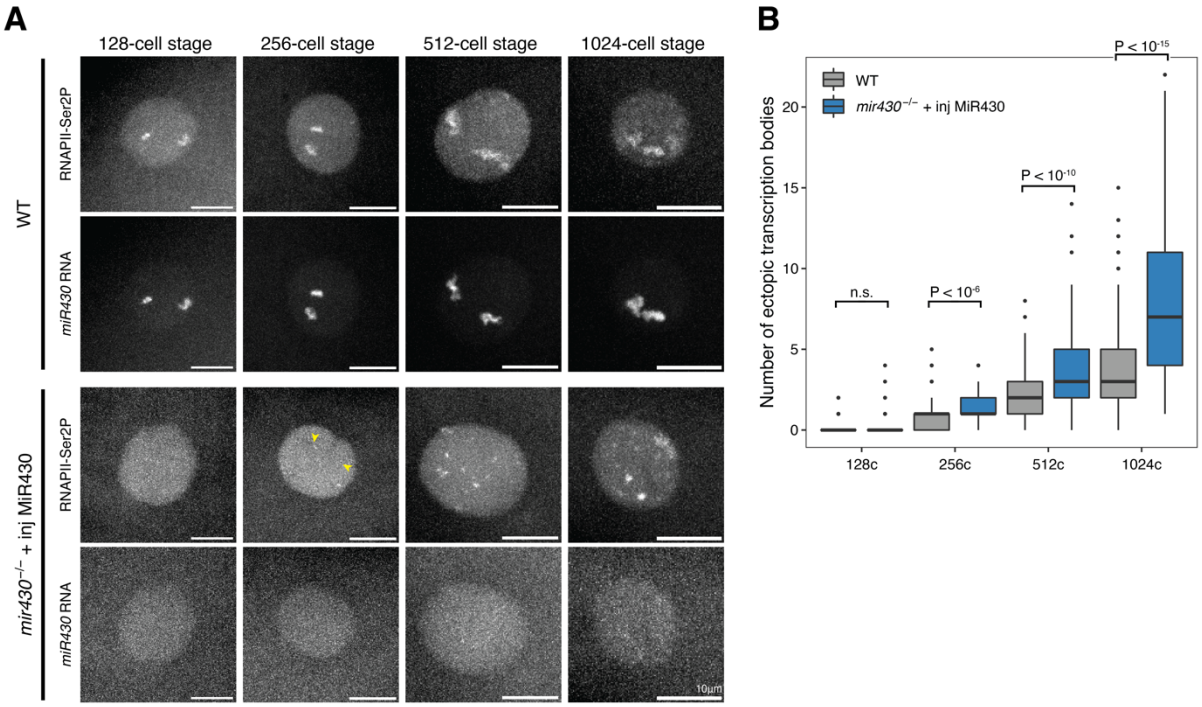

Fig. 3. In the absence of *mir430* transcription bodies, ectopic transcription bodies form. A. Visualization of RNAPII-Ser2P with antigen-binding fragments (Fabs), and *miR430* RNA with MOVIE in WT and *mir430*$^{-/-}$ + inj MiR430 embryos across stages. Shown are representative micrographs of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. See Figs. S7-10 for complete cell cycles. **B.** Quantification of ectopic transcription bodies as shown in A. N=3, 14≤n≤168.

Figure 4

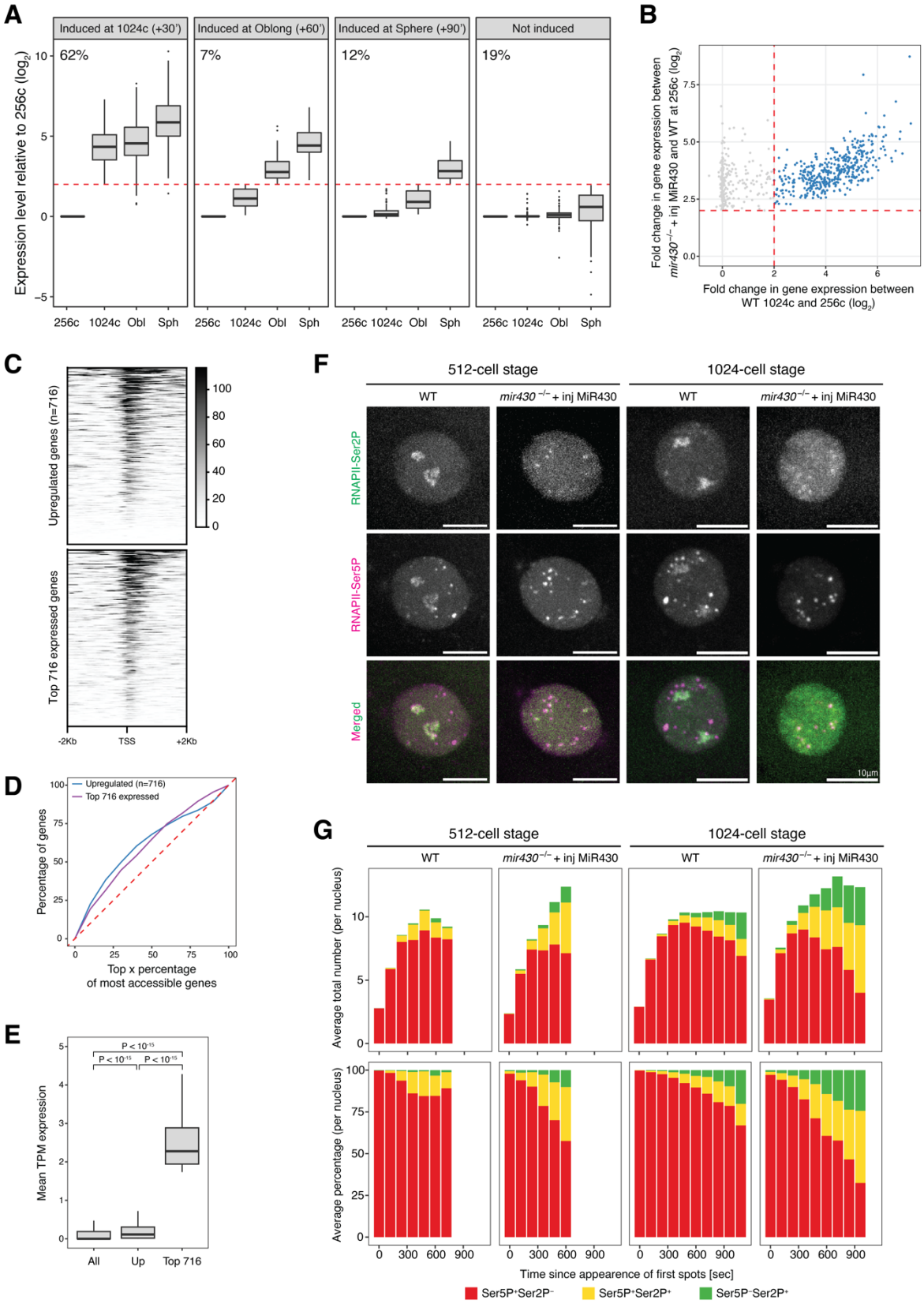
A
Induced at 1024c (+30')
Induced at Oblong (+60')
Induced at Sphere (+90')
Not induced
62%
7%
12%
19%
Expression level relative to 256c (log2)
256c 1024c Obl Sph
B
Fold change in gene expression between mir430−/− + inj MiR430 and WT at 256c (log2)
Fold change in gene expression between WT 1024c and 256c (log2)
C
Upregulated genes (n=716)
Top 716 expressed genes
-2Kb TSS +2Kb
D
Upregulated (n=716)
Top 716 expressed
Percentage of genes
Top x percentage of most accessible genes
E
P < 10-15
Mean TPM expression
All Up Top 716
F
512-cell stage
1024-cell stage
WT
mir430 −/− + inj MiR430
RNAPII-Ser2P
RNAPII-Ser5P
Merged
10µm
G
Average total number (per nucleus)
Average percentage (per nucleus)
Time since appearence of first spots [sec]
Ser5P+Ser2P-
Ser5P+Ser2P+
Ser5P-Ser2P+

Fig. 4. Loss of *mir430* transcription bodies causes premature gene activation. A. Time of induction in WT embryos for the 716 genes that are upregulated in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT embryos (at 256-cell stage). Expression at 256-cell is used as a reference. The genes are split in four groups based on when they are induced. The percentage of genes in each group is indicated. **B.** Scatterplot representing the fold change in expression between *mir430*$^{-/-}$ + inj MiR430 and WT embryos at 256-cell on the y-axis, and between 1024-cell and 256-cell in WT embryos on the x-axis. All upregulated genes are shown in gray, genes induced at 1024-cell (62%) are shown in blue. **C.** Heatmap of chromatin accessibility at 256-cell of the 716 genes that are upregulated in *mir430*$^{-/-}$ + inj MiR430 embryos compared to WT embryos at 256-cell stage, and – for comparison – the 716 most expressed genes in WT embryos at 256-cell stage. Genes are ranked by accessibility. **D.** Graph showing the fraction of promoters of the 716 upregulated (in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT at 256-cell stage), and the 716 most expressed genes (in WT at 256-cell) amongst the x% most accessible promoters at 256-cell in WT embryos. **E.** Average expression level of the 716 upregulated genes (*mir430*$^{-/-}$ + inj MiR430 *vs.* WT at 256-cell stage), the 716 most expressed genes, and all annotated genes, at the 256-cell stage, in WT. Outliers are not shown. **F.** Visualization of RNAPII-Ser2P and RNAPII-Ser5P with antigen-binding fragments (Fabs) in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 512- and 1024-cell stage. Shown are representative micrographs of individual nuclei. See Figs. S13,14 for complete cell cycles. **G.** Quantification of Ser5P-positive/Ser2P-negative (red), Ser5P-positive/Ser2P-positive (yellow) and Ser5P-negative/Ser2P-positive (green) transcription bodies during the cell cycle at 512-cell and 1024-cell, as shown in F. Number (upper panel) and percentage (lower panel) are shown.

