Pluripotency factors determine gene expression repertoire at zygotic genome activation

Meijiang Gao*[1,2], Marina Veil*[1], Marcus Rosenblatt[3], Anna Gebhard[1], Helge Hass[3], Lenka Buryanova[1], Lev Y. Yampolsky[4,5] Björn Grüning[6,7], Sergey V. Ulianov[8,9], Jens Timmer[2,3], Daria Onichtchouk**[1,2,10]

***contributed equally**

[1]Department of Developmental Biology, University of Freiburg, 79104 Freiburg, Germany

[2]Signalling Research Centres BIOSS and CIBSS, 79104 Freiburg, Germany

[3]Institute of Physics, University of Freiburg, 79104 Freiburg, Germany

[4]Department of Biological Sciences, East Tennessee State University, Johnson City, Tennessee 37614-1710, USA

[5]Zoological Institute, Basel University, Basel CH-4051, Switzerland

[6]Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany

[7]Center for Biological Systems Analysis (ZBSA), University of Freiburg, 79104 Freiburg, Germany

[8]Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

[9]Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, Russia

[10]Koltzov Institute of Developmental Biology RAS, 119991 Moscow, Russia

**Correspondence to: daria.onichtchouk@biologie.uni-freiburg.de

Abstract

Awakening of zygotic transcription in animal embryos relies on maternal pioneer transcription factors. The interplay of global and specific functions of these proteins remains poorly understood. Here, we analyzed nucleosome positioning, H3K27 acetylation and time-resolved transcription in zebrafish embryos lacking pluripotency factors Pou5f3 and Sox19b. We show that Pou5f3 and Sox19b modify chromatin in a largely independent manner. The bulk transcriptional onset does not require Sox19b and Pou5f3, but is sensitive to their balance. Pou5f3 docks H3K27ac on the enhancers of genes involved in gastrulation and ventral fate specification. Sox19b facilitates Pou5f3 access to a half of these enhancers. The genes regulating mesendodermal and dorsal fates are primed for activation independently of Pou5f3 and Sox19b. Strikingly, simultaneous loss of both factors leads to premature expression of differentiation genes. Our results uncover how independent activities of maternal Pou5f3 and Sox19b synergize or antagonize to determine the early gene expression repertoire.

Introduction

Following fertilization, the differentiated cells, egg and sperm, are reprogrammed into the totipotent state of the zygote. The zygotic genome initially remains silent. It awakens through a process known as maternal-to-zygotic transition (MZT), during which the degradation of maternal transcripts is coordinated with zygotic genome activation (ZGA). In the current model of ZGA, the gradual increase in the ratio of transcriptional activators to transcriptional repressors, accompanied with local changes of chromatin accessibility create a permissive environment for ZGA to occur[1]. In zebrafish, *Xenopus* and *Drosophila,* where development starts with rapid cell cycles, excessive maternal core histones serve as general transcriptional repressors before ZGA[2-4]. Several types of activators are translated before ZGA and reach critical levels at ZGA, including basal transcription factors[5], the regulators of H3K27ac enhancer mark[6] and maternal enhancer-binding transcription factors (TFs). TFs that broadly activate zygotically expressed genes have been identified in *Drosophila*[7], zebrafish, *Xenopus* and mammals[8]. In lower vertebrates, zygotic transcription is activated by homologs of mammalian pluripotency factors: Pou5f3, Sox19b and Nanog in zebrafish[9,10], Pou5f3 and Sox3 in *Xenopus*[11].

Nucleosome positioning plays a dominant role in regulating genome access by TFs. The widespread action of genome activators is thought to result from their ability to act as pioneer factors, first displacing nucleosomes so that other TFs can bind[12]. Indeed, reduction or loss of genome-activating TFs in *Drosophila*, zebrafish and *Xenopus* resulted in the decreased chromatin accessibility on their binding sites[11,13-16]. Out of them, direct pioneer binding to nucleosomes was demonstrated that far only for *Drosophila* genome activator Zelda[17]. The mechanisms underlying nucleosome-displacing activity of zebrafish and *Xenopus* activators are less clear: they may bind to nucleosomes similarly to their mammalian homologs[18], or compete with nucleosomes for DNA binding[19], or both.

Mammalian POU5F1 and SOX2 reprogramme somatic cells to pluripotency and are in several cases sufficient for reprogramming[20]. The mechanisms underlying their partnership *in vivo* are still not resolved. Until recently, POU5F1 and SOX2 were thought to act cooperatively, binding as heterodimers to bipartite *pou:sox* cognate motifs[21]. This view was challenged by Soufi *et al.*(2015), who demonstrated that POU5F1 and SOX2 target distinct motifs on the nucleosome-wrapped DNA[18], and by four studies that suggested different scenarios of how POU5F1 and SOX2 interact with each other and with chromatin in embryonic stem (ES) cells. These scenarios are: 1) assisted loading, whereby SOX2 first engages the target DNA, then assists the binding of POU5f1[22]; 2) negative reciprocity, where POU5F1 and SOX2 sometimes help and sometimes hinder each other in binding to the genome[23]; 3) conditional cooperativity of POU5f1 and SOX2 binding, depending on the motif positions in the nucleosomal context[24] and 4) independent binding, even at co-occupied sites[25].

In the case of zebrafish genome activators Pou5f3 and Sox19b, it remains an open question how their broad nucleosome-displacing activity at ZGA relates to their different functions later in development, as judged by their distinct loss-of-function phenotypes. Maternal-zygotic Pou5f3 null mutants MZ*spg* have abnormal epiboly and arrest during gastrulation[26]. The quadruple morpholino knockdown (QKD) of redundant SoxB1 family members (Sox19b, Sox19a, Sox3 and Sox2) leads to severe

defects during organogenesis, with the first morphological defects visible at the end of gastrulation[27]. The relatively late QKD phenotype is at odds with the earlier roles for SoxB1 proteins suggested by dominant-negative approaches[28], and by combined knockdowns of SoxB1 genes with Nanog and/or Pou5f3[9]. One possible reason for this discrepancy could be the presence of maternal Sox19b protein in SoxB1 QKD[10], which may mask the early requirement for Sox19b.

The mechanisms of SoxB1 activity at ZGA and its molecular connection to Pou5f3 remain poorly understood. In this study, we used maternal-zygotic (MZ) Sox19b, Pou5f3 and the double mutants to investigate how two zygotic genome activators interact *in-vivo*.

Results

Maternal-zygotic Sox19b mutants are delayed in gastrulation
To abolish the expression of Sox19b in zebrafish, we introduced a mutation in *sox19b* using gene disruption via TALEN[29] (Fig. 1a). The MZ*sox19b* embryos lacking both maternal and zygotic Sox19b, and M*sox19b* embryos lacking maternal Sox19b developed into fertile adults, albeit more slowly than controls (Movie S1-1, S1-2, S1-5).

The zebrafish midblastula transition (MBT) begins at cell cycle 10, at 3 hours postfertilization (hpf). MBT is characterized by cell cycle lengthening, loss of cell synchrony, activation of zygotic transcription (referred as the major wave of ZGA) and appearance of cell motility[30]. The duration of the pre-MBT cell cycles was the same the MZ*sox19b* and wild-type embryos (Fig.1b, Fig. S1). However, the appearance of morphological landmarks of subsequent development was delayed in MZ*sox19b* (Fig.1c). The time gap between MZ*sox19b* and wild-type development increased during gastrulation, with no further delay during organogenesis (Fig. 1c, Fig.S2, summarized in Fig.1d).

Normal development of MZ*sox19b* embryos can be plausibly explained by the presence of zygotic SoxB1 members, Sox19a, Sox3 and Sox2 (Fig.2a). The TALEN-induced *sox19b* mutation resulted in premature stop codon before the first intron of *sox19b*; nonsense mediated mRNA decay in this type of mutants can trigger compensatory response by upregulation of the genes that exhibit sequence similarity with the mutated gene's mRNA[31]. To investigate if the transcription of SoxB1 genes is changed in the MZ*sox19b* mutant, we quantified the levels of *sox19b*, *sox19a*, *sox3* and *sox2* by RNA-seq(Fig. S3a). *sox19b* maternal message was reduced 15-fold already before MBT, indicating that nonsense mediated decay takes place. Although *sox19a*, *sox3* and *sox2* bear the closest sequence similarity to *sox19b*, we did not detect compensatory upregulation of these genes which were rather delayed in MZ*sox19b* (Fig.S3a-c).

To address if maternal Sox19b protein masked early requirement for SoxB1 proteins in published SoxB1 quadruple knockdown (QKD) experiments[27], we injected Sox3, Sox19a and Sox2 morpholinos into MZ*sox19b* mutant embryos (triple knockdown, or TKD). MZ*sox19b*-TKD and wild-type- QKD embryos showed similar developmental defects in tailbud formation, anterior–posterior axis elongation and neural system development (Fig. 2b, Fig.S3d-f). The MZ*sox19b*-TKD phenotype could be completely rescued by co-injection of *sox19b* mRNA (Fig. 2c, Fig.S3g). We concluded that combined zygotic activity of Sox2/3/19a/19b proteins becomes critical for the embryo starting from the end of gastrulation. The developmental delay in MZ*sox19b* may reflect non-essential earlier requirement for maternal Sox19b. Sox19b may act redundantly with Nanog or Pou5f3, during major ZGA onset, as suggested previously[9].

Double MZ*sox19bspg* mutants are dorsalized but not delayed
To investigate the early requirements for maternal Sox19b and Pou5f3, we obtained a double mutant MZ*sox19bspg* by crossing MZ*sox19b* to Pou5f3 null-mutant MZ*spg*^{m793} [26]. MZ*spg*^{m793} mutants develop severe epiboly defects[32,33] and are weakly dorsalized[34]. Epiboly defects in double mutant were similar to MZ*spg* while

dorsalization was stronger. Unexpectedly, the double mutants were less delayed during the first part of gastrulation than the MZ*sox19b* (Fig.2d, Movie S1-2, 3, Fig. S3h). We produced maternal-only M*sox19bspg* mutants by fertilizing the mutant eggs with wild-type sperm and used these to determine if a combined maternal contribution of Pou5f3 and Sox19b is critical for developmental timing, epiboly, or dorso-ventral (D/V) patterning. Single maternal mutants were delayed in gastrulation (Movie S1-4, 5) but developed normally. The double mutant embryos were not delayed, but were severely dorsalized, as judged by radially expanded domains of dorsal markers *noggin1* and *chordin*[35,36] (Fig. 2e), and severe embryonic phenotypes (Fig. 2f and Movie S 1-4, 5).

In zebrafish, similar to other vertebrates, Chordin blocks the flow of BMPs to the dorsal side of the embryo. The action of multiple gene products within dorso-ventral self-regulatory network converges on defining the size of the Chordin domain[37 38]. Reduction of Chordin levels by morpholinos was sufficient to rescue M*sox19bspg* phenotype to normal (Fig. 2g). This result suggests combinatorial action of maternal Pou5f3 and Sox19b, i.e. two TFs act additively inducing ventral regulators, and/or repressing dorsal regulators. The combined changes override the self-regulatory capacities of dorso-ventral gene network, while changes in single mutants can be buffered (Fig. 2h). Apart of that, Pou5f3 becomes essential at the beginning of gastrulation, and SoxB1 factors at the end (Fig. 2i). Phenotypic suppression of developmental delay in MZ*sox19bspg* may indicate compensatory relationships between Pou5f3 and Sox19b, which were not appreciated previously.

Bulk transcription at ZGA onset is sensitive to the Sox19b/Pou5f3 balance

To characterize the mutant transcriptomes, we performed time-resolved RNA-seq analysis of wild-type, MZ*sox19b*, MZ*spg* and MZ*sox19bspg* embryos. The embryos were collected starting from 2.5 hpf (pre-MBT) every 30 min until 6 hpf (Fig. 3a).

Two processes shape the transcriptional landscape of embryos at ZGA: the burst of zygotic transcription and regulated decay of maternal mRNAs. About 70% of zebrafish zygote mRNAs are maternally loaded[39], so that the mRNA present in the embryo early time points is a mixture of maternal and zygotic transcripts for most genes. To account for the maternal and zygotic differences between the wild-type and mutants, we developed a novel tool for dynamic RNA-seq data analysis, which we called RNA-sense. The RNA-sense 3-step analysis is explained in the Movie S2 and the Methods. At the first step, RNA-sense builds a time profile for each expressed transcript in one condition (i.e. wild-type), and tests if the transcript abundance grows or decays significantly. Dynamic transcripts are then sorted to non-overlapping groups by the time point of switch UP, or switch DOWN. For all the genes in switch UP groups, zygotic increase in the transcript levels exceeds maternal RNA decay.

The switch UP and switch DOWN groups in the wild-type (Table S1) were in agreement with the zygotic and maternal transcript groups identified in three previous studies[9,39,40](Fig. S4a), and are referred below as “zygotic” and “maternal”.

Analysis of differential expression revealed delays in zygotic transcription and maternal mRNA degradation in all mutants (Fig. 3b, Fig. S4b-g). Unexpectedly, we detected stronger effects in single mutants than in the double mutant. Out of 4643

zygotic transcripts, 51% were at least two-fold downregulated in MZ*sox19b* and 32% in MZ*spg,* versus 24% in the double mutants (Fig. 3b, see also Fig. S4b-g).

We compared the bulk zygotic transcription rise and bulk maternal RNA decay across the genotypes, using the same scale for all transcripts, log2 ratio of expression level to pre-MBT (2.5 hpf). As early as 3.5 hpf, zygotic transcription differed as WT > MZ*sox19bspg* > MZ*sox19b* = MZ*spg* (Fig. S5a). This result implied that Pou5f3 and Sox19b compensate some of each other's effects on the zygotic transcription already at the beginning of major ZGA. At 5 hpf, zygotic transcription and maternal RNA decay were at strongest delayed in MZ*sox19b* (Fig. S5a, b).

Taken together, our data suggested unanticipated antagonism between two pluripotency factors. The balance of Pou5f3 and Sox19b appeared to have stronger effects on zygotic transcription, than simply the presence of both factors in the embryo.

Sox19b and Pou5f3 activate ventral genes and are dispensable for dorsal genes

We divided all zygotic genes to 8 groups, A-H, based on the patterns of downregulation in the mutants (Fig.3b, Table S1, Fig.S6 shows selected examples). Groups A and B were enriched in distinct developmental regulatory gene categories (Fig. 3c) which we discuss below; the other groups were enriched in general housekeeping categories or showed no enrichment.

Group A transcripts (17% of zygotic genes, downregulated in all mutants) included major regulators of ventro-posterior development: BMP pathway ligands *bmp2b*[36] and *bmp4*[41]*,* the ventral transcription factors *vox, vent* and *ved*[42-44], regulators of ventro-lateral mesoderm and ectoderm patterning *draculin*, *gata2*, *tbx16* and *vgll4l*, which start their expression at the ventral side or in the ectoderm[45]. Group A also included known direct transcriptional targets of Pou5f3: *mych*[46] and *klf17*[47]*.* Additional genes related to the BMP pathway were repressed in the double mutant and in one of the single mutants: the second major BMP ligand *bmp7a*[48] was repressed in the double mutant and MZ*sox19b* (Group D), BMP targets *klf2b*[47] and *foxi*[49] in the double mutant and MZ*spg* (group F).

Group B transcripts (8% of all zygotic genes, downregulated only in the single mutants) were enriched in the category "endoderm development". Group B included the main critical components of the early endomesoderm and dorsal specification network: *mixl1*[50], *sox32*[51], *sebox*[52] *dusp4*[53], *Brachyury* homolog *tbxta*[54], and *foxa3*[55], which are activated by the maternal transcription factors Eomesodermin, FoxH1 and by Nodal signaling[56-58]. Group B also included double Nodal and maternal β-catenin targets *chordin*[38] and *noto*[59], and the maternal β-catenin target *hhex*[60]. In contrast to group A, group B transcripts start their expression at the dorsal side of the embryo, or in the yolk syncitial layer (YSL).

To characterize the earliest effects of Pou5f3 and Sox19b in all groups, we tested if the expression in the wild-type was significantly different from the mutants from 3 to 4.5 hpf (Fig. 3d, Fig. S5d-f). The group A transcripts were significantly downregulated in all mutant genotypes starting from the beginning of the major ZGA wave. The downregulation was non-additive: the transcripts levels were not statistically different in three mutant genotypes until 4.5 hpf (Fig.3d). We concluded that Pou5f3 and

Sox19b are required for activating group A zygotic transcription on the ventral side and act either redundantly or sequentially.

The levels of group B transcripts were significantly downregulated from at 3 to 4.5 hpf in the single mutants but not in the double mutants (Fig. 3d). We concluded, that group B transcripts, do not require Pou5f3 and Sox19b for their expression, but are sensitive to Pou5f3/Sox19b balance starting from ZGA onset. Similar analysis of the groups C-F (Fig. S5d-f) suggested non-additive and compensatory effects of Pou5f3 and Sox19b on early transcription.

Ventral (group A) zygotic transcripts were expressed to significantly higher level then dorsal (group B) in the wild-type embryos at 3-4.5 hpf (black in Fig. 3e,f). In contrast, in the double mutants expression of dorsal transcripts was higher than that of the ventral (purple in Fig. 3e,f). We wondered if the severe dorsalization phenotype of the double mutants could be explained by the increase in the ratio of *chordin (*group B) to *bmp2b (*group A*)* and *bmp7a* zygotic transcripts at ZGA onset. Indeed, the transcription of BMPs was reduced relative to Chordin in the MZ*sox19bspg* embryos, but not in the single mutants, from the beginning of major ZGA (Fig. 3g). The ratio of Chordin to BMPs is critical for the size of dorsal domain[37], hence, the early transcriptional disbalance between these factors in MZ*sox19bspg* embryos plausibly explained their phenotypic dorsalization.

In sum we demonstrated, that Pou5f3 and/or Sox19b activate 24% of zygotic transcripts (groups A,D,F,G) including but not restricted to the components of BMP signaling pathway, ventral genes and ectodermal genes. Pou5f3 and Sox19b are dispensable for the activation of mesendodermal regulators, targets of Eomesodermin/FoxH1/Nodal, starting their expression on the dorsal side of the embryo.

Differentiation genes are prematurely expressed in MZ*sox19bspg*.
We next inquired if Pou5f3, Sox19b or both of them suppress zygotic transcription for certain genes. We scored zygotic transcripts which are upregulated in the mutants (Table S2), and divided them to three non-overlapping groups (I-K) by the strongest upregulation in one of the genotypes (Fig. 4a). Transcripts upregulated in MZ*sox19bspg* (Group I) were enriched in developmental and regulatory ontologies (Fig. 4b) and included two categories: transcription factors such as *pax8, dlx5a, nkx6.3, tbx2b*, *tbx2b* and *tbx6* that are involved in tissue differentiation and were normally expressed in the wild-type starting later than 6 hpf[45], and dorsal genes like *nog1, noto* and *hhex,* which were expressed in the wild type at lower levels. In MZ*spg* and in MZ*sox19bspg*, but not in MZ*sox19b*, selected transcripts were upregulated starting from the ZGA onset, as compared to the wild-type (Fig. 4c,d, Fig.S5h, Fig. S6e-g for examples).

In sum we show that Pou5f3 and Sox19b together prevent the premature expression of late differentiation genes after major ZGA onset. Together with early bias in the dorsal to ventral transcription, this effect contributes to the change of zygotic gene expression repertoire in the double mutants.

Sox19b and Pou5f3 act as independent pioneer factors
In order to connect the effects of Pou5f3 and Sox19b on transcription with their effects on chromatin, we chose 4.3 hpf time point in the middle of transcription time

curve, before the end of MZT, to make a snapshot of chromatin state in the single mutants. Choosing 4.3 hpf enabled us to profile chromatin marks, which are low in earlier stages. Throughout the analysis, we focused on the genomic regions previously shown to be bound by Pou5f3 and/or SoxB1 by ChIP-seq[10]. We divided these regions into three groups: Pou5f3-only (P), SoxB1- only (S), and SoxB1 - Pou5f3 (SP) peaks (Table S3), and ranked them by descending TF occupancy, as shown at Fig.5a.

To profile nucleosomes, we performed MNase-seq experiment on 4.3-hpf MZ*sox19b* mutants and compared MNase signals on P, S and SP peaks in MZ*sox19b* with MZ*spg* and the wild type[16]. The strongest nucleosome displacement in both mutants occurred at SP peaks and it was proportional to TF occupancy (Fig. 5c,d). Weaker changes were observed at P and S peaks in both genotypes (Fig. 5d, Fig. S7a). We concluded that Sox19b binding contributes to establishment of chromatin accessibility during MZT, confirming the results of a previous study[13].

We next asked which binding cues contribute to nucleosome displacement by each TF. SoxB1 and Pou proteins can recognize their consensus *sox* or *pou* motifs, respectively, or bipartite *pou:sox* motifs, which they are thought to bind together[61] (Fig. 5e, Table S3). Apart of that, TFs occupy DNA with specific shape ("shape motifs") regardless of whether or not these correspond to high information content sequence motifs. This phenomenon was called non-consensus binding and is not fully understood[62,63]. We have previously shown that Sox19b and Pou5f3 bind to high nucleosome affinity regions (HNARs), featuring high predicted DNA shape parameter propeller twist (PT°) values and high *in vitro* predicted nucleosome occupancy[16]. To characterize the possible changes in nucleosome landscape caused by Sox19b and Pou5f3 consensus and non-consensus binding, we included sequence-specific motifs and non-consensus binding cues in our analysis. Fig. 5b and Fig. S7b show enrichment for non-consensus binding cues on Pou5f3 and SoxB1 binding regions. The example genomic region of *pou5f3* gene shows distribution of non-consensus binding cues (HNARs), consensus binding motifs, TF binding and nucleosome displacement (Fig.5f).

We found, by comparing nucleosome displacement on the SP peaks with or without motifs of each type, that nucleosome displacement in MZ*sox19b* depends mostly on *sox* motifs, and in MZ*spg* on *pou* and *pou:sox* motifs (Fig. 5g,h). We concluded, that two factors modify nucleosome landscape by binding to separate motifs. Analysis of S and P peaks supported this conclusion (Fig. S7c).

To estimate non-consensus binding effects of Sox19b and Pou5f3 on chromatin, we ranked the SP genomic regions by ascending PT° values to four quartiles, from low (q1) to high (q4), and compared the mean nucleosome displacement values. Nucleosome displacement in MZ*sox19b* mutant increased proportionally to PT°, while nucleosome displacement in MZ*spg* showed a weak opposing tendency (Fig.5i). This effect was not restricted to the TF-binding regions and occurred genome-wide, also on the random control regions (Fig. 5j, see also Fig. S7d-f). We propose, but we have not proven, that opposing genome-wide changes of nucleosome landscape in MZ*sox19b* and MZ*spg* may be due to the widespread non-consensus binding of the TFs (see Discussion).

In sum, our results suggest that Sox19b and Pou5f3 are both involved in establishment of chromatin accessibility during MZT. They act as independent pioneer factors, displacing nucleosomes on different consensus motifs even on co-occupied (SP) regions.

Pluripotency factors regulate H3K27 acetylation on three enhancer types.
In parallel with the increase of chromatin accessibility during the ZGA, zebrafish genome acquires H3K27ac histone tail mark on active enhancers[6,64]. To test if H3K27ac deposition depends on Pou5f3 and Sox19b, we immunoprecipitated chromatin from wild-type, MZ*spg* and MZ*sox19b* 4.3 hpf embryos with H3K27ac antibodies, sequenced and compared the mapped data (Fig. S8a,b).

To infer the putative enhancer activity changes in the mutants, we selected Pou5f3 and/or Sox19b-binding peaks that were flanked by H3K27ac in at least one genotype, and then clustered them by H3K27ac signal in three genotypes. This resulted in four clusters of enhancers, named by downregulation of H3K27ac in the mutants (Table S4): CD = codependent, PD = Pou5f3-dependent, SD = Sox19b-dependent, and U = unchanged (see the heatmap in Fig.6a and examples in Fig.6d). In wild-type embryos, the CD cluster was at most and the SD cluster at least H3K27 acetylated (Fig.6b). H3K27 acetylation on some SD enhancers was reduced in the absence of Sox19b, but increased in the absence of Pou5f3 (Fig.6c).

CD, PD and SD enhancers differed by associated gene ontologies (Fig. 6d), by the extent of nucleosome displacement in the mutants, and by density of consensus motifs (Fig. S8c-h). We found, that both H3K27 acetylation and nucleosome displacement on PD enhancers depended on Pou5f3 binding to *pou:sox* and *pou* motifs (Fig. S8i). On SD enhancers, both H3K27 acetylation and nucleosome displacement depended on Sox19b binding to *sox* motifs (Fig. S8j). On codependent (CD) enhancers, Pou5f3 effects depended on *pou* and *pou:sox* motifs, while Sox19b effects were non-sequence specific (Fig.6f,g). The most parsimonious explanation for the last observation is that Sox19b non-consensus binding competes with histones and facilitates Pou5f3 loading on CD enhancers. Pou5f3 binding to *pou* or *pou:sox* motifs promotes local H3K27 acetylation.

In sum, we delineated three enhancer types by differential regulation of the H3K27ac mark and nucleosome displacement in MZ*spg* and MZ*sox19b* mutants: CD (codependent), PD (Pou5f3-dependent) and SD (Sox19b-dependent).