Materials and Methods

Zebrafish and molecular biology approaches

Zebrafish husbandry and manipulation

Zebrafish were maintained and raised under standard conditions, and according to Swiss regulations (canton Vaud, license number VD-H28). Wildtype (TLAB) and mutant (*mir430*$^{-/-}$) (*9*) embryos were collected immediately upon fertilization and allowed to develop to the desired stage at 28°C. *mir430*$^{-/-}$ embryos were rescued by injecting 2nL of rescue solution (10µM of MiR430a, MiR430b and MiR430c duplex in 1x siRNA solution (300mM KCl, 30mM HEPES, 1mM $MgCl_2$, brought to pH 7.5 with KOH)), as described before (*9*, *24*). If the chorion was chemically removed, this was done by incubating embryos right after fertilization in 1.5 mg/mL Pronase E (Sigma-Aldrich, #107433) for approximately five minutes. Developmental stage was determined by morphology (according to (*36*)), the distance between nuclei, as well as cell size, cell cycle duration and synchronicity (*9*). Fabs were injected into the cell at 0.44ng/embryo (αRNAPII-Ser2P-Alexa488), or 0.66ng/embryo (αRNAPII-Ser5P-Cy5). Lissamine-labeled anti-*miR430* morpholino (5'-TCTACCCCAACTTGATAGCACTTTC-3'-Lissamine, Gene Tools) was injected into the cell at 14 fmole/embryo (*21*). s^4UTP (TriLink, tebu-bio GmbH, #N-1025-1, diluted to 50mM in 10mM Tris-HCl pH 7.5) was injected into the yolk at 50nM/embryo.

Genotyping

Embryos (live and fixed) were genotyped using the HotSHOT method (*37*).

mRNA production and injection

eGFP with three perfect miR430 target sites in its 3'UTR (eGFP-3xPT-miR430b) was inserted in the pCS2$^+$ plasmid (Addgene) as in (*24*). A pCS2$^+$-RFP plasmid was used as a control. mRNA was synthesized using the mMessage mMACHINE Transcription Kit (ThermoFisher, #AM1340) and purified with the RNeasy MinElute Cleanup Kit (QIAGEN, #74204). 50pg of the eGFP mRNA and 100pg of the RFP mRNA were injected into each embryo at the 1-cell stage.

Preparation of antigen-binding fragments (Fabs)

Fluorescently labeled Fabs specific to RNAPII-Ser5P and RNAPII-Ser2P were prepared from monoclonal antibodies specific to RNAPII-Ser5P and Ser2 phosphorylation (*38–40*). Monoclonal antibodies were digested with Ficin (ThermoFisher Scientific), and Fabs were purified through protein A-Sepharose columns (Cytiva) to remove Fc and undigested IgG. After passing through desalting columns (PD MiniTrap G25; Cytiva) to substitute the buffer with PBS, Fabs were concentrated up to >1 mg/ml using 10 k cut-off filters (Amicon Ultra-0.5 10 k; Merck). Fabs were conjugated with Alexa Fluor 488 (Sulfodichlorophenol Ester; ThermoFisher Scientific) or Cy5 (N-hydroxysuccinimide ester monoreactive dye; Cytiv) to yield ~1:1 dye:protein ratio. After the buffer substitution with PBS, the concentration was adjusted to ~1 mg/ml.

Probe production

A DIG-labeled antisense probe against *goosecoid* (*gsc*) was generated by *in vitro* transcription from a linearized plasmid using 10x Transcription buffer (Roche, #11465384001), 10x DIG RNA labeling mix (Roche, #11277073910), T7 RNA polymerase (Roche, #10881775001) and RNase OUT (Invitrogen, #10777019). Probe was purified with the RNeasy MinElute Cleanup Kit (QIAGEN, #74204), diluted in Hybridization buffer and stored at -20°C.

Whole mount in situ hybridization

Whole mount in situ hybridization (WM-ISH) was performed as described before (*41*), with some differences. In brief, embryos were fixed at the desired stage in 4% (wt/vol) PFA in PBS pH 7.4 overnight at 4°C. Embryos were then manually dechorionated with sharp forceps (Dumont No. 5) in PBS with 0.1% Tween-20 (PBST) and dehydrated by rinsing them in 50% and then 100% methanol in PBST on a nutator. At this point, embryos could be stored for several weeks at -20°C. To continue, embryos were gradually rehydrated on a nutator, incubated in Hybridization buffer (50% formamide, 1.3x SSC pH 5.0, 5mM EDTA, 50µg/mL torula yeast RNA, 0.2% Tween-20, 0.1% SDS, 100µg/mL Heparin) for 1 hour at 70°C, and incubated with anti-*goosecoid* DIG-labeled probe overnight at 70°C. The next day, the probe was washed off with Hybridization buffer and embryos were incubated for 15 minutes in 50% (v/v) Hybridization buffer in TBST (25mM Tris-HCl pH 7.8, 137mM NaCl, 2.7mM KCl, 0.1% Tween-20) at 70°C. After rinsing with TBST, embryos were incubated for 1 hour in TBST and then in Blocking Buffer (2% blocking reagent in maleic acid buffer pH 7.5 (Roche, #11096176001) and 10% sheep serum (Sigma Aldrich, #S2263)

in TBST) at room temperature on a nutator. anti-DIG antibody (Roche, #11093274910) was added in a dilution of 1:2.000 and embryos incubated overnight at 4°C on a nutator. The next day, the antibody was washed off with TBST and then NTMT (100mM NaCl, 100mM Tris-HCl pH 9.5, 50mM MgCl2, 0.1% Tween-20), and stained in NTMT-NBT-BCIP solution (4.5µL/mL of NBT (Roche, #11383213001) and 3.5µL/mL of BCIP (Roche, #11383221001) in NTMT) in the dark. Once sufficiently stained, the staining solution was washed off with PBST, and embryos were briefly washed in 100%, 50% and 30% methanol in PBST at room temperature on a nutator. Embryos were stored at 4°C in PBST for up to a few days before imaging.

Enriched SLAM-Sequencing (eSLAM-Seq)

Rationale

Early embryos contain large amounts of maternally loaded RNA (*19*) which mask the nascent RNAs in total RNA sequencing approaches. We therefore developed a method to sensitively detect nascent transcripts. Methods to enrich nascent transcripts have previously been described. Typically, an uracil/UTP analog (EU/s^4UTP) is injected at the 1-cell stage and is incorporated into newly made RNAs. Such labeled transcripts can then be biotinylated and enriched by pull-down (*20*, *42*). SLAM-Seq is another approach in which nascent transcripts are labeled by s^4UTP incorporation, but here the label is used to convert 4-thio-uracils to carboxyamido-methylated 4-thio-uracils, which will be misread as a cytosine during reverse transcription, introducing specific point mutations that enable the identification of reads derived from labeled transcripts by sequencing (*43–45*). We combined these two approaches in enriched SLAM-Seq (eSLAM-Seq), which is conceptually similar to TT-TimeLapse-Seq and TT-SLAM-Seq (*46*, *47*). Specifically, we used a combination of s^4UTP labeling (*42*) with improved biotinylation and enrichment chemistry (*48*, *49*), followed by SLAM-Seq chemistry.

Sample preparation

Because of the light-sensitive nature of s^4UTP, all steps until the treatment of RNA with iodoacetamide (IAA) were performed in the dark. When staging embryos at the stereomicroscope, for example, we used the lowest light-setting that allowed us to still see the embryos, and a UV filter was set between the injection plate and the light source to protect the s^4UTP. Successive steps

were performed protecting the samples from the light by using low light settings in the laboratory room, by working inside a cardboard box, and by covering the samples with aluminum foil.

Synchronized s^4UTP-injected embryos were collected and left to develop to the desired stage. They were collected in as little 0.3x Danieau's Buffer as possible, homogenized in 500µL Qiazol (QIAGEN, #79306) with a 5mm stainless steel bead (QIAGEN, #69989) using a TissueLyser (30 seconds at 30Hz) and stored at -80°C. 100 embryos (256-cell stage) or 50 embryos (1024-cell, Oblong and Sphere stage) were used. Chloroform was added and RNA was extracted by phase separation, using a MaXtract tube (QIAGEN, #129046), after which it was precipitated with isopropanol supplemented with 0.2mM DTT and glycogen. The pellet was washed with 80% ethanol supplemented with 1mM DTT, air-dried, and resuspended in 40µL of 1mM DTT. DNase treatment was performed in 100µL of 1x DNAse reaction buffer with 4U of TURBO DNase (ThermoFisher, AM2238) for 30 minutes at 37°C. The volume was brought to 250µL with nuclease-free water and RNA was extracted with 25:24:1 Phenol-Chloroform-Isoamyl alcohol (Invitrogen, #15593-031) using a MaXtract tube, precipitated with isopropanol supplemented with 0.2mM DTT, glycogen and sodium acetate. The pellet was washed with 80% ethanol supplemented with 1mM DTT, air-dried, and resuspended in 40µL of nuclease-free water.