PD and CD enhancers are associated with the major ZGA wave.
Our next step was to address whether H3K27ac changes at three regulated enhancer types correlate with transcriptional outcomes in the wild-type and mutants. We first tested if CD, PD or SD enhancers were overrepresented in the putative regulatory regions of the previously defined transcript groups (zygotic A-K and maternal, Fig. S9a-d). CD and PD, but not SD enhancers were overrepresented in the group of zygotic genes downregulated in all mutants (Fig.7a). SD enhancers were overrepresented in the group of zygotic genes upregulated in MZ*spg* (Fig. 7b), in agreement with their increased H3K27 acetylation in MZ*spg*. We next grouped the zygotic transcripts by enhancer presence and type (Table S5). Example genes linked to 3 or more enhancers of the same type are shown in Fig. 7c (genomic regions), and Fig. 7d (transcription profiles). CD- and PD-linked genes were expressed early in the wild type and downregulated or delayed in all mutants. SD-linked genes *vrtn* and

pcdh18b (direct SoxB1 target[27]) were expressed late in the wild type and showed compensatory effects in the mutants in that they were upregulated in MZ*spg*, downregulated in MZ*sox19b* and expressed at the intermediate level in MZ*sox19bspg*. To assess if these differences could be generalized, we compared the mean relative expression profiles. As shown in Fig. 7e,f, transcripts linked to PD and CD groups were downregulated in MZ*sox19bspg* starting from ZGA onset. PD-linked transcripts were the earliest to be expressed in the wild-type (Fig.7e,g). SD-linked transcripts showed compensatory effects in the mutants, in that they were significantly downregulated only in MZ*sox19b*, but not in MZ*sox19bspg*. We concluded that Pou5f3-dependent and -codependent enhancers, but not Sox19b-dependent enhancers are, as a rule, involved in the transcriptional activation by TFs at major ZGA onset.

Discussion

Gene expression starts as a result of a multi-step regulatory programme involving a rapid succession of changes in chromatin accessibility, histone modifications and transcription[65]. In this study, we dissected the contributions of Pou5f3 and Sox19b to each of these three steps during ZGA, which yielded five novel mechanistic and developmental insights.

First, on the mechanistic level, we found that Sox19b and Pou5f3 modify chromatin independently, even at the regions they bind together. This result is in agreement with a recent study of chromatin accessibility in ES cells, which found that the mammalian pluripotency factors POU5F1 and SOX2 operate in a largely independent manner even at co-occupied sites, and that their cooperative binding is mostly mediated indirectly through regulation of chromatin accessibility [25].

Second, we show that Pou5f3 sequence-specific binding activates early zygotic genes by increasing chromatin accessibility and promoting H3K27 acetylation on Pou5f3- dependent and -codependent (CD) enhancers (Fig. 6g, Fig. 7). We made a snapshot of chromatin state at 4.3 hpf, when PD enhancer-linked transcription was already active, but CD enhancer-linked transcription was just switching on (Fig. 7e,g). The reduction of H3K27ac on CD enhancers in MZ*sox19b* could be transient, and could be due to delayed access of Pou5f3 to its binding sites. We speculate that Sox19b assists Pou5f3 loading on CD enhancers (Fig. 8a). Sox19b protein is highly abundant in the early embryo[9], and it is possible that high frequency of non-specific Sox19b–DNA interactions may, for instance, increase nucleosome turnover and facilitate Pou5f3 access to chromatin. It is also possible that Sox19b facilitates Pou5f3 binding at the distance, by allosteric mechanism which was recently proposed for Sox2[66].

Third, our results strongly suggest that the establishment of transcriptional competency at the onset of major ZGA is regional, mediated by Pou5f3 and Sox19b at the ventral side of the embryo, and independent of Pou5f3 and Sox19b at the dorsal side (Fig. 8b). This view is supported by several observations.1) Pou5f3 and Sox19b promote zygotic expression of the regulators of ventral development (Fig.3, group A). 2) Pou5f3 and Sox19b are dispensable for the zygotic activation the targets of Eomesodermin/FoxH1/Nodal and maternal β-catenin pathways which are first expressed on the dorsal side of the embryo (Fig 3, group B). 3) In the double MZ*sox19bspg* mutant, the zygotic transcription ratio of “ventral” group A to “dorsal” group B is reversed (Fig. 3d,e,g), resulting in the dorsalized phenotype (Fig. 2e,f,g). The most likely candidates for priming dorsal and mesendodermal genes for activation during the major ZGA wave are maternal transcription factors Eomesodermin, FoxH1 and Nanog[67,68]. The idea of independent zygotic activation of mesendodermal genes is compatible with recent studies in mouse and *Xenopus*. In *Xenopus*, Foxh1, together with two other maternal factors, binds to the endodermal enhancers prior to ZGA and primes them for zygotic activity[69]. In mouse, Eomesodermin and Brachyury are the sole factors responsible for establishing the competence for activation of mesodermal enhancers, and they repress the pluripotency and neuroectodermal programs driven by Pou5f1 and Sox2[70].

Fourth, we observed an unanticipated antagonism of Pou5f3 and Sox19b at major ZGA onset (Fig. 8c). The transcriptional changes in each of MZ*sox19b* and MZ*spg*

single mutants exceeded those in the MZ*sox19bspg* (Fig.3b). This was in odds with the results of of Lee et al.,[9] who reported that Sox19b loss-of-function (LOF) has the mildest effects on transcription, compared with combined LOFs. We explain this discrepancy by the presence of maternal Sox19b protein in SoxB1 QKD which was used in this study. ZGA timing is defined by competition between histones and transcription factors[3], and we expected that ablation of two major activators would additively delay the major ZGA wave. We observed exactly the opposite: delays in zygotic transcription, most prominent in MZ*sox19b,* were compensated in the double mutant. Taking the compensatory transcriptional changes together with opposing genome-wide changes of the nucleosome landscape in the single mutants (Fig. 5j, Fig. S7), we hypothesize here that maternal Sox19b may restrict non-consensus binding of Pou5f3 and *vice versa*; e.g. in MZ*sox19b*, Pou5f3 binds ectopically and changes nucleosome positioning across the genome, causing broad transcriptional delay. Two TFs may compensate each other's effects across the genome by *i.e.* promoting displacement of the same nucleosome in the opposing directions, or by recognizing different shape motifs within the same nucleosome footprint. Indeed, mammalian Sox2 and Pou5f1 bind on the different locations to nucleosome-wrapped DNA *in-vitro*: Pou5f1 binds DNA at the entry- or exit- positions of the nucleosomes, SoxB1 proteins tend to bind closer to the nucleosome dyad [24,71]. Our hypothesis is in agreement with a theoretical model, originating from analysis of Sox2 and Pou5f1 binding in ES cells[23]. The model predicts that two factors hinder each other's binding across the genome in the majority of cases (e.g. genome-wide), and help each other only in the minority of cases (e.g. on codependent enhancers).

Fifth, we show that Pou5f3 and Sox19b control the order of transcriptional onset for zygotic genes, by shutting down the expression of genes which elicit later developmental programs. Early studies of Alexander Neyfakh described two distinct periods of zygotic gene function in teleost fish[72,73]. The genes expressed from mid to late blastula (first period) provided the instructions for gastrulation. The genes expressed starting from midgastrula (second period) provided the instructions for organogenesis. We show that maternal Pou5f3 and Sox19b ensure the proper time gap between the first and second period in two situations. First, Pou5f3 suppressed H3K27 acetylation of some SD enhancers and premature transcription of associated genes (Fig. 8d). Similar phenomenon was recently documented in mouse ES cells, where Pou5f1 suppresses Sox2- dependent enhancers of neural differentiation genes[25]. Second, the group of transcripts enriched in regulators of differentiation and patterning for all tissues is synergistically induced in the absence of both Pou5f3 and Sox19b at ZGA onset, thereby significantly changing the early gene expression repertoire (Fig. 8e).

Balancing of early and late developmental programs by Pou5f3 and Sox19b directly parallels reprogramming in mouse fibroblasts, where Pou5f1 and Sox2 shut down the somatic gene expression programs and activate early embryonic genes[74]. The mechanism by which this is achieved is currently unclear. Answering this question will require identification of TFs which activate transcription of the late genes in the absence of Pou5f3 and Sox19b. As the activation may be region-specific, these studies will require a combined single-cell analysis of transcription and enhancer accessibility.

Acknowledgments
We are grateful to Sebastian Arnold, Rainer Duden and Christina Gross for commenting on the manuscript, to Sabine Götter for excellent fish care, and to Andrea Buderer and Cornelia Wagner for administrative support. This work was supported by DFG-ON86/4-2 for DO, DFG-EXC2189 – Project ID: 390939984 for DO and JT, and RFBR 20-54-12022 for SU.The Freiburg Galaxy Team is funded by DFG grant SFB 992/1 2012 and BMBF grant 031 A538A RBC**.**

Author Contributions
GM, MV, AG and LB and SU performed the experiments; GM, BG, LY and DO analysed the data, MR and HH wrote RNA-sense program, DO - design of the study, JT and DO – supervision, funding acquisition, MV and DO wrote the manuscript, GM, MR, LY and JT edited the manuscript.

Declarations of Interests
The authors declare no competing interests.

Figure titles and legends

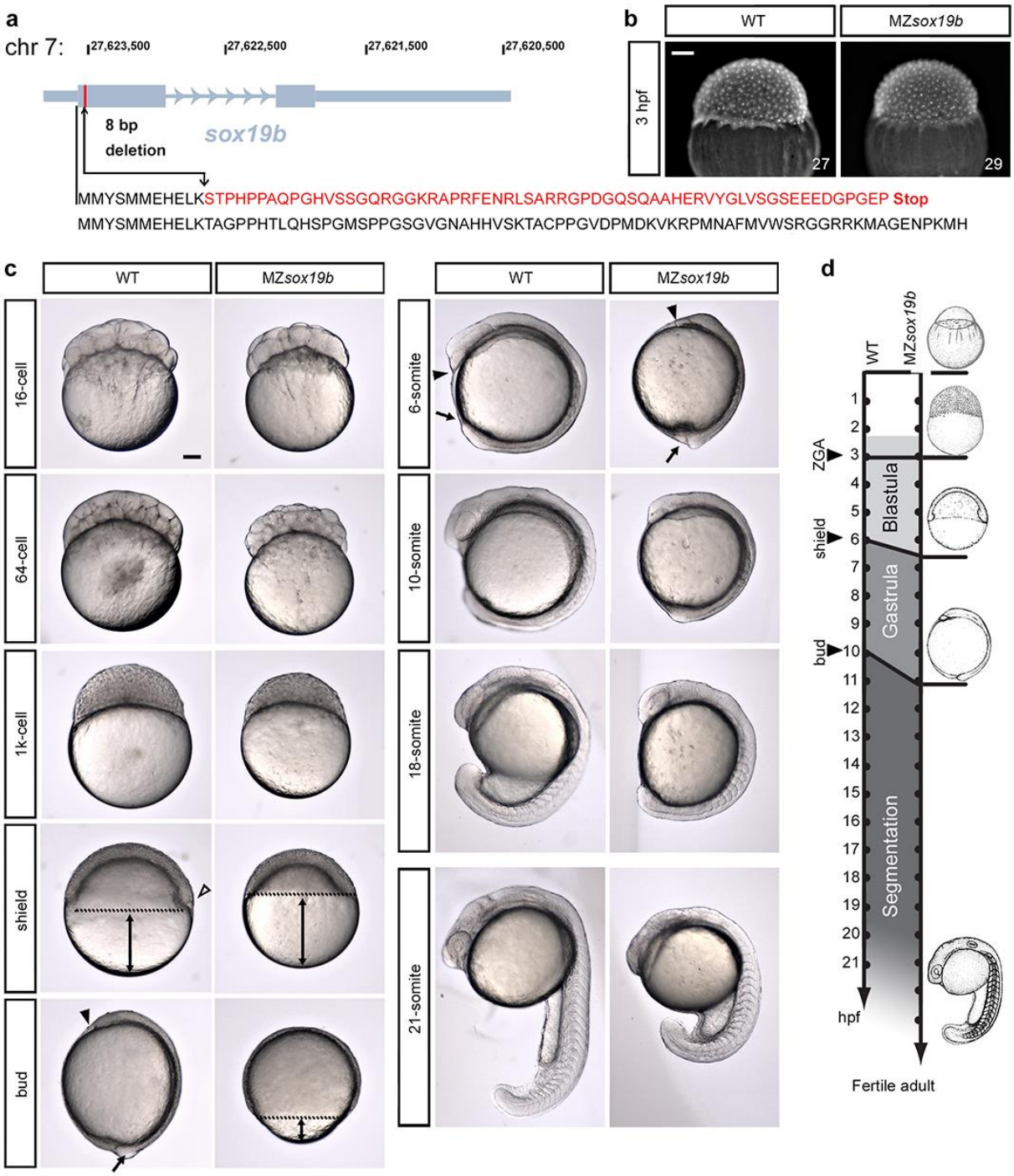

Figure 1: Developmental delay in MZ*sox19b* mutants. a. Disruption of the *sox19b* gene on chromosome 7 by introducing an 8 bp deletion. **b**. No difference in cell division rates between WT and MZ*sox19b* was observed prior to MBT (see also Fig. S1). **c**. MZ*sox19b* embryos are delayed in gastrulation. Simultaneously collected WT and MZ*sox19b* embryos were let to develop at 28,5°C, pictures of representative embryos were taken at the indicated time points. The developmental stage of the wild-type is indicated. In zebrafish embryos, embryonic shield forms at 6 hpf at the dorsal side (shield stage, hollow arrowhead in the wild-type). MZ*sox19b* embryos are still phenotypically at 40% epiboly (blastula stage). Double black arrows show the distance from epiboly border (black dotted line) to the vegetal pole. Gastrulation ends

with tail bud formation at 10 hpf (black arrow - tailbud, black arrowhead- head process at the bud stage in the wild-type). MZ*sox19b* embryos are still at 80-90% epiboly stage. Black arrow – tail bud, black arrowhead- head process. Scale bar 200 µm. **d**. Scheme of the altered developmental timing in MZ*sox19b*, based on multiple observations (see also Fig. S2).

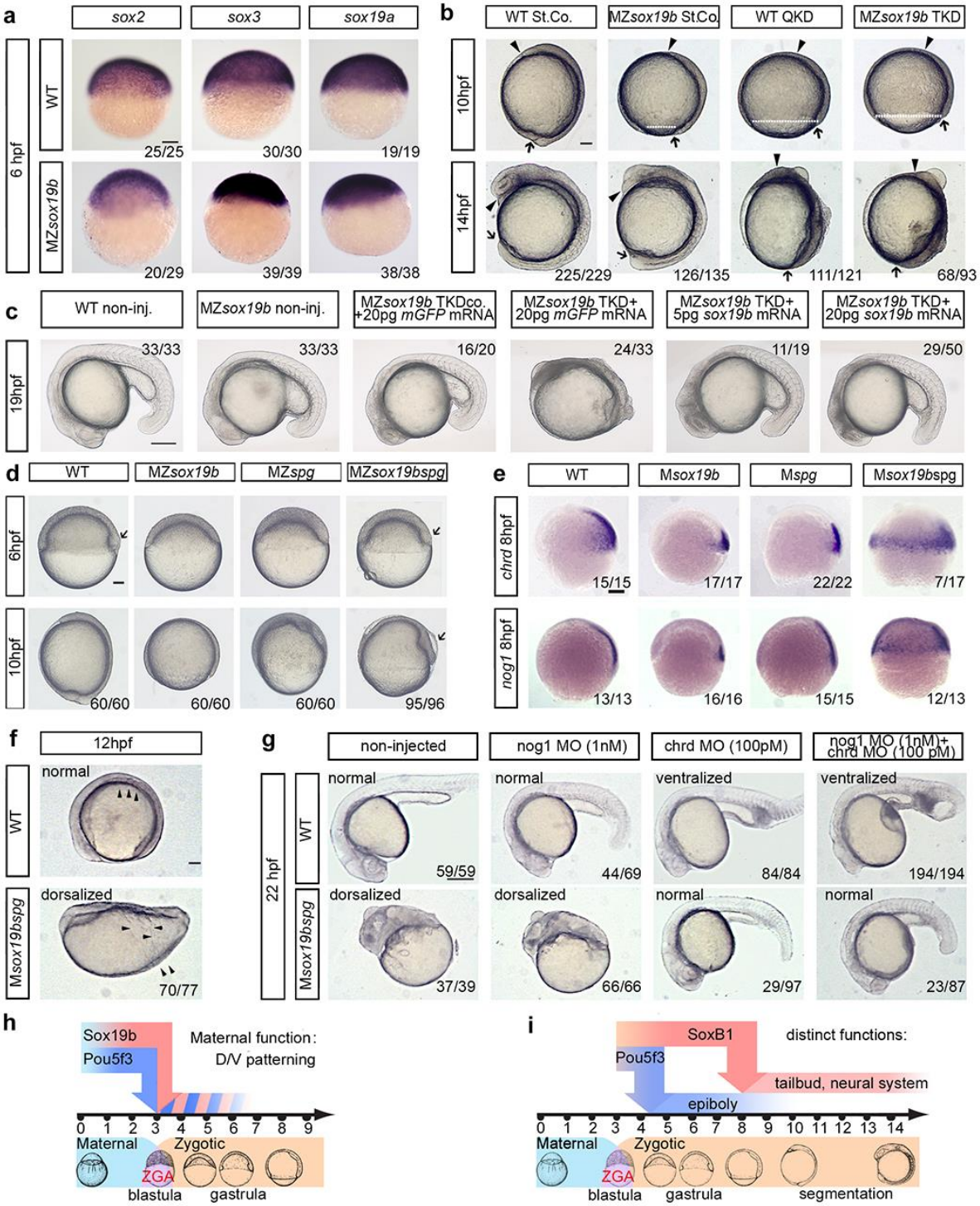

Figure 2: Common and separate biological functions of Pou5f3 and SoxB1 factors. a. *In situ* hybridization for *sox2, sox3 and sox19a*, lateral views. **b.** SoxB1 knockdown embryos complete gastrulation, but show later defects in tailbud formation and axis elongation. Arrow - tailbud position; arrowhead -head position; white dotted line - epiboly border. 1-cell stage wild-type or MZ*sox19b* embryos were injected with control morpholino (StCo), or QKD (quadruple knockdown) mix (Sox2, Sox3, Sox19a and Sox19b morpholinos), or TKD (triple knockdown, Sox2, Sox3, Sox19a morpholinos), as indicated. **c.** Axis elongation defects in MZ*sox19b*-TKD can be rescued by injection of *sox19b* mRNA. 1-cell stage MZ*sox19b* embryos were

injected with either TKD mix or TKDco mix (Sox2, Sox3, Sox19b Morpholinos), together with Sox19b or control GFP mRNA, as indicated (see Fig 3d-g for additional statistics). **d**. Comparison of the single mutants, double mutant and wild-type embryos. 6 hpf: shield formation in the double mutant (arrow) is not delayed in MZ*sox19bspg*. 10 hpf: MZ*sox19bspg* mutants are arrested in gastrulation similarly to MZ*spg*. Arrow shows abnormally enlarged shield in MZ*sox19bspg*. **e-f.** Double maternal mutants M*sox19bspg* are dorsalized. **e** *In-situ* hybridization for dorsal markers *noggin1* and *chordin*, lateral views, dorsal to the right. Note the circumferential expansion in the M*sox19bspg*. **f.** Somites (arrowheads*)* form on the dorsal side in WT, but spread over the M*sox19bspg* embryo. Dorsal up, anterior to the left. **g.** Normal development of M*sox19bspg* mutants can be rescued by reducing Chordin, but not Noggin levels. The wild-type or M*sox19bspg* embryos were injected with the indicated morpholinos or non-injected. The numbers show the ratio of embryos with indicated phenotype/ all embryos alive at 22 hpf. The arrows show abnormally expanded blood progenitor cells in the ventralized wild-type embryos. Anterior to the left, dorsal up. **h,i.** Combinatorial and distinct functions of Pou5f3 and SoxB1. **h.** Maternal Sox19b and Pou5f3 safeguard correct D/V patterning. **i.** Distinct functions: Pou5f3 is critical for epiboly and gastrulation, redundant action of zygotic SoxB1 (Sox19a, Sox19b, Sox2 and Sox3) is critical for organogenesis. Scale bars in a,b,d,e,f: 100 µm, in c,g – 200 µm.

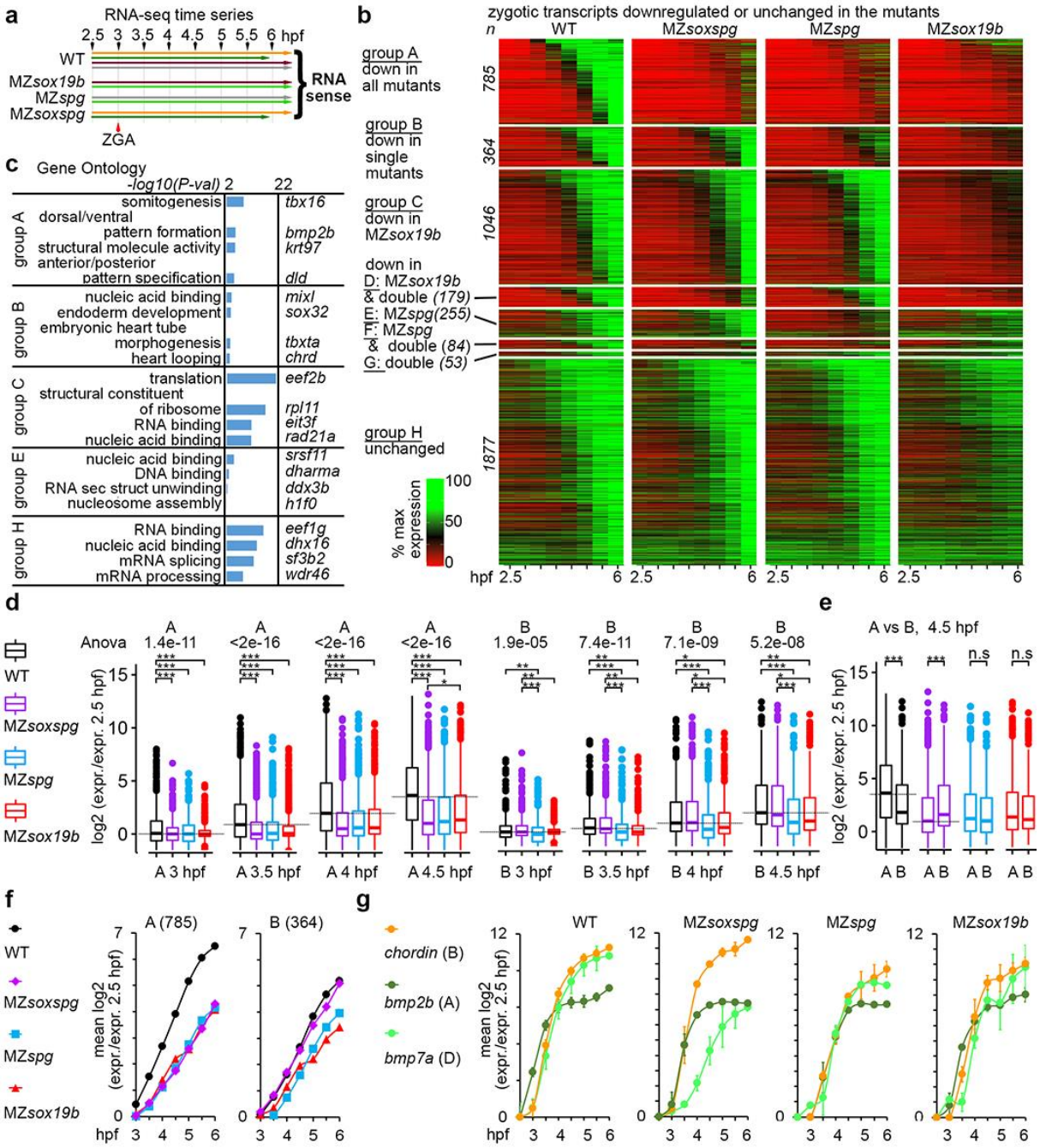

Figure 3: Pou5f3 and Sox19b balance the zygotic transcription of two groups of developmental genes. a. Experimental setup: RNA-seq time series included 8 time points, 4 biological replicates for the wild-type and 2 biological replicates for each MZ*spg*, MZ*sox19b*, and double mutant MZ*sox19bspg*. Time series, where material was collected in the same experiment, have the same color. **b.** Heatmap of all 4643 zygotic transcripts in the indicated genotypes. 2766 transcripts were downregulated in the mutants (groups A-G),1877 transcripts were not (group H). Zygotic transcripts were sorted by ascending switch time and switch p-value in the wild-type. n- number of transcripts. **c.** Enrichment in Gene Ontology terms (DAVID). Top four categories per group and example genes are shown. GO: enrichment for "structural molecule activity" in group A was due to the battery of 8 keratins, activated by Klf17 in the epithelial layer[47]**. d.** Non-additive (group A) and compensatory (group B) effects of Pou5f3 and Sox19b mutations on the earliest zygotic transcription. Box plots show

the distribution of zygotic expression levels at 3-4.5 hpf relative to expression at 2.5 hpf. Only significant pairwise differences are shown. WT median - dotted line, p-values in Tukey-Kramer test (* $p<0.05$, ** $p<0.005$, *** $p<0.0005$). **e.** Disbalance of zygotic transcription in MZ*sox19bspg*: A-group transcripts are expressed in a higher level than B-group transcripts in the wild-type, reverse is true for the MZ*sox19bspg*. Box plots for 4.5 hpf. Median - dotted line. p-values in Student's t-test, (*** $p<0.0005$, n.s-$p>0.05$). **f.** Mean zygotic transcription profiles for the groups A and B, relative to 2.5 hpf. Note that group A has the earliest and steepest expression curve in the WT, but not in MZ*sox19bspg.* **g.** Chordin/BMPs ratio is increased in MZ*sox19bspg*, but not in the single mutants throughout the time curve. Note that in the double mutants the transcription start of *bmp2b* and *bmp7a,* but not *chordin* is delayed in MZ*sox19bspg* compared to the WT.