Labeled RNA was biotinylated in 100µL of 20mM HEPES pH 7.4, 1mM EDTA pH 8.0, 10ng/µL MTSEA-biotin-XX (Biotium, #90066) and 20% DMF) by incubation at room temperature for 30 minutes while rotating. The volume was brought to 250µL with nuclease-free water and RNA was extracted with 24:1 Chloroform-Isoamyl alcohol (Sigma-Aldrich, #227056) using a MaXtract tube, and precipitated with isopropanol supplemented with glycogen and NaCl. The pellet was washed with 80% ethanol, air-dried, and resuspended in 50µL of nuclease-free water.

To enrich for biotinylated RNA, Dynabeads MyOne Streptavidin C1 beads (Invitrogen, #65001) were used. Per sample, 10µL of beads were washed twice in nuclease-free water and twice in high-salt wash buffer (100mM Tris-HCl pH 7.4, 10mM EDTA pH 8.0, 1M NaCl, 0.05% Tween-20), incubated for at least 1 hour at room temperature in bead-blocking buffer (40ng/µL glycogen in high-salt wash buffer) and washed twice with high-salt wash buffer. RNA was denatured at 65°C for 10 minutes, incubated for 5 minutes on ice and then added to the blocked beads. After

incubation for 15 minutes while rotating at 30 rpm, the supernatant was removed and beads were washed 3 times with high-salt wash buffer, twice with 5.9M guanidinium chloride (with 5-minute incubation at room temperature between consecutive washes), once with TE buffer and three times with TE buffer (with 5-minute incubation at 55°C between consecutive washes). Bead-bound RNA was finally released by adding 25µL of freshly made elution buffer (100mM DTT, 20mM HEPES pH 7.4, 1mM EDTA pH 8.0, 100mM NaCl, 0.05% Tween-20) twice. Volume was brought to 100µL with nuclease-free water and RNA was precipitated with isopropanol supplemented with 0.2mM DTT, glycogen and NaCl. The pellet was washed with 80% ethanol supplemented with 1mM DTT, air-dried, and resuspended in 20µL of 1mM DTT.

Iodoacetamide-mediated alkylation was performed in 10mM IAA (freshly dissolved in 100% ethanol, BioUltra, #I1149), 50mM $NaPO_4$ pH 8.0 and 50% DMSO for 15 minutes at 50°C. Samples were then transferred on ice and the reaction stopped by quickly adding 1.4µL of 1M DTT. RNA was precipitated with 100% ethanol supplemented with glycogen and sodium acetate overnight. The pellet was washed with 80% ethanol, air-dried, and resuspended in 8.5µL of nuclease-free water.

Libraries were prepared with the TruSeq stranded mRNA Kit (Illumina, #20020594) and TruSeq RNA UD Indexes (Illumina, #20022371). The polyA enrichment step was not performed. Fragmentation was performed for 1 minute at 94°C. Paired end 150bp sequencing was performed on an Illumina NovaSeq 6000 System, loading each library on two lanes (Illumina Reagent Kit v1.5). Sequencing data were demultiplexed with bcl2fastq 2.20.

Quality controls

To assess the integrity of total RNA (Fig. S4C, left panel), 1µg/sample was diluted in 2x RNA Gel Loading Dye (Thermo Scientific, #R0641), denatured for 90 seconds at 85°C, chilled for 5 minutes on ice and loaded onto an agarose gel alongside RiboRuler High Range RNA Ladder (ThermoFisher, #SM1823). The presence of two prominent bands (rRNA) are an indicator of RNA integrity.

To test the specificity of s^4UTP biotinylation (Fig. S4C, right panel), the RNA was then transferred from the gel onto a Nylon Hybond-N+ membrane (Amersham, #RPN203B) by capillary transfer using a Whatman Turboblotter in 20X SSC according to manufacturer's instructions overnight. After drying, the membrane was crosslinked twice with 120mJ/cm^2 using a UV Stratalinker 1800 (Stratagene). The membrane was then incubated for 20min in Blocking Buffer (10% SDS and 1mM EDTA in 1xPBS pH 7.5), for 1h in 1:10.000 Streptavidin-HRP conjugate (Perkin, #NEL750) in Blocking buffer, 2x for 10min in 10% SDS in 1xPBS pH 7.5, 2x for 10min in 1% SDS in 1xPBS pH 7.5, and 2x for 10min in 0.1% SDS in 1xPBS pH 7.5, all while nutating (*50*). Finally, the signal was visualized with ECL Western Blotting Detection Reagents (Amersham, #RPN2106) according to manufacturer's instructions, and documented with a Fusion FX (Vilber).

Bioinformatic analyses

eSLAM-Seq

To map the raw sequencing data we concatenated Fastq files that came from the same library but that were run on different lanes with the 'cat' command (GNU core utilities, linux terminal), trimmed them using Trimmomatic (Paired-end mode, LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (*51*), and mapped them using HISAT-3N (--base-change T,C --repeat --rna-strandness RF --fr --no-discordant) (*52*), which is a HISAT variant specifically designed for methods like SLAM-Seq that introduce specific point-mutations. Uniquely mapped fragments were filtered with Samtools (*53*) and mate information was fixed with Picard FixMateInformation (--IGNORE_MISSING_MATES FALSE) (*54*).

To distinguish between naturally occurring SNPs and SLAM-Seq-specific mismatches, we first identified naturally occurring SNPs by analyzing four IAA-minus samples (one per stage) in which the iodoacetamide treatment – and therefore the base conversion – was not performed. The bam files were converted to a suitable format with the Samtools mpileup function and then SNPs were identified by Varscan pileup2snp (--min-coverage 20 --min-reads2 5 --min-avg-qual 15 --min-var-freq 0.25 --p-value 0.01). The union of the SNPs that were identified was used as a consensus list. Then the splbam.py script from pulseR (*55*) was used to select fragments with at least one SLAM-Seq-specific mismatch.

To assess gene expression levels of annotated genes, raw read counts were obtained using FeatureCounts (-p --countReadPairs -B --minOverlap 10 -s 2 -t exon -g gene_id) (*56*). Normalization, Principal Component Analysis (PCA) and call for differential expression (alpha = 0.01 and $\log_2$(Fold Change) threshold = 2) was performed using DESeq2 (*57*). When performing PCA, only the top 500 genes (selected by highest variance in expression levels) were considered. When counting the number of differentially expressed genes, mitochondrial and *mir430*-coding genes were excluded.

Motif analysis

To assess the presence and enrichment of specific motifs in promoters of upregulated genes, we started with the non-redundant JASPAR 2022 Vertebrate core database. From this, only motifs whose cognate transcription factors are expressed during ZGA (Riboprofiling dataset, >10rpkm, (*30*)) were kept. The Nanog motif (*29*) and Pou5f1 motif (*31*) were added, because it is known that these factors are important during zebrafish ZGA. Next, we extracted promoter sequences (4000bp centered around the TSS) for upregulated genes as well as non-expressed genes (less than 10 raw reads in total across all samples, background group). Motif enrichment was then assessed using SEA (MEME Suite, (*58*)). To estimate depletion of motifs, the background and input sample were inverted. Only motifs with a P-value < 0.05 and an E-value ≤ 10 were considered.

Promoter accessibility analysis

To analyze promoter accessibility of the upregulated genes in WT, we used an ATAC-Seq dataset that we previously generated (*29*). Replicates were merged and the signal in 4000bp centered around the TSS was plotted using Deep Tools (*59*). Promoter signal was used to rank all promoters by accessibility and to assess their enrichment amongst the most accessible promoters. The 716 most expressed (mean TPM) genes (excluding mitochondrial genes and *mir430*-coding genes) were used as a comparison.

Graph plotting for eSLAM-Seq

Plots were generated using R version 4.0.5 (*60*) with the ggplot2 (*61*) and Ggbio package (*62*).

Microscopy

Fluorescence microscopy

To determine MiR430-activity using the GFP sensor assay, injected embryos were manually dechorionated at 24hpf and kept in Petri dishes. Fluorescence signal was measured on an OLYMPUS SZX16 stereo microscope with a MicroPublisher 5.0 RTV-R-CLR-10C color camera (5x magnification, 400ms exposure time).

Bright field imaging

For bright field time lapse imaging, a mold (*63*) was used to create agarose wells using as little as possible 1% agarose in 0.3x Danieau's solution in an Ibidi µ-Dish 35mm, high (#81158). The dish was mounted on a 28°C heated stage (by a Warner automatic temperature controller) and dechorionated embryos were loaded to the single wells at the 2-cell stage. Movies with 10-minute time intervals were taken on a Zeiss Axio Observer.Z1 microscope (Zeiss 5x/0.25 Fluar air objective, Zeiss AxioCam MRm Monochrome CCD camera). A z-stack (400µm in 9 steps) was taken for each embryo. A lid was used to reduce evaporation of buffer during imaging.

To assess developmental progression of live embryos past somitogenesis, snapshots of manually dechorionated embryos were taken at 24hpf, 48hpf and 72hpf on a Zeiss Axio Observer.Z1 microscope (Zeiss 2.5x/0.08 EC Plan-Neofluar air objective, Zeiss AxioCam ICc 1).

The gene expression pattern in fixed embryos after WM-ISH was visualized on a LEICA M165C stereomicroscope with LEICA MC170 HD color camera (4x magnification). Embryos were oriented such that the dorsal region of the embryo was visible.