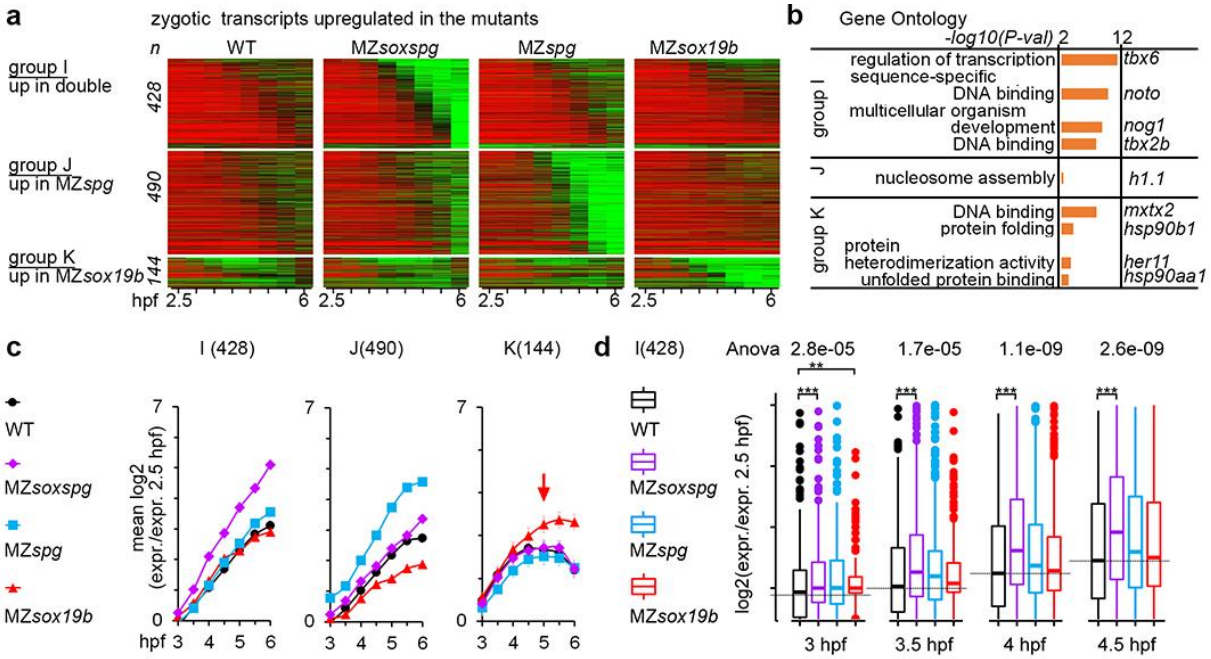

Figure 4: Simultaneous loss of Pou5f3 and Sox19b leads to increase in dorsal genes and premature expression of differentiation factors.
a. Heatmap for 1062 zygotic transcripts, upregulated in the mutants compared to the wild-type. The upregulated transcripts were grouped as indicated in the left (groups I, J, K), and sorted by ascending switch time and switch p-value in the respective mutant. **B.** The groups I,J,K were tested for enrichment in Gene Ontology terms using DAVID. Top four categories per group /example genes are shown, when present. Note the that the group I upregulated in MZ*sox19bspg* is enriched for the regulators of transcription and for developmental genes. **c.** Mean zygotic transcription profiles for the groups which are upregulated in the mutants compared to the wild-type (I,J,K), relative to 2.5 hpf. Note the early upregulation of the transcripts over the WT in MZ*sox19bspg* (group I) and MZ*spg* (group J), but not in MZ*sox19b* (K), where the transcripts are mostly upregulated at 5 hpf (red arrow).

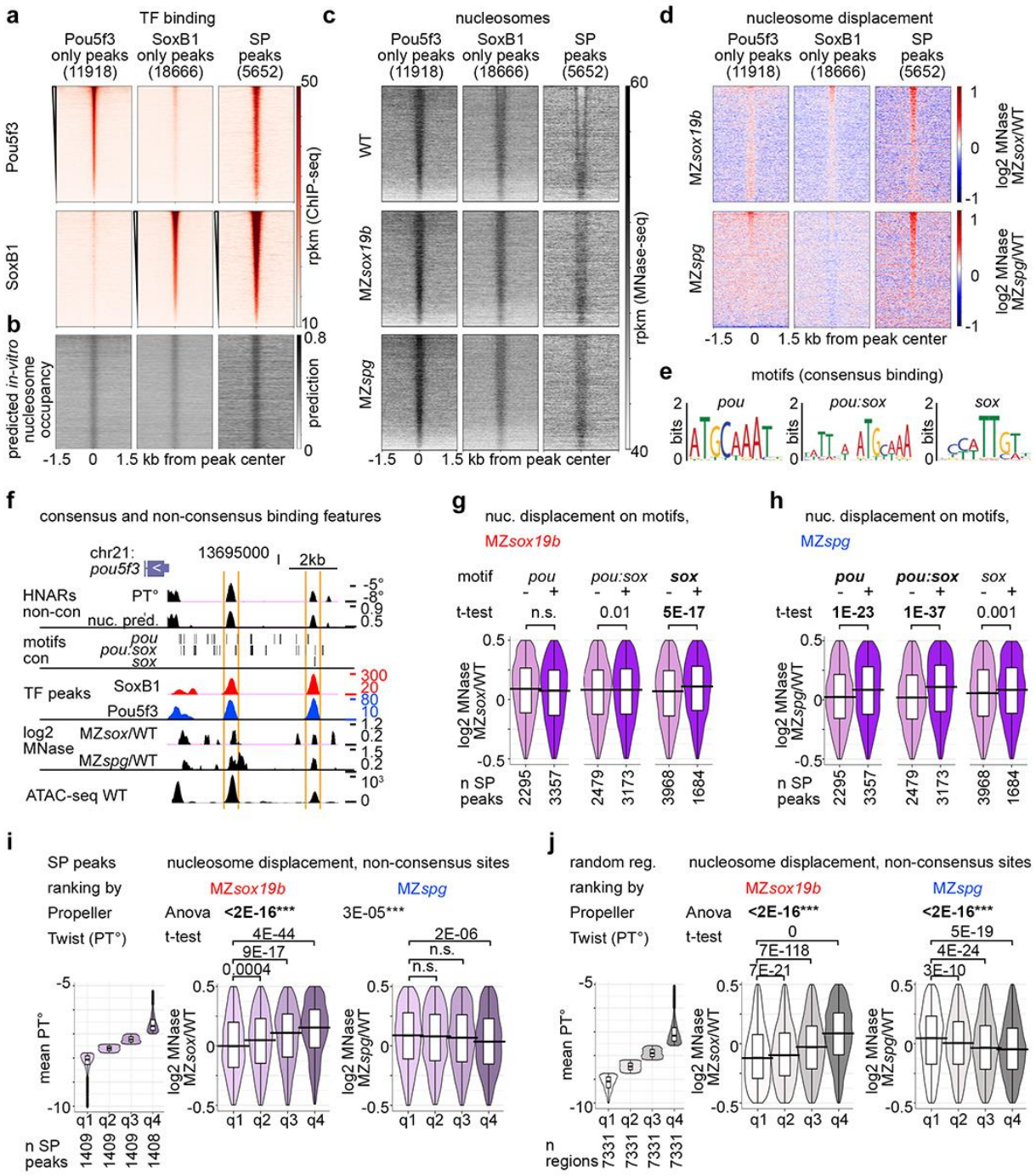

Figure 5. Pou5f3 and Sox19b displace nucleosomes using different binding cues. a-d. Heatmaps of 3 kb genomic regions around Pou5f3-only (P), SoxB1-only (P) or SoxB1-Pou5f3 binding peaks, sorted by descending TF occupancy, as shown in (a). **a.** TF occupancy. **b**. *in-vitro* predicted nucleosome occupancy. **c.** Real nucleosome occupancy (MNase) at 4.3 hpf in the indicated genotypes. **d**. Nucleosome displacement at 4.3 hpf in MZ*sox19b* and MZ*spg*. **e.** Consensus binding cues: *pou*, *pou:sox* and *sox* motifs. **f.** Genomic region upstream of *pou5f3* gene with non-consensus binding features (HNARs), consensus binding features (motifs), TF binding, nucleosome displacement in the mutants at 4.3 hpf (as log2 MNase mut/WT) and ATAC-seq signal in the WT at 4.3 hpf[75] are shown. Two enhancers upstream of *pou5f3* gene are highlighted by orange lines. **g,h.** Nucleosome displacement was compared on SP peaks with (+) or without (-) the motifs indicated above. p-values in

2-tailed Student t-test. **g.** MZ*sox19b:* nucleosome displacement depends mostly on *sox* motifs. **h.** MZ*spg*: nucleosome displacement depends mostly on *pou* and *pou:sox* motifs. **i,j**. Nucleosome displacement on SP peaks (i) and random control genomic regions (j) increases with predicted PT° value in MZ*sox19b*, but decreases in MZ*spg*. Left: SP peaks or random regions were sorted by increasing PT° and split into 4 quartiles q1 (low)- q4 (high). Right: Nucleosome displacement changes between quartiles were statistically tested using 1-way Anova (numbers above the graph). Differences between the first and other quartiles were tested using 2-tailed Student's t-test. Nucleosome displacement, motif density and PT° per 300 bp region (+/- 150 bp from the peak summit). n – number of peaks.

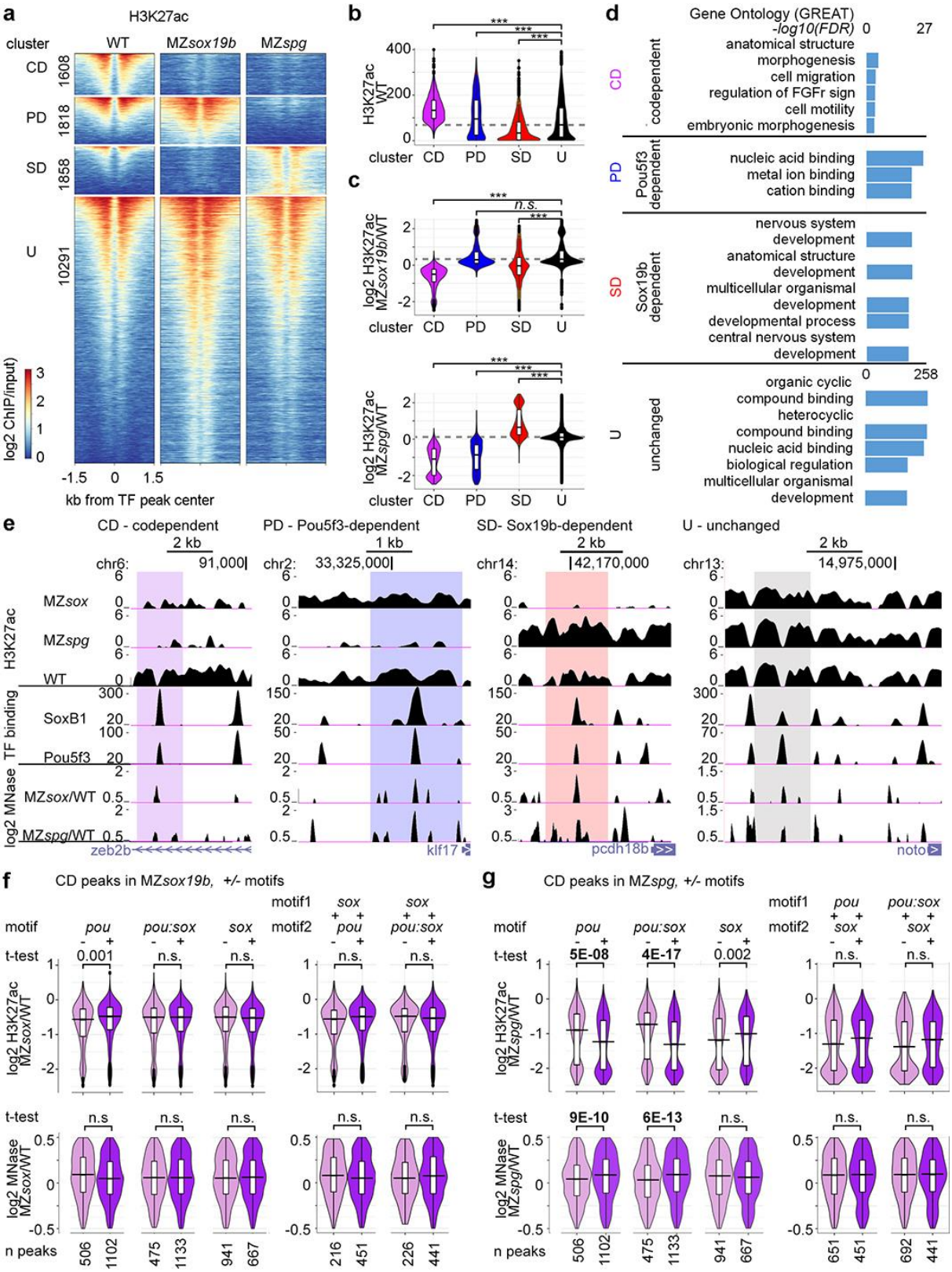

Figure 6: Three types of differentially H3K27-acetylated enhancers at 4.3 hpf. a. Pou5f3 and/or SoxB1- binding enhancers clustered by regulation in the mutants. CD-codependent, PD – Pou5f3 – dependent, SD – Sox19b dependent, U – unchanged. Number of enhancers in each cluster is indicated at the left, genomic regions within each cluster are sorted by descending H3K27ac. **b,c** Summary statistics for 4

enhancer clusters: Violine plots of H3K27ac in the wild-type (b) or H3K27ac change in each mutant (c). 1-way Anova <2e-16 for all graphs. *** - p-value<0.0005, n.s.- p-value >0.05 in Tukey-Kramer Test. **d.** GREAT[76] enrichment in Gene Ontology terms. Top five categories per cluster are shown, if present. Note that SD enhancers are associated with the neural system development. **e.** Enhancer example for each cluster (UCSC genome browser, H3K27ac signal units are log2 ChIP/Input for each genotype). **f,g.** H3K27ac change and nucleosome displacement were compared on CD enhancers with (+) or without (-) the motifs indicated above. H3K27ac change and nucleosome displacement on CD enhancers do not depend on motifs in MZ*sox19b* (f), but both values depend on *pou:sox* and *pou* motifs in MZ*spg* (g). p-values in 2-tailed Student's t-test, n.s. – non-significant.