Spinning Disk microscopy

For imaging on a spinning disk microscope, dechorionated embryos were mounted at 32-cell stage in small droplets of 1% low-melting agarose (Invitrogen, #16520-050) in 75% v/v 0.3x Danieau's solution and 25% v/v Optiprep Density Gradient Medium (Sigma Aldrich, #D1556) (*64*) on a Ibidi µ-Dish 35mm, high (#81158). Embryos were brought closer to the coverslip surface by keeping

the dish upside down until the agarose solidified (~20 minutes). More agarose was then added to avoid drying out.

Transcription bodies in live embryos were imaged on a Nikon eclipse Ti2 with a Yokogava CSU-W1 Spinning Disk unit at 28°C (Okolab temperature control and microscope enclosure). Multi-position, z-stack (20µm in 41 steps, NIDAQ Piezo Z) time-lapse (2 minutes time interval) imaging was performed with a Nikon 60x/1.2 Plan Apochromat VC water objective. Images were captured in parallel on two Photometrics Prime 95B cameras with 100ms exposure time at 10% laser intensity.

Image processing and analysis

Software used for image analysis

Microscopy image handling was done using FIJI (*65*). Further data processing was carried out using R version 4.0.5 (*60*) and plots were generated with the ggplot2 package (*61*).

Analysis of eGFP reporter and RFP control

To assess *mir430* activity, mean fluorescence of eGFP (targeted by MiR430) and RFP (not targeted, and used as a normalizer) were measured in a rectangle of equal size in the head region and a background region outside of the embryo. The signal ratio was then calculated as follows: $(eGFP_{head} - eGFP_{background}) / (RFP_{head} - RFP_{background})$.

Detection and classification of mir430 and ectopic transcription bodies

We identified transcription bodies by the presence of RNAPII-Ser2P signal. To classify them into *mir430* and ectopic transcription bodies, we analyzed *miR430* MOVIE signal. In case of overlap between RNAPII-Ser2P and MOVIE signal, bodies were classified as *mir430* transcription bodies. In case of no overlap between RNAPII-Ser2P and MOVIE signal, bodies were classified as ectopic transcription bodies. We used FIJI plugin ComDet (*66*), which detects spots using consistent intensity thresholding parameters that were manually set and used for all experiments. We note that frames taken after the disassembly of the nuclear envelope were not analyzed because at that time transcription bodies also disassembled.

Detection and classification of transcription bodies by transcription state

Nuclei were segmented using the RNAPII-Ser5P channel, and RNAPII-Ser5P and Ser2P bodies detected using ComDet. This was used to classify transcription bodies into initiating, initiating-elongating and elongating bodies. We then identified and removed *mir430* transcription bodies from the analysis based on their morphological features (early appearance during cell cycle, larger size, growth behavior) using the RNAPII-Ser2P channel. Here too, frames taken after the disassembly of the nuclear envelope were not analyzed. For each nucleus, t=0 is set as the first frame at which spots appear, and the last time point is set as the last frame before prophase. Nuclei in which spots were detected for only one frame were discarded.

Sample size and statistics

A minimum of 3 biological replicates (N) was acquired for each experimental condition. Lower-case n (n) refers to the number of embryos or nuclei for each experimental condition.

Kruskal-Wallis tests were performed to determine if there is a statistically significant difference between the three genotypes in the degree of MiR430 activity as determined by the eGFP-Sensor assay (Fig. 1E), the time of completion of Epiboly (Fig. 1G) and the time of appearance of Kupffer's vesicle (Fig. 1H). If this test was statistically significant (P-value < 0.05), pairwise comparisons with Bonferroni correction were performed using a pairwise Wilcoxon rank-sum test. A comparison was considered significant when adjusted P-value < 0.05. Dotplots were used to display the data, and adjusted P-values were reported using WT as reference. Red rhombuses represent the median.

Wilcoxon rank-sum tests were performed to determine if there is a statistically significant difference in the number of detected ectopic transcription bodies between WT and *mir430*$^{-/-}$ + inj MiR430 (Fig. 3B). A comparison was considered significant when P-value < 0.05.

A pairwise Wilcoxon rank-sum test with Bonferroni correction was performed to determine if the upregulated genes and the top 716 most expressed genes are significantly higher expressed

compared to all genes (based on TPM expression values) (Fig. 4E). A comparison was considered significant when adjusted P-value < 0.05.

Figure S1

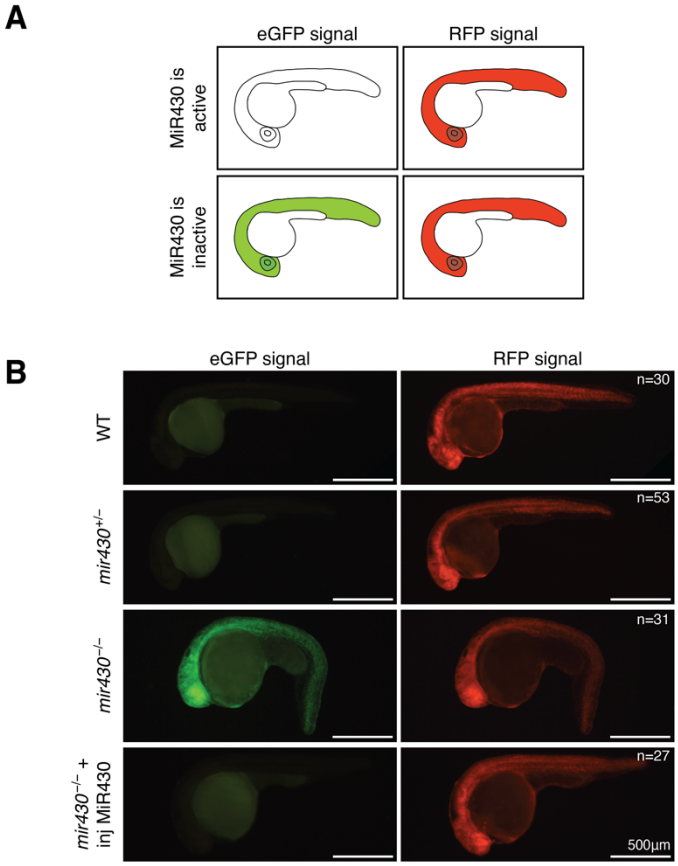

Fig. S1. MiR430 activity is restored by injecting MiR430 RNA in *mir430*$^{-/-}$ embryos. A. Schematic representation of expected eGFP and RFP expression in embryos with active and inactive MiR430 microRNA activity. **B.** Representative micrographs showing eGFP and RFP expression in WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos at 24hpf.

Supplementary Figure 2

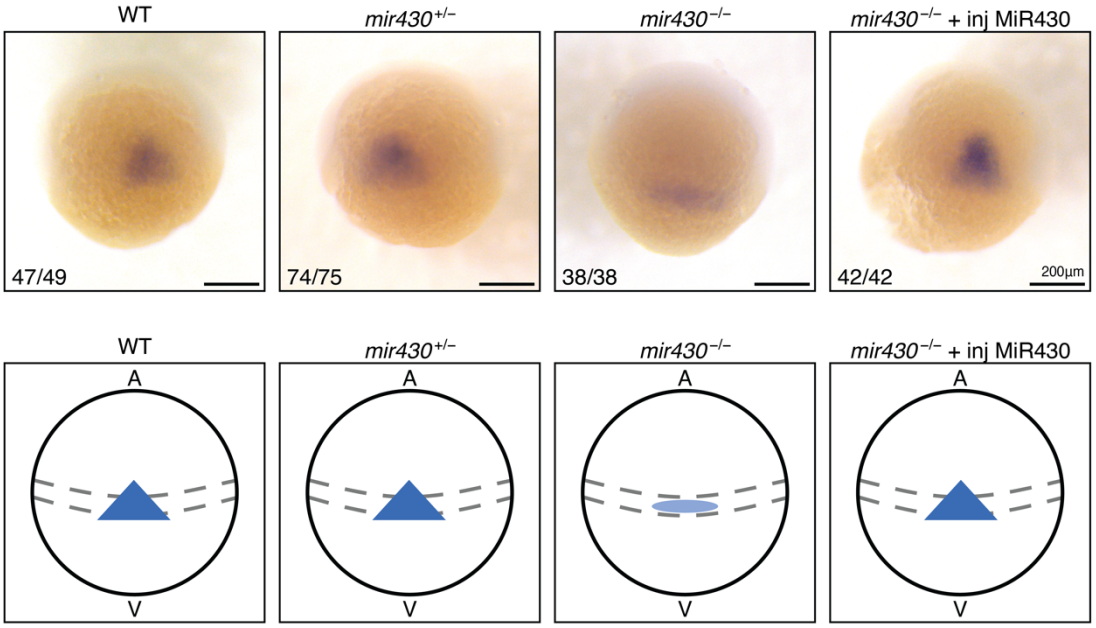

Fig. S2. Phenotypic analysis of *mir430*$^{-/-}$ embryos with and without injected MiR430 (*gsc* expression). Whole mount *in situ* hybridization (WM-ISH) of *Goosecoid (Gsc)* in WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos at Shield stage. Representative micrographs show the embryos in dorsal view. Misexpression (reduced expression levels and reduced anterior extension of expression pattern) of *Gsc* can be observed in *mir430*$^{-/-}$ embryos. The fraction in the lower left corner represents the frequency with which the shown phenotype was observed. A schematic representation of the phenotypes is shown below.

Supplementary Figure 3

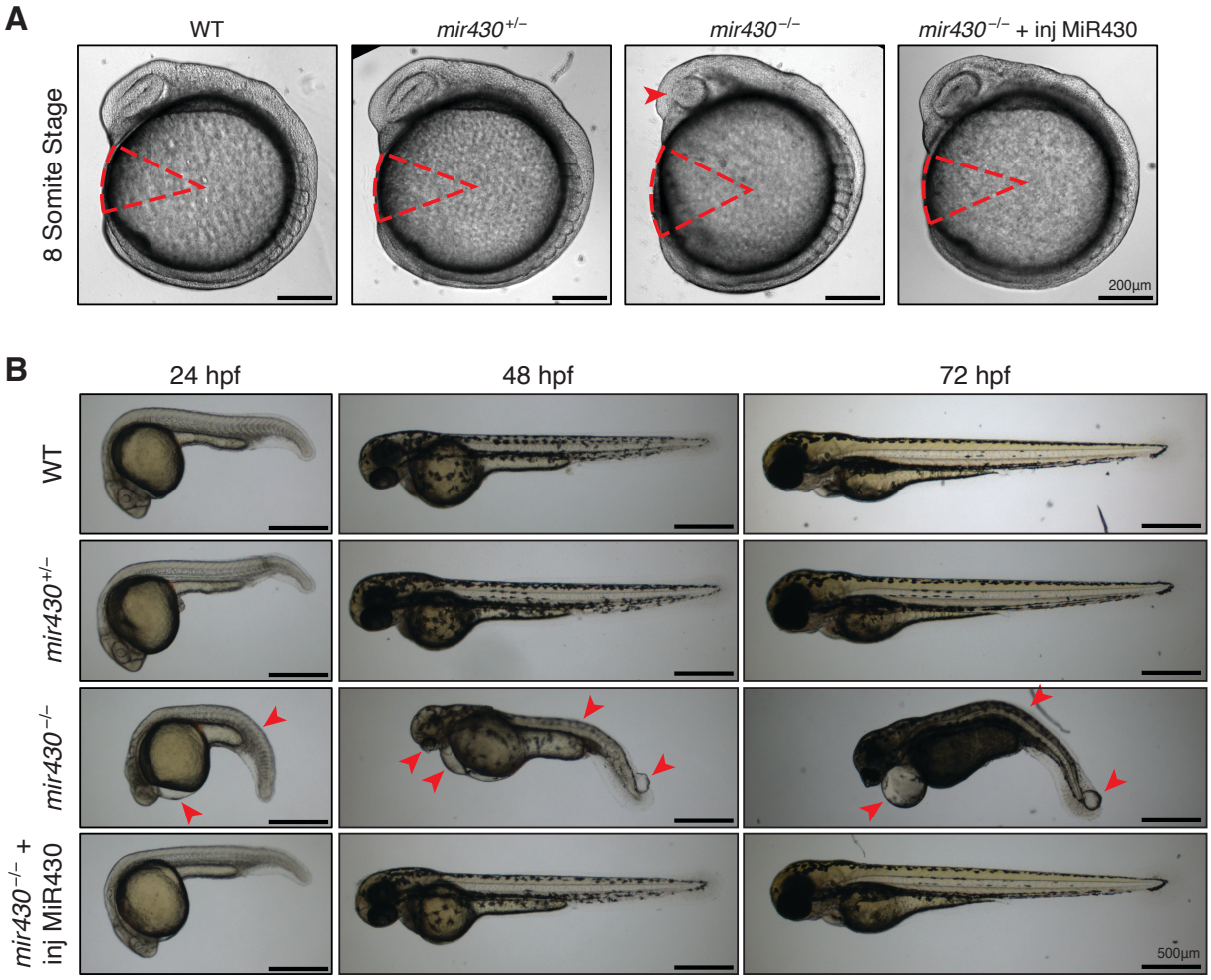

Fig. S3. Phenotypic analysis of *mir430*$^{-/-}$ embryos with and without injected MiR430 (developmental progress). A. Embryos at 8-Somite stage for WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. The reduced body axis extension and the malformation of the optic primordium in *mir430*$^{-/-}$ embryos are highlighted. Representative micrographs are shown. **B.** Larvae at 24, 48 and 72 hours post fertilization (hpf) for WT, *mir430*$^{+/-}$, *mir430*$^{-/-}$ and *mir430*$^{-/-}$ + inj MiR430 embryos. The malformation of trunk morphology and of the eye, the development of heart edema and the appearance of blisters at the tail tip in *mir430*$^{-/-}$ embryos are highlighted. Representative micrographs are shown. The 48hpf stage is also shown in Fig. 1I.

Supplementary Figure 4

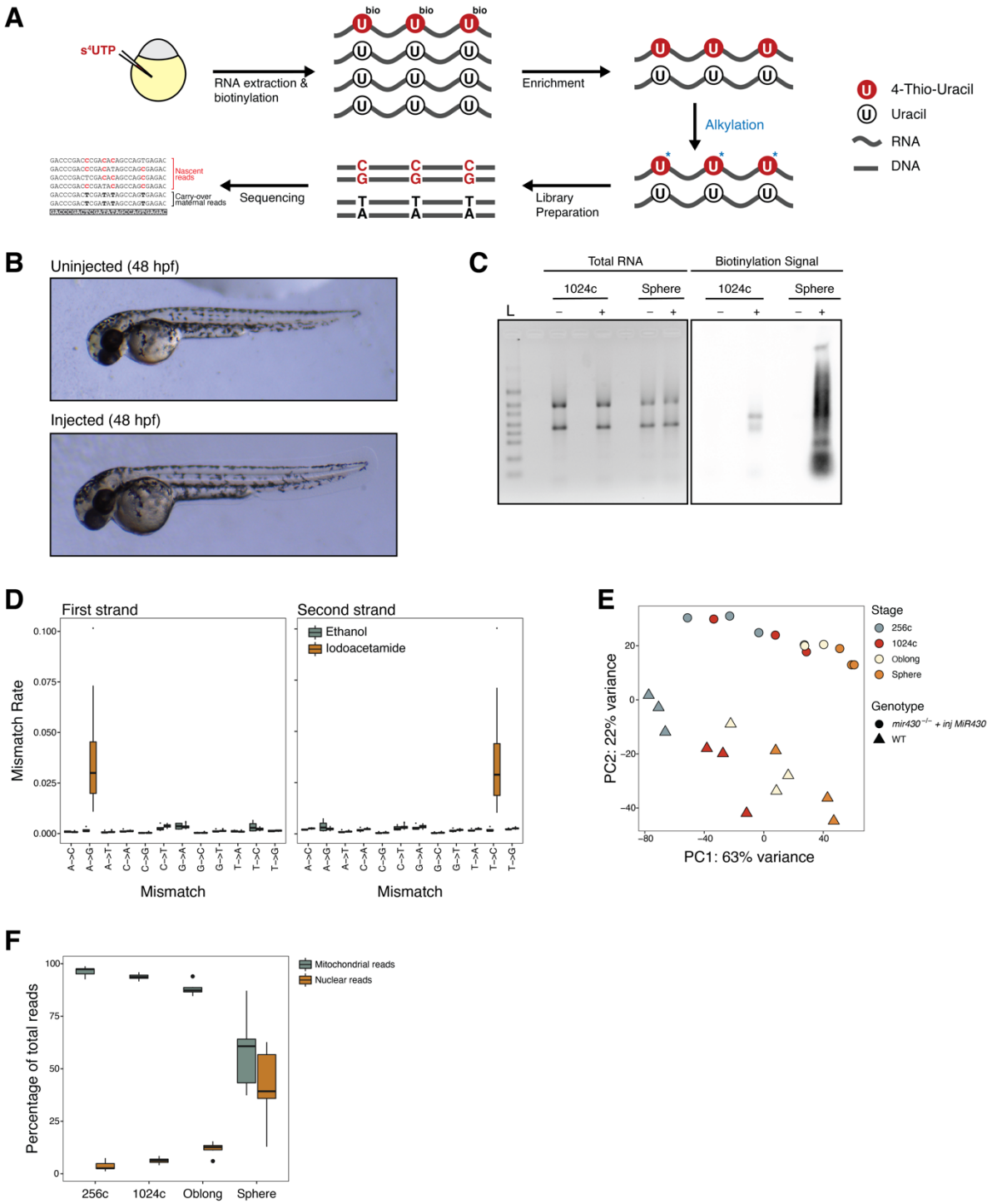

Fig. S4. Establishment of enriched SLAM-Sequencing (eSLAM-Seq) in zebrafish embryos.

A. Schematic representation of eSLAM-Seq protocol. See Materials and Methods for more detail.

B. Representative micrographs of larvae that were/were not injected with 50nM/embryo of s^4UTP

at the 1-cell stage. No developmental malformations were observed upon injection. **C.** Equal amounts of biotinylated and heat-denatured total RNA derived from injected (+) and uninjected (-) embryos at 1024-cell and Sphere stage were run on an agarose gel (left panel). The RNA was then transferred on a nylon membrane and biotinylation detected by Streptavidin-HRP (right panel). Biotinylation was only observed in RNA derived from injected embryos, showing that biotinylation is specific for s^4UTP-labeled RNA. L: RNA Ladder. **D.** Box plot representation of mismatch rate in Iodoacetamide- (N=24) and Ethanol-treated (N=4) samples. As expected, only Iodoacetamide-treated samples show a high level in A>G (First strand) and T>C (Second strand) mismatch rate. **E.** Principal Component Analysis (PCA) plot based on the top 500 genes (selected by highest variance in expression levels). Samples separate according to their developmental stage along the first PC and according to their experimental condition (WT and *mir430*$^{-/-}$ + inj MiR430) along the second PC. **F.** Box plot representation of percentage of mitochondrial and nuclear reads, separated by developmental stage. Most reads are of mitochondrial origin at early ZGA, as shown before (*42*), with a gradual increase in the nuclear fraction during later stages.