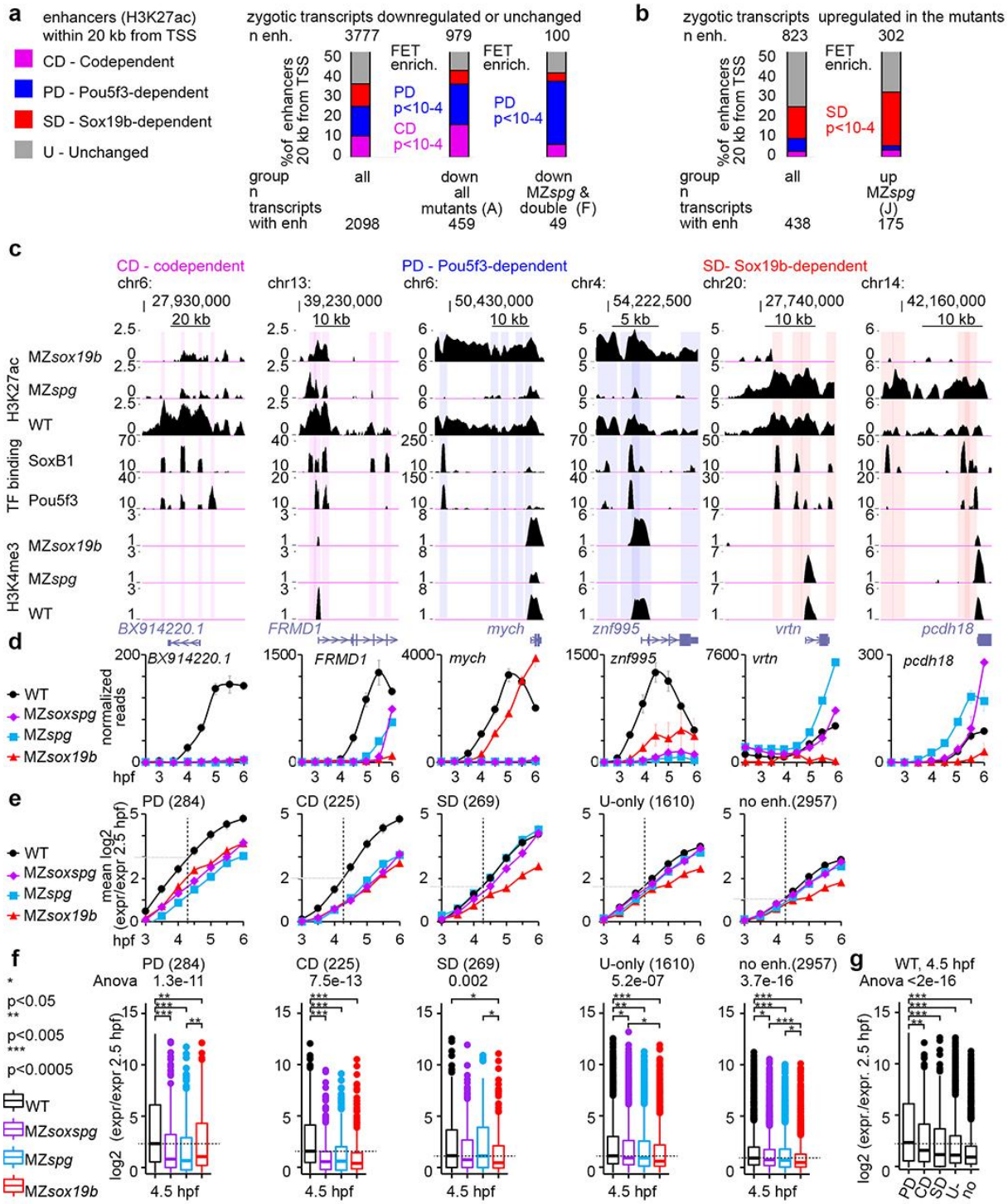

Figure 7: Pou5f3-dependent and codependent enhancers are associated with the major ZGA wave. a,b. Enhancers were linked to the closest gene within +/-20 kb. Zygotic transcript groups A-K were tested for enrichment in CD, PD and SD enhancers. FET - 2-tailed Fisher exact test. Number of enhancer-linked transcripts below the graph, number of enhancers above the graph. **a.** CD and PD enhancers were overrepresented around zygotic genes, downregulated in all mutants (group A). PD enhancers were also overrepresented around zygotic genes, downregulated in MZ*spg* and double mutants (group F at Fig.3b). **b**. SD enhancers were overrepresented around the genes, upregulated in MZ*spg* (group J, Fig.4a), relatively to the list of zygotic genes, upregulated in the mutants. **c.** Examples of genes linked

to three or more enhancers of the same type (CD- PD- and SD). 1 kb regions around enhancers are highlighted. UCSC genomic browser view, H3K27ac and H3K4me3 are shown as log2 ChIP/input ratio, TF binding - normalized reads. Note the changes in H3K4me3 promoter mark. **d.** Expression of genes shown in (c) in the WT, single and double mutants, as indicated, at 2.5 to 6 hpf. Error bars – SEM (standard error of the mean). Note that CD and PD- linked genes were downregulated or delayed in all mutants. SD- linked genes *vrtn* and *pcdh18b* were upregulated over the wild-type levels in MZ*spg*, downregulated in MZ*sox19b* and expressed at the intermediate level in MZ*sox19bspg*. **e.** Mean zygotic expression of transcripts, grouped by the presence and type of enhancer, relative to 2.5 hpf. PD, CD and SD transcripts are linked to at least one enhancer of respective type, U-only transcripts are linked to unchanged enhancers only, no enh group has no enhancers within +/- 20 kb. The dotted lines indicate 4.3 hpf time point. **f.** Summary statistics at 4.5 hpf time point in (e), in all genotypes. 1-way Anova values and all significant pairwise differences (Tukey-Kramer test) are shown. **g.** PD group is expressed earlier than the others. Summary statistics at 4.5 hpf time point in (e), in the wild-type.

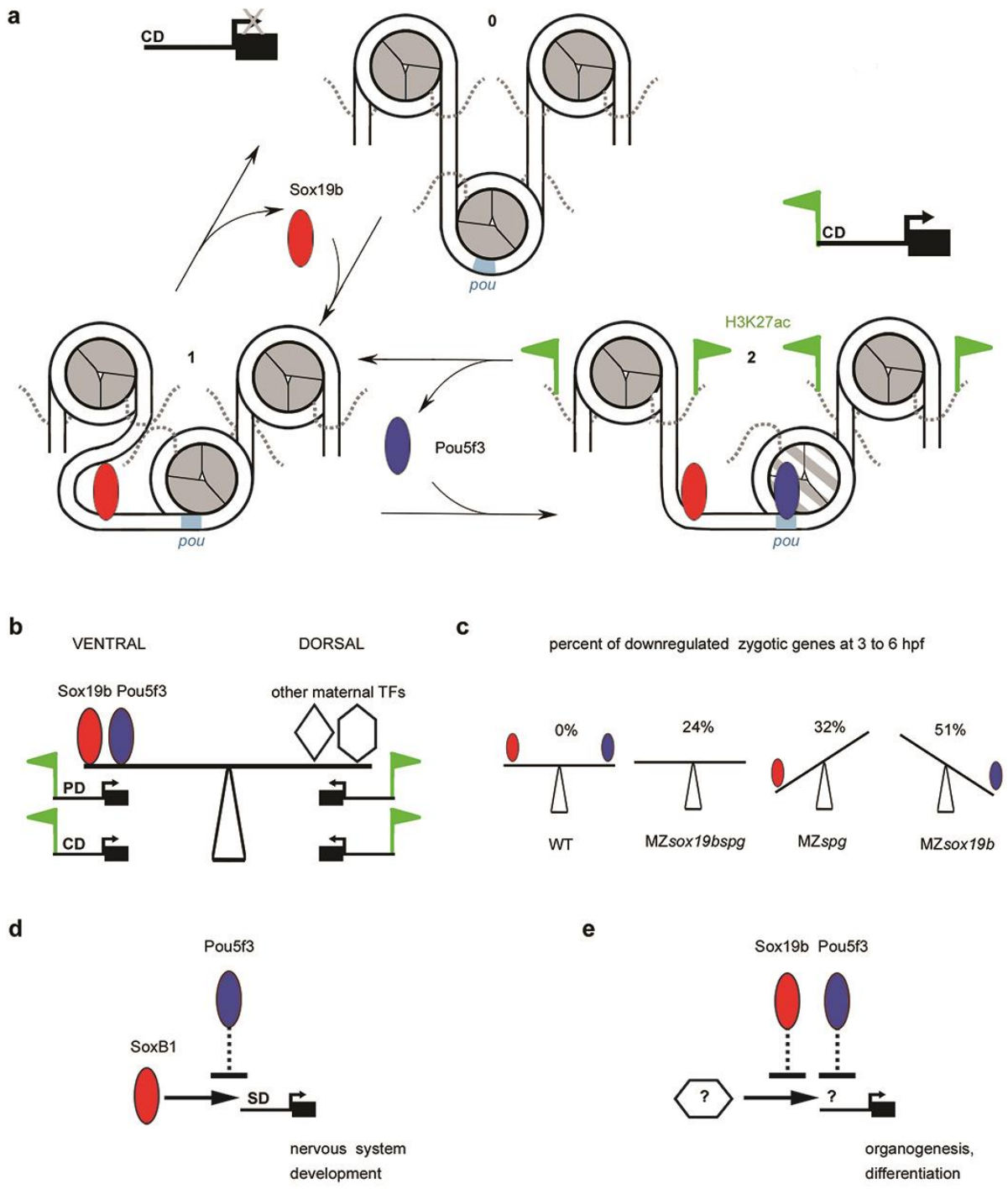

Figure 8: Summary of Pou5f3 and Sox19b activities. a. Model of Sox19b-assisted Pou5f3 loading on codependent enhancers. Sox19b binding competes with histones (0-1), which allows Pou5f3 binding (2) to *pou* (or *pou:sox*) motif nearby or at a distance. DNA-bound Pou5f3 promotes H3K27 acetylation of the neighboring nucleosomes and promotes gene expression. Nucleosome drawings are adapted from[77]. **b.** The dorso-ventral balance at ZGA: Pou5f3 and Sox19b prime ventral and ectodermal genes for activation, acting on PD and CD enhancers. Dorsal and mesendodermal genes are primed by other maternal TFs. **c.** Compensation of zygotic transcriptional delays in MZ*sox19bspg* mutants. **d.** Pou5f3 suppresses

premature activation of SD enhancers by SoxB1 activity**. e.** Pou5f3 and Sox19b suppress premature expression of transcription factors involved in organogenesis and differentiation.

Methods

Experimental model and subject details

Wild-type fish of AB/TL and mutant sox19b^{m1434} strains were raised, maintained and crossed under standard conditions as described by Westerfield[78]. The progeny of MZ*spg*^{m793} and MZ*sox19b*^{m1434}*;spg*^{m793} homozygous lines was rescued to viability by microinjection of Pou5f3 mRNA into 1-cell stage embryos as described previously[26]. Embryos were obtained by natural crossing (4 males and 4 females in 1,7 l breeding tanks, Techniplast) or by *in-vitro* fertilization, as indicated. Wild-type and mutant embryos from natural crosses were collected in parallel in 10-15 minute intervals and raised in egg water at 28.5°C until the desired stage. Staging was performed following the Kimmel staging series[30]. Stages of the mutant embryos were indirectly determined by observation of wild-type embryos born at the same time and incubated under identical conditions. All experiments were performed in accordance with German Animal Protection Law (TierSchG) and European Convention on the Protection of Vertebrate Animals Used for Experimental and Other Scientific Purposes (Strasburg, 1986). The generation of double mutants was approved by the Ethics Committee for Animal Research of the Koltzov Institute of Developmental Biology RAS, protocol 26 from 14.02.2019.

Genomic DNA isolation and PCR for genotyping

Genomic DNA was isolated from individual tail fin biopsies of 3 months old fish or from the pools of 24 hours post-fertilization (hpf) embryos. Tail fin biopsies or embryos were lysed in 50µl lysis buffer (10 mM Tris pH 8, 50 mM KCl, 0.3% Tween20, 0.3% NP-40, 1mM EDTA) and incubated at 98°C for 10 min. After cooling down Proteinase K solution (20 mg/ml, A3830, AppliChem) was added and incubated overnight at 55°C. The Proteinase K was destroyed by heating up to 98°C for 10 min. The tail fin biopsies material was diluted 20x with sterile water. 2µl of was used as a template for PCR. PCR was performed in 25-50 µl volume, using MyTag polymerase (Bioline GmbH, Germany) according to the manufacturer instructions, with 30-35 amplification cycles. The primer sequences are listed in the subsequent sections and in the Data Resources Table at the end of this chapter.