Supplementary Figure 5

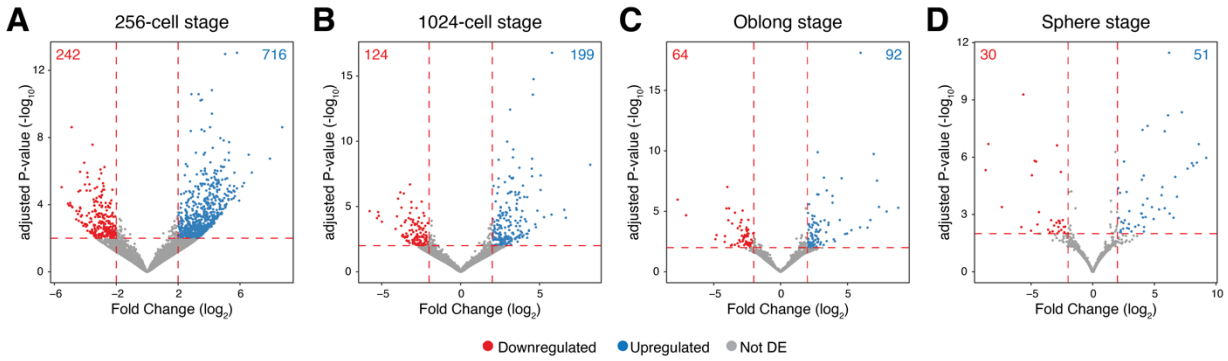

Fig. S5. The disruption of two large endogenous transcription bodies causes short-term misregulation of gene expression. Volcano plots showing up- and downregulated genes in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT embryos at 256-cell **(A)**, 1024-cell **(B)**, Oblong **(C)** and Sphere **(D)** stage. Panel A is also shown in Fig. 2A.

Supplementary Figure 6

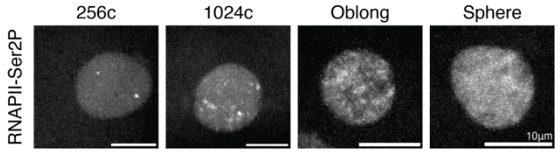

Fig. S6. RNAPII-Ser2P signal at 256-cell, 1024-cell, Oblong and Sphere stage in *mir430*$^{-/-}$ + inj MiR430 embryos. Visualization of RNAPII-Ser2P signal with Fabs. Representative micrographs are shown.

Supplementary Figure 7

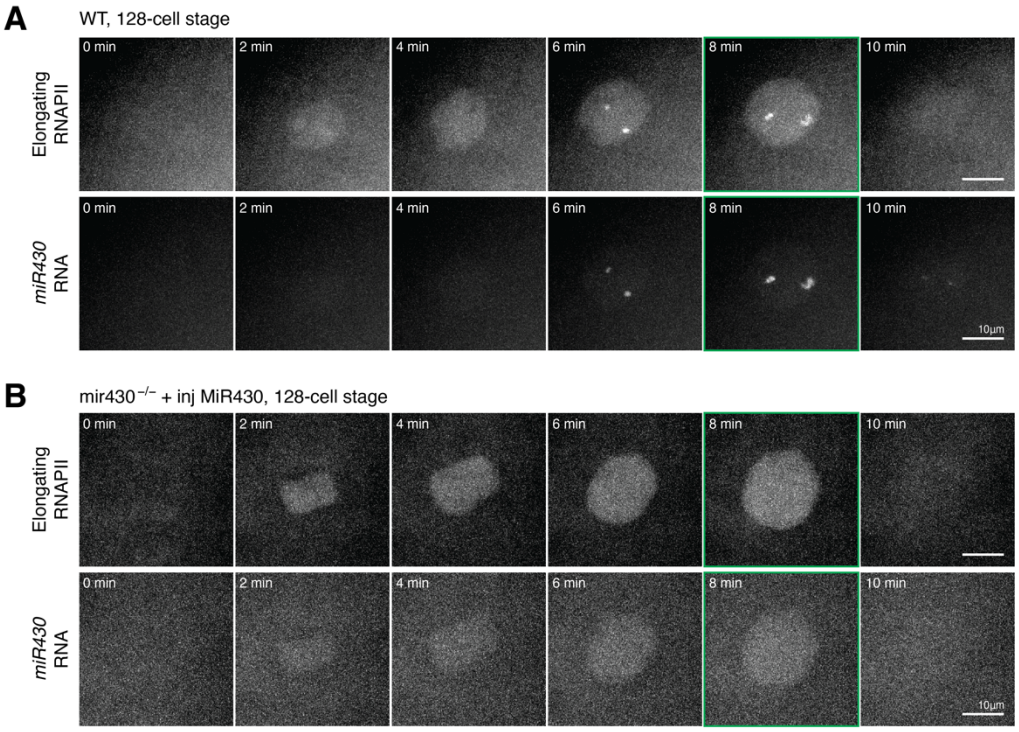

Fig. S7. RNAPII-Ser2P signal in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 128-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 128-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 8

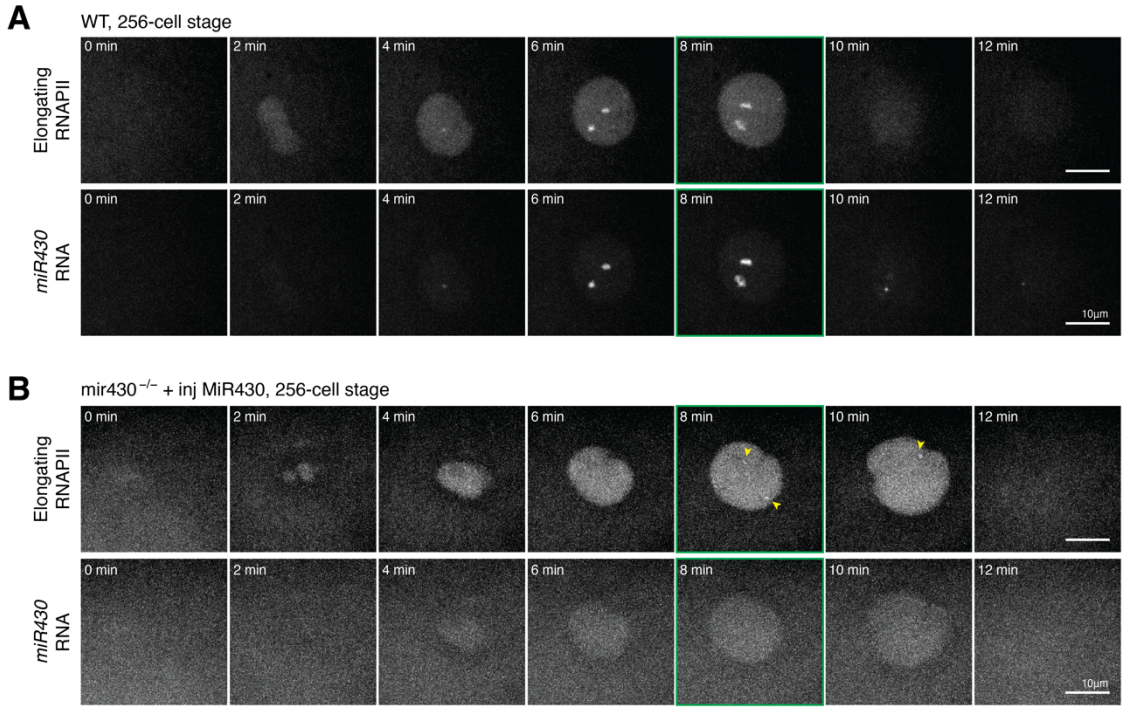

Fig. S8. RNAPII-Ser2P signal in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 256-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 256-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 9

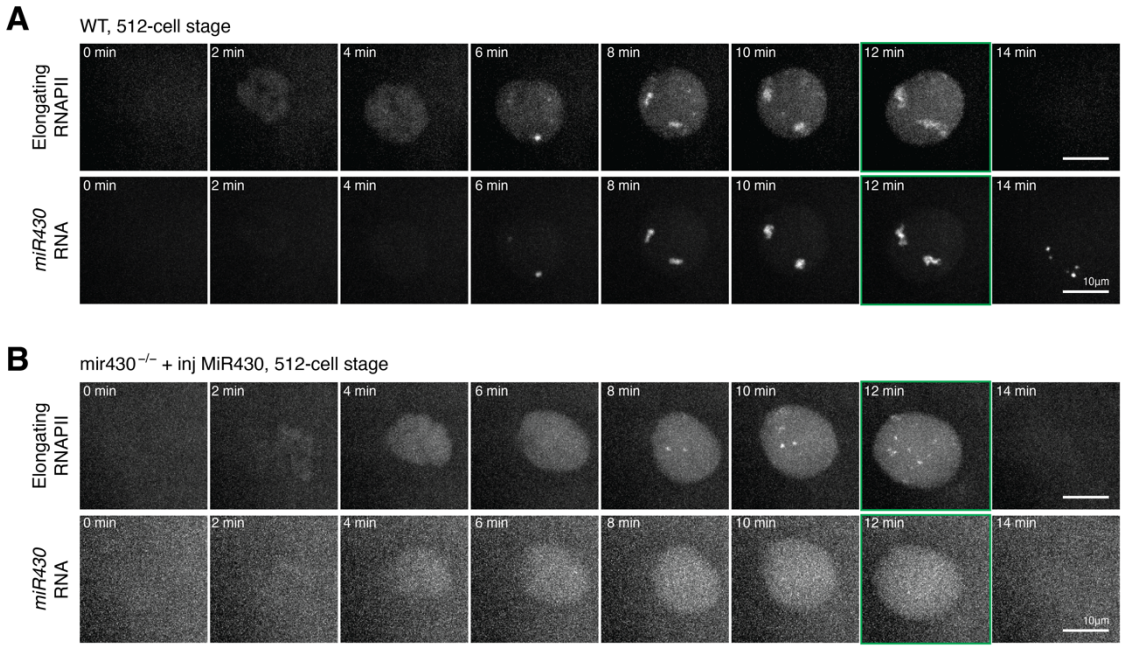

Fig. S9. RNAPII-Ser2P signal in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 512-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 512-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 10