Generation and maintenance of MZ*sox19b*^{m1434} mutant line

To generate *sox19b* zebrafish mutants, we used the TALEN technique and targeted the first exon for the mutation. The square brackets indicate the spacer sequence, flanking sequences should bind to TAL1 and TAL2[29]: 5'-GATGGAGCACGAGCT[GAAGACCGCTGGTCCA]CCCCACACCCTCCAGC-3'. For restriction digest the enzyme BbsI was selected with the corresponding restriction site 5'-GAAGAC-3'. After injecting the TALENs (100 ng/µl each) into 1-cell stage wild-type embryos we tested the proper activity of TALENs. We extracted genomic DNA from 20 of 24 hpf old wild-type injected embryos and 20 non-injected embryos for the control, and used it as a template for PCR. We used the following primers for PCR: Sox19b-f1 5'-ATTTGGGGTGCTTTCTTCAGC-3' and Sox19b-r1 5'-GTTCTCCTGGGCCATCTTCC-3'. This gives a product of 362 bp length,which contains two restriction sites for BbsI. 5µl of the PCR mix was digested overnight with 5 units of BbsI (New England Biolabs) in 30 µl volume. The digestion of the wild-type PCR product resulted in three bands with sizes of 40 bp, 132 bp and 190 bp. In the successfully mutated embryosone of the BbsI sites was destroyed, and additional 322 bp band appeared. After successful TALEN injection we let the fish grow and found through an outcross with wild-type two founders out of 19 tested fish. We chose the founder with an 8 bp deletion resulting in a frameshift and a stop codon after 62 amino acids (5'-GATGGAGCACGAGCTGAAGA|CCACCCCACACCCTCCAGC-3', the line shows the position where the deletion occurred). We incrossed the heterozygous progeny of this founder (*sox19b*^{m1434}) and selected homozygous fish by PCR-BbsI restriction digest (322 bp band, but no 190 bp and 132 bp bands should be present). To obtain MZ*sox19b* ^{m1434} line, we incrossed the homozygous fish. MZ*sox19b* ^{m1434} line has been maintained by incrossing over 6 generations. To confirm that all the fish carry *sox19b*^{m1434}, allele, we genotyped tail fin biopsies of 3 months old fish in each generation. In all generations the fish were viable, fertile and phenotypically normal, except that the embryos were somewhat smaller than the wild-type and exhibited the gastrulation delay. To obtain Maternal mutant progeny (M*sox19b),* MZ*sox19b* females were crossed to the wild-type (AB/TL) males.

Generation of MZ*sox19bspg* double mutants

We obtained MZ*sox19bspg* mutants in three subsequent crossings. First, MZ*spg* homozygous males were outcrossed with MZ*sox19b* homozygous females. The double heterozygous progeny developed into phenotypically normal and fertile adults. The heterozygous fish were incrossed to obtain the double mutants. To bypass the early requirement for Pou5f3 in the *spg793* homozygous mutants, one-

cell stage embryos were microinjected with synthetic Pou5f3 mRNA. The fish were raised to sexual maturity (3 months), genomic DNA from tail fin biopsies was isolated and used it for genotyping. We first selected *sox19b* homozygous mutants, by PCR-with Sox19b-f1/Sox19br1 primers followed by restriction digest with BbsI, as described in the previous chapter. To select the double homozygous fish, we used the genomic DNA from *sox19b* homozygous mutants to PCR-amplify the region, flanking the *spg*^{m793} allele. *Spg*^{m793} allele carries an A->G point mutation in the splice acceptor site of the first intron of Pou5f3 gene, which results in the frameshift starting at the beginning of the second exon, prior to the DNA-binding domain. *Spg*^{m793} is considered to be null allele[79]. We used the following PCR primers: spg-f1 5'-` GTCGTCTGACTGAACATTTTGC -3' and spg-r1 5'-` GCAGTGATTCTGAGGAAGAGGT -3'. Sanger sequencing of the PCR products was performed using commercial service (Sigma). The sequencing traces were examined and the fish carrying A to G mutation were selected. To obtain the Maternal-Zygotic (MZ) homozygous mutants, *Zsox19bspg* were in-crossed. MZ*sox19bspg* fish have been maintained by in-crossing over three generations. Each generation of MZ*sox19bspg* was rescued to viability by microinjection of Pou5f3 mRNA into 1-cell stage embryos. Fish in each generation were genotyped to confirm that they carry *sox19b* homozygous allele. To obtain Maternal (M) mutants, MZ*sox19bspg* females were crossed to the AB/TL males.

Preparation of syntetic *sox19b* mRNA

Sox19b open reading frame was amplified from zebrafish total cDNA (4.3 hpf), using the PCR primers according to *sox19b* mRNA sequence in UCSC: sox19Bf1 with BamHI: 5'-GGGGGATCCATGGAGCACGAGCTGAAGAC-3', Sox19bR1 with Xho1: 5'-GGGGCTCGAGTCAGATGTGAGTGAGGGGAAC-3'. The PCR product was cloned in pCRII-TOPO vector using TOPO TA Cloning® Kit (Invitrogen), and further sub-cloned into PCS2+ vector via BamH1/Xho sites. mRNA was *in-vitro* transcribed with mMESSAGE mMACHINE® SP6 Kit (Ambion) according to the user manual. *Sox19b* mRNA was cleaned up with QIAGEN RNeasy® Mini Kit.

SoxB1 Morpholino knockdown

Two translation-blocking morpholinos per each SoxB1 gene (Sox2, Sox3, Sox19a and Sox19b (see the Data Resources Table for the sequences) were as designed and validated by Okuda et al, 2010[27] and provided by Gene Tools. LLC (Philomath, USA). To reproduce SoxB1 quadruple knockdown phenotype (QKD), 1-cell stage wild-type embryos were microinjected with the mix of 8 morpholinos, (0.9 ng each morpholino, 7.2 ng per embryo total, as in[27]). Standard Morpholino control (Gene Tools) injection at 7.2 ng per embryo was used as a control. To generate the triple knockdown of zygotic SoxB1 genes in MZ*sox19b* mutants, 1-cell stage MZ*sox19b* embryos were microinjected with Sox2, Sox3 and Sox19a morpholino mix. Microinjection of 0.9 ng per morpholino (5.4 ng in total) resulted in the MZ*sox19b* TKD phenotypes similar to WT QKD in two experiments: the tail bud and head were not properly formed and anterior-posterior axis was severely truncated (Fig.2b).

To investigate the dose-dependent effects of the morpholinos, we injected different amounts of TKD mix and TKD control mix (TKDco) into MZ*sox19b* embryos. In TKDco mix, Sox2 and Sox3 morpholino pairs were present as in TKD, while Sox19a was replaced with Sox19b morpholino pair, so that only three out of four SoxB1 genes would be blocked upon injection. As it was previously shown that triple Sox2/3/19b knockdown causes mild defects detectable only after 31 hpf[27], the main purpose of using TKDco mix was to better control for non-specific effects of TKD mix. Non-specific developmental defects, especially cell death in the nervous system, are common caveats of the morpholino use in zebrafish, which depend on the morpholino concentration and composition[80]. TKD injection in of 0.9 ng each morpholino (5.4 ng total) per MZ*sox19b* embryo, or 50% of this amount (0.45 ng each) resulted in the range of three phenotypic classes (II-IV, Fig.S3d,e) in two experiments. The "middle" phenotype (class III) was similar to previously published QKD phenotype resulting from either 0.45 and 0.9 ng morpholino injections into the WT[27]. The TKDco mix injection into MZ*sox19b* embryos resulted in the embryos, which were often somewhat shorter than non-injected (class I and normal in Fig. S3d,e); non-specific axial defects were observed in less than 5% of MZ*sox19b*-TKDco injections (0.9 ng, Fig. S3f). Since TKD injection into MZ*sox19b* caused different and much more severe phenotypes than TKDco, we concluded that the MZ*sox19b*-TKD phenotype specifically results from the reduction of zygotic SoxB1 activity.

MZ*sox19b*-TKD rescue experiments

For the rescue experiments, 5 or 20 pg of *sox19b* mRNA or 20 pg GFP mRNA was co-injected into 1-cell stage MZ*sox19b* embryos together with 2.7 ng TKD Morpholinos. For the control, MZ*sox19b* embryos were injected with 2.7 ng TKDco mix and 20 pg control GFP mRNA. The embryos of all

experimental groups were let to develop until 19 hpf at 28,5°C. At 19 hpf, the phenotypes were scored, representative pictures of living embryos were taken, and the whole experiment was fixed in 4% Parafolmaldehyde (PFA) in PBS. The rescue extent was quantified by measuring the body axis length according to Okuda et al, 2010. Namely, we took lateral images of all fixed embryos, draw the midline from head to tail for each embryo (Fig. S3g) and measured it using ImageJ. The representative phenotypes of one of two experiments (rep1) are shown in Fig.2c, body length statistics for two independent experiments is shown in the Fig. S3g.

Noggin and Chordin Morpholino knockdown

Noggin and Chordin Morpholinos were designed by Dal-Pra et al (2006)[35]. To block Noggin translation, 1 nM Noggin MO was injected. To moderately reduce Chordin translation, 100 pM Chordin Morpholinos were injected, as recommended by Dal-Pra et al. (2006). All morpholinos were provided by Gene Tools. LLC (Philomath, USA), sequences are provided in the Data Resources Table.

In vitro fertilization

Adult male fish were anesthetized with Tricain (4% 3-Aminobenzoic acid ethyl ester, PH = 6.7) and then positioned with anal area above. The sperm was taken using a capillary, mixed with 5 µl E400 buffer (9.7 g KCl, 2.92 g NaCl, 0.29 g $CaCl_2$ -2H_2O, 0.25 g $MgSO_4$ -7H_2O, 1.8 g D-(+)-Glucose, 7.15 g HEPES in 1L dH_2O, PH = 7.9) and then with 150µl SS300 buffer (0.37 g KCl, 8.2 g NaCl, 0.15 g $CaCl_2$ -2H_2O, 0.25 g $MgSO_4$ -7H_2O, 1.8 g D-(+)-Glucose 20ml 1M Tris-Cl, PH=8.0, in 1L dH_2O). The collected sperm was kept at room temperature. Adult females were anesthetized, placed into 35mm petri dish and to squeeze the eggs. For fertilization, we activated the sperm by adding 200µl sterile water and transferred it onto the freshly collected eggs. After 2 minutes, the dish was filled with sterile water and transferred to 28,5°C incubator.

Quantification of the pre-MBT cell cycle duration in the wild-type and MZ*sox19b* mutant embryos.

To obtain synchronously developing wild-type and MZ*sox19b* embryos, *in-vitro* fertilization of approx. 200-400 eggs per genotype was performed by two people in parallel. The fertilized embryos were scored and left to develop at 28.5℃. 30~35 embryos from each genotype were simultaneously fixed in 4% PFA every 15 minutes starting from 1.75 hpf (5th cell cycle, 32 cell stage) till 3hpf (10th cell cycle, 1K stage). The embryos were incubated at 4℃ overnight, washed 3x in PBST (PBS with 0,1% Tween-20) and manually dechorionated. The embryos were then permeabilized in PBSDT (PBS with 0,1% Tween-20, 1% DMSO and 0.3% TritonX-100) for 48 hours at 4℃, stained in 0.025µM SYTOX Green (ThermoFisher SCIENTIFIC) for 1.5 hours and washed 3x in PBST. The pictures were taken using LEICA MZ16F stereomicroscope equipped with epifluorescence. The pictures were contrasted in ImageJ 1.53c., and used to estimate the stage and the cell cycle phase (Fig.S1a). The embryos where the most nuclei were mitotic (metaphase to anaphase) were scored as intermediate between the cycles (i.e. cycle 5-6, 6-7 etc.,Fig.S1a). We did not detect the differences in the pre-MBT cell cycle lengths in two experiments. The quantification of representative experiment out of two is shown in the Fig.S1b.

Live imaging

Zebrafish embryos were imaged using Leica MZ APO stereomicroscope with Axiocam 305 camera and AxioVision or Zen2 software (Carl Zeiss).