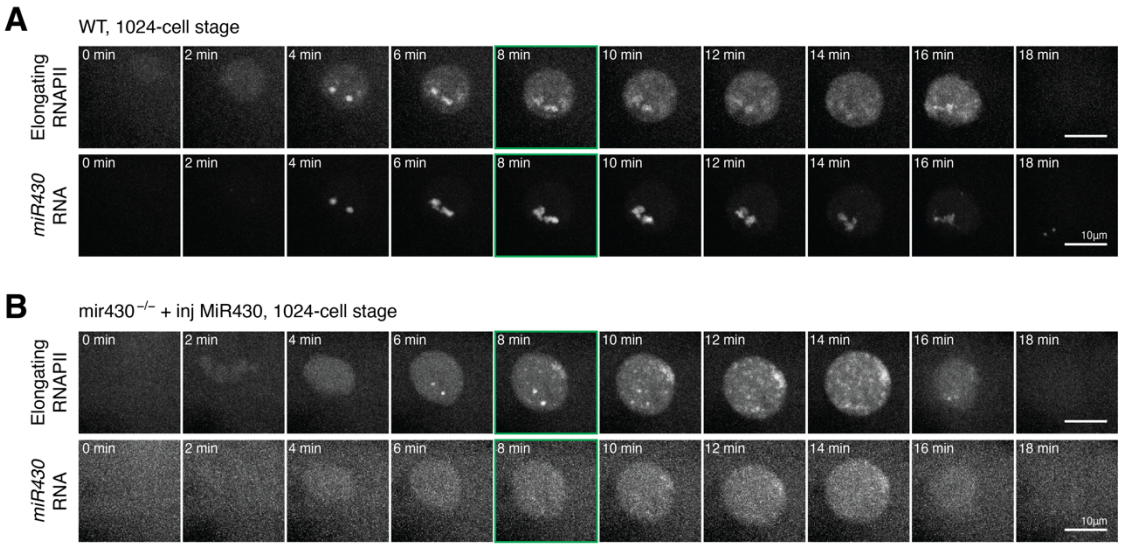

Fig. S10. RNAPII-Ser2P signal in WT and *mir430*$^{-/-}$ + inj MiR430 embryos at 1024-cell stage. Visualization of RNAPII Ser2P (with Fabs) and *miR430* RNA (with MOVIE) in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 1024-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 11

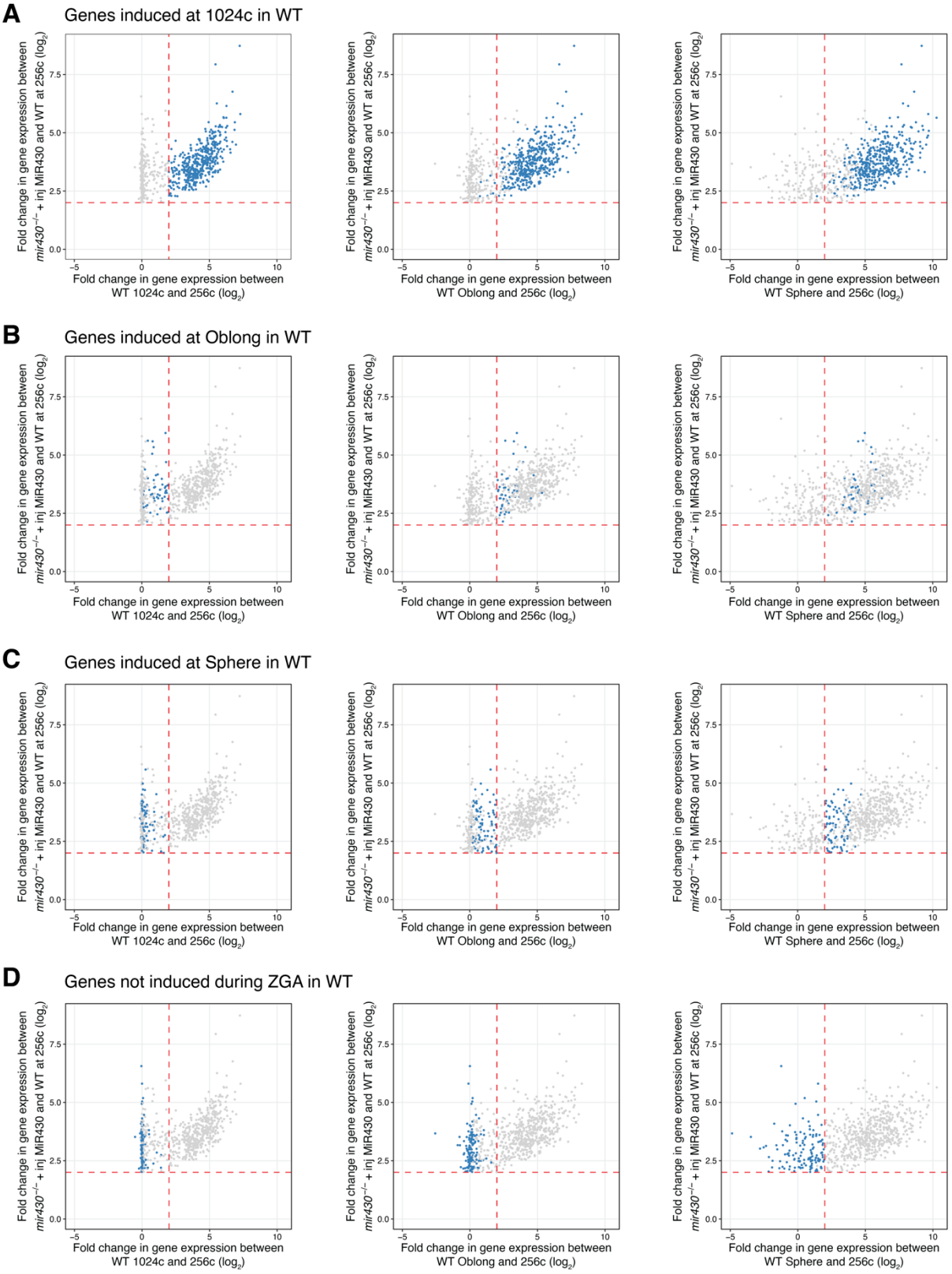

Fig. S11. Correlation between expression level in *mir430*$^{-/-}$ + inj MiR430 *vs.* WT and 1024-cell, Oblong and Sphere *vs.* 256-cell in WT embryos. A. Scatterplots representing the fold change in gene expression ($\log_2$ scale) between *mir430*$^{-/-}$ + inj MiR430 and WT at 256-cell (y-axis), and the fold change in gene expression ($\log_2$ scale) between 1024-cell and 256-cell (left), Oblong and 256-cell (central) and Sphere and 256-cell (right) in WT embryos (x-axis). All upregulated genes are shown in gray, and genes induced at 1024-cell are highlighted in blue. The left graph is also shown in Fig. 3B. **B-D.** Same as in (A) but genes that are highlighted represent genes that are induced at Oblong **(B)**, Sphere **(C)**, or not induced during ZGA **(D)**.

Supplementary Figure 12

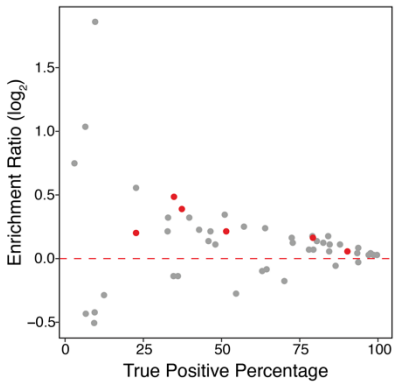

Fig. S12. Transcription factor motif analysis of upregulated genes. Scatter plot representing the Enrichment Ratio (log_2) of transcription factor motifs in the promoters (TSS ± 2Kb) of upregulated genes compared to the promoters of not expressed genes (y-axis), and the percentage of promoters of the upregulated genes that have the motif (x-axis). Only motifs whose corresponding protein is translated during early development (*30*) were considered, and only motifs with a p-value < 0.05 and E-value ≤ 10 are shown. Motifs corresponding to the three pluripotency factors Nanog, Sox19b and Pou5f3 (POU5F1, Pou5f1, Pou5f1::Sox2, POU5F1B, POU2F1, POU2F1::SOX2) are labeled in red.

Supplementary Figure 13

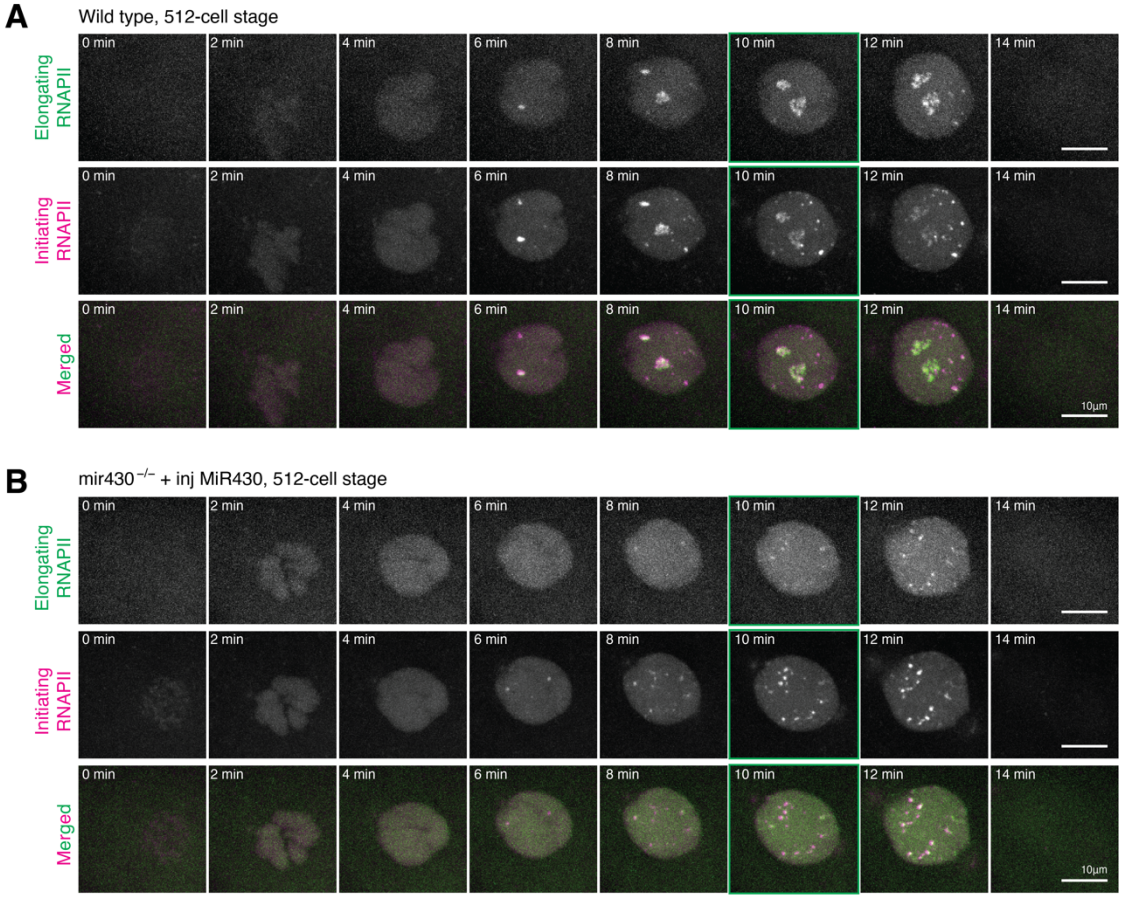

Fig. S13. Stalled transcription bodies go into productive transcription in *mir430*$^{-/-}$ + inj MiR430 embryos (512-cell stage). Visualization of initiating (Ser5P) and elongating (Ser2P) RNAPII with Fabs in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 512-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 14

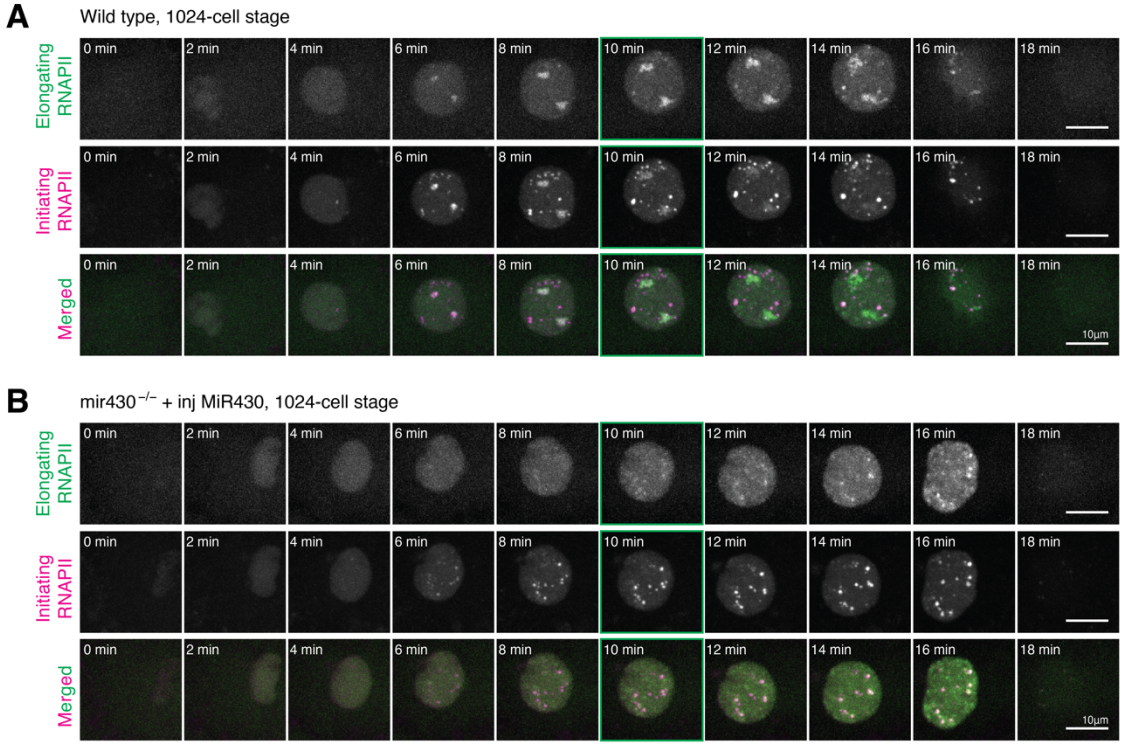

Fig. S14. Stalled transcription bodies go into productive transcription in *mir430*$^{-/-}$ + inj MiR430 embryos (1024-cell stage). Visualization of initiating (Ser5P) and elongating (Ser2P) RNAPII with Fabs in WT **(A)** and *mir430*$^{-/-}$ + inj MiR430 **(B)** embryos at 1024-cell stage. Shown are representative images of individual nuclei, extracted from a spinning disk confocal microscopy time lapse. The micrographs shown in the main figure are highlighted in green.

Supplementary Figure 15

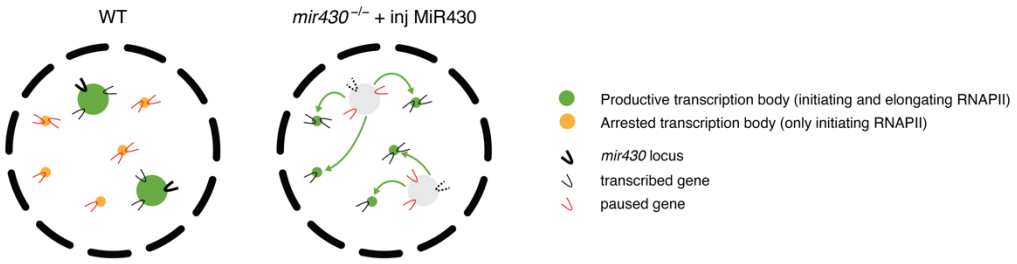

Fig. S15. Model for the role of transcription bodies in transcription regulation. In WT nuclei, two large transcription bodies are nucleated by the *mir430* locus. These facilitate the expression of genes that localize there. They also sequester components of the transcription machinery that are required for the transition from transcription initiation to elongation, thereby stalling ectopic transcription bodies in the initiation state. The specific disruption of the two *mir430* transcription bodies leads to a redistribution of these components which results in the downregulation of genes that localize to the *mir430* transcription bodies and the upregulation of genes elsewhere in the nucleus.

Supplementary Figure 16

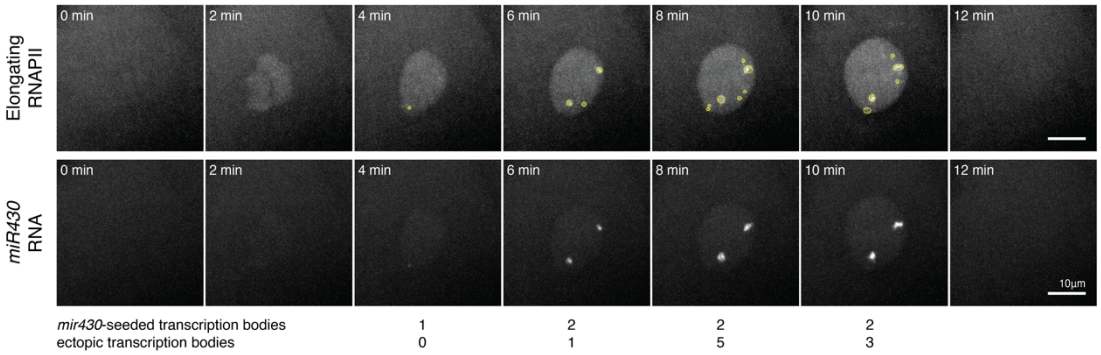

Fig. S16. Classification of *mir430* and ectopic transcription bodies. Transcription bodies are identified by detecting elongating RNAPII signal. If this overlaps with MOVIE signal (detecting *miR430* RNA) they are classified as *mir430*, and if they do not overlap with MOVIE signal, they are classified as ectopic transcription bodies. See Materials and Methods for more detail.