Parallel time-lapse recording of developmental rates in the wild-type and mutants

The wild -type and MZ or M mutant embryos were obtained by natural crossings in the mass crossing cages (4 males and 4 females of each genotype). The freshly laid eggs were collected in 10-15 minute intervals. At 4-8 cell stage, one embryo per genotype was manually dechorionated and placed on a 1.5% agarose chamber filled with 0.3x Danieau's buffer. Images were taken with 3 min intervals for 24 hours by either Leica MZ APO stereomicroscope, Axiocam 305 camera and AxioVision software (Carl Zeiss). The video recording started between 1.5 -2 hpf. Time after fertilization (hh.mm) was added in ImageJ 1.50i. Time lapse recordings from five independent experiments are included in the Movie S1, (Part1-Part5), which show parallel development of the embryos with following genotypes:
Part1: WT, MZ*sox19b*, M*sox19b*
Part2: WT, MZ*sox19b*, MZ*spg*, MZ*sox19bspg*
Part3: WT, MZ*spg*, MZ*sox19bspg*
Part4: WT,M*spg*, M*sox19bspg*

Part5: WT, M*sox19b*, M*sox19bspg*

Whole mount *in situ* hybridization (WISH)

To visualize the expression pattern of some chosen genes we performed whole mount *in situ* hybridization as previously described[68]. The plasmids for anti-sense RNA probe synthesis were kind gifts of Wolfgang Driever, Matthias Hammerschmidt, Yusuke Kamachi and Liliana Solnica-Krezel. Embryos were fixed at proper stages with 4% paraformaldehyde at 4°C overnight, 3 times washed in PBST for 5 min, dechorionated manually in PBST and dehydrated with ascending series of methanol to 100%. Embryos in 100% methanol were stored at -20°C. After rehydrating with descending series of methanol to PBST, embryos were washed 3 times in PBST, pre-hybridized in 300µl Hyb-Mix for 3 hours at 65°C. 2 µl of prepared probes were added to 300µl Hyb-Mix for hybridization overnight at 65°C. Emrbyos were placed into 24-well plates and washed with in situ robot BioLane™ HTI by series washing steps: firstly embryos were washed three times for 20min in 300 µl 50% Formamide at 65°C, in 500 µl 2x SSCT for 15 times at 65°C, three times in 500 µl 0.2x SSCT for 20 min at 65°C, twice in 1 ml PBST for 10 min at room temperature. Then embryos were incubated in blocking solution (2% goat serum (heat inactivated) in PBST/BSA (2 mg/ml) for 2 hours at RT and incubated overnight with anti-DIG (1:5000 diluted in PBST) at 4°C. After washing 6 times for 20 min in 1ml PBST and once for 10min in 100Mm Tris-HCl, (pH=9.5), embryos were incubated with staining buffer for 15 min and the robot program is finished. We replaced staining buffer with 500 µl staining solution and stained for a proper time on a shaker. To stop the staining process embryos were washed with stop solution. Stained embryos were fixed in an increasing series of glycerol (25%, 50% and 75% in PBST), finally stored in 100% glycerol at 4°C. These embryos were imaged with Leica MZ APO stereomicroscope using AxioVision SE64 Rel. 4.9.1 software and the images were processed in Adobe Photoshop CS4.

MNase-seq, data processing and visualization.

MNase-seq and data processing was performed as previously described[16]. Briefly, 200-400 MZ*sox19b* embryos from mass-crossing cages were collected within 10-15 minutes. The embryos were enzymatically dechorionated and fixed 10 minutes in 1% Formaldehyde at dome (4.3 hours post-fertilization) stage. The nuclei were isolated and digested with MNase. The yield of and degree of digestion was controlled using Agilent High Sensitivity DNA Kit on Agilent Bioanalyzer, according to manufacturer instructions. Chromatin was digested so that it contained 80% mono-nucleosomes. Libraries were prepared using the Illumina sequencing library preparation protocol and single-end sequenced on an Illumina HiSeq 2 500 by Eurofins Company (Germany). All further data processing was done using European Galaxy server usegalaxy.eu[81]. Sequenced reads were mapped to the zebrafish danrer11 assembly using Bowtie2[82]. Resulting BAM files were converted to BED format, all mapped reads were extended to 147 bp in their 3' direction, truncated to the middle 61 bp and converted back to BAM format using BED Tools[83]. Alternate loci and unassembled scaffolds were excluded from the further analysis. To create MNase Bigwig files for the visualization of nucleosome density, BAM files for MZ*sox19b* (this work), MZ*spg* and WT (remapped from[20]) were converted to bigwig format using BAM coverage program in deepTools[84], bin size=10bp, normalization- rkpm(reads per million reads per one kilobase). To create log2 MNase Bigwig files for the visualization of nucleosome displacement in MZ*sox19b* and MZ*spg*, the log2 mutant MNase/wild-type MNase ratio was obtained using BigWigcompare program in deepTools, bin size=10. The heatmaps or profiles of selected genomic regions were plotted using plotheatmap or plotprofile programs in deepTools. Visualization of selected genomic regions was done in UCSC browser. For the violine plots and nucleosome displacement statistics on the TF-binding or control peaks, average MNase signals per 300bp were calculated using Bigwig files for each region, and log2 (mutant/WT) ratio was taken (data available as Source data file).

Chromatin immunoprecipitation (ChIP) for histone marks.

The embryos were obtained from natural crossings in mass-crossing cages (4 males + 4 females). 5-10 cages were set up per genotype, the eggs from different cages were pooled. The freshly laid eggs of MZ*sox19b*, MZ*spg* mutants and wild-type were collected in 10-15 min intervals. Unfertilized eggs were removed at 2-4 cell stage. Collected embryos were transferred to 0.5x Danieau's solution (for 1L of 30X stock: 1740 mM NaCl, 21 mM KCl, 12 mM MgSO4, 18 mM Ca(NO3)2, 150 mM HEPES buffer, ph 7.6; dilute 60X before use) followed by enzymatic dechorionation with pronase E (0.3 mg/ml). The reaction was stopped by adding 1 % BSA (final conc. 0.04 %) followed by two to three washing steps with 0.5x x Danieau's. The eggs were cultured in glass petri dishes to prevent the embryos from adhere to the dish and thus eventually rip. They were incubated at 28 °C until the 4.3 hpf stage was reached. In order to fix the chromatin state at developmental stage 4.3 hpf (dome) and avoid nucleosome shifts, the dechorionated embryos were homogenized in 10 ml 0.5 % Danieau's

containing 1x protease inhibitor coctail (PIC) using a Dounce tissue grinder and immediately treated with 1 % (v/v) Methanol-free Formaldehyde (Pierce) for exactly 10 min at room temperature. The homogenizate was transferred into a 15 ml falcon tube and shaken on a rotating platform for the rest of the 10 min. The fixation was stopped with 0.125 M Glycine by shaking for 5 min on a rotating platform. Subsequently the homogenizate was centrifuged for 5 min, 4700 rpm at 4 °C, whereupon a white pellet formed. The supernatant was discarded, and the pellet was resolved in Wardle cell lysis buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 0.5 % NP-40, 1-4 ml/1000 embryos). The lysate was distributed upon 2 ml eppendorf tubes, followed by 5 min incubation on ice with subsequent 1 min centrifugation, 2700 g at 4 °C. The supernatant was discarded again and the pellet was washed 2 times with 1 ml ice cold 1x PBST (for 1 L: 40 ml PO4 buffer (0.5 M), 8 g NaCl, 0.2 g KCl, 0.1% Twen20, pH 7.5). In order to count the obtained nuclei, the pellet was resolved in 1 ml ice cold 1x PBST, of which 10 µl were diluted 1:1 with 12 µM Sytox® green. The nuclei were scored under fluorescence microscope using the Neubauer counting chamber.
The residual nuclei were again pelleted by 1 min centrifugation at 2700 g and 4 °C, subsequently snap frozen in liquid nitrogen and stored at -80 °C. The nuclei collected in different days were pooled together to reach the total number of 2.5-3 million, which was used to start one ChIP experiment.

The chromatin was thawn and resolved in 2 ml of Wardle nuclei lysis buffer (50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1 % SDS) and incubated 1 h on ice. In order to shear the chromatin to 200 bp fragments (on average), the chromatin was sonicated using the Covaris S2 sonicator (DC 20 %, Intensity 5, Cycles of Burst 200, Time = 3*40 cycles with 30 sec each (3*20 min)). To ensure that the sonication was successful, 30 µl of the sheared chromatin was de-crosslinked with 250 mM NaCl over night at 65 °C and then analyzed using the Agilent Expert 2100 Bioanalyzer® and Agilent high sensitivity DNA Chip kit.

The lysed and sheared samples were centrifuged for 10 min at 14,000 rpm and 4 °C. 60 µl of each sample were kept as input control. The chromatin was then concentrated to 100 µl using the Microcon centrifugal filters (Merck Millipore MRCF0R030) and diluted 1:3 by adding ChIP dilution buffer (16.7 mM Tris-HCl pH 7.5, 167.0 mM NaCl, 1.2 mM EDTA) containing protease inhibitors. The antibody (listed in the Data resources table at the end of the section) were added and incubated overnight at 4 °C on a rotating wheel. 150 µl of magnetic Dynabeads coupled to protein G (Stock 30 mg/ml; invitrogen DynaI 10003D) were transferred into a 2 ml eppendorf tube and placed on a magnetic rack in order to remove the liquid from the beads. Subsequently the beads were washed 3x with 5 mg/ml specially purified BSA in PBST and 1x with 500 µl ChIP dilution buffer. After removing the ChIP dilution buffer, the chromatin-antibody mix was added and incubated with the beads at 4 °C overnight on a rotating wheel. Beads were pulled down by placing the eppendorf tubes on the magnetic rack in order to discard the supernatant. The beads were resuspended in 333 µl RIPA buffer containing PIC. The Protein G-antibody-chromatin complex was washed 4x5 min on a rotating platform with 1 ml of RIPA buffer (10 mM Tris-HCl (pH 7.6), 1 mM EDTA, 1 % NP-40, 0.7 % sodium deoxycholate, 0.5M LiCL), followed by 1x1 ml TBST buffer (25 mM Tris-HCl,150 mM NaCl, 0.05% Tween 20, pH 7.6). The beads were pulled down again and the supernatant was removed. In order to elute the chromatin, 260 µl elution buffer (0.1 M NaHCO3, 1% SDS) was added and incubated for 1 h at 65 °C in a water bath. The samples were vortexed every 10 - 15 min. Afterwards the supernatant was transferred to a fresh eppendorf tube by pulling the beads down, using the magnetic rack. 12.5 µl 5M NaCl was added to de-crosslink the chromatin and incubated overnight at 65 °C in a water bath. The input samples were treated as control in parallel (230 µl elution buffer per 30 µl input).

Purification of the de-crosslinked chromatin was performed using the QIAquick PCR Purification Kit from Qiagen. The concentration was determined using the Qubit fluorometer and Quanti-iT™ PicroGreen® dsDNA Kit according to manufacturer instructions.

ChIP quality control and library preparation for histone mark ChIP-seq

To estimate the signal to background ratio in each ChIP experiment, we have chosen the positive and negative reference genomic regions, enriched in or devoid of chromatin marks. According to previously published data[85], the chromatin region near *tiparp* gene was highly enriched in H3K27ac and H3K4me3 histone marks at 4.3 hpf, while genomic region near *igsf2* gene was not enriched in any of these marks. We performed quantitative PCR in ChIP and Input control material, using the primers for these regions. PCR primers used were: tiparp_f_1 5' CGCTCCCAACTCCATGTATC-3', tiparp_r_1 5'-AACGCAAGCCAAACGATCTC-3',igsf2_f_2 5'-GAACTGCATTAGAGACCCAC-3', igsf2_r_2 5'-CAATCAACTGGGAAAGCATGA-3'.QPCR was carried out using the SsoAdvanced™ Universal SYBR® Green Supermix from BIO-RAD. ChIP and input were normalized by ddCT method, using

negative reference region *(igsf2*). The ChIP experiment was considered successful, if the enrichment in ChIP over input control on the positive reference region (*tiparp*) was more than 5-fold.

In order to convert a small amount of DNA into indexed libraries for Next Generation Sequencing (NGS) on the Illumina platform, we used the NEBNext® Ultra™ DNA Library Prep Kit. As the DNA outcome of individual WT K27ac Chip and MZ*spg* K27ac ChIP experiments did not reach the input DNA limit for this kit (5 ng), we pooled together the material from two successful ChIP experiments, as well as corresponding inputs, for the library preparation in these genotypes. Two libraries were prepared from two single MZ*sox19b* K27ac ChIP experiments. Single libraries were prepared from single K4me3 experiments in three genotypes. The library preparation was carried out according to manufacturer instructions, with the modifications indicated below. The library preparation follows a 4-step protocol including end-repair (5' phosphorylation, dA-tailing), adaptor ligation, PCR enrichment including barcoding and clean up. Since the DNA input was <100 ng, in the adaptor ligation step, the NEBNext Adaptor for Illumina® (15 µM) was diluted 10-fold in 10 mM Tris-HCl (pH 7.8) to a final concentration of 1.5 µM and used immediately. At the final clean-up step, the reaction was purified by mixing the samples with AMPure XP Beads (45 µl) before incubating at room temperature for 5 min. After a quick spin using the tabletop centrifuge, samples were placed on a magnetic rack and supernatant was discarded. 200 µl of 80 % Ethanol were added and removed after 30 sec, two times. The beads were air dried for 3 min. and the DNA target was subsequently eluted by adding 33 µl of 0.1x TE (pH 8) and incubation at room temperature for 2 min. 28 µl of the library were transferred to a fresh PCR tube and stored at -20 °C. 2 µl of the sample were diluted 5-fold with 0.1x TE and used to check the size distribution of the library using Agilent Expert 2100 Bioanalyzer® and Agilent high sensitivity DNA Chip kit. In order to reduce the peak of residual unligated adaptors, the reaction was re-purified, by adding H2O up to 50 µl and 45 µl of AMPure XP Beads. The concentration was determined using the Qubit™ Fluorometer and Quanti-iT™ PicroGreen® dsDNA Kit. The 7 ChIP-seq libraries were sequenced at 70 mln paired end 150bp reads each: WT K27ac Chip, MZ*spg* K27ac Chip, MZ*sox19b* K27ac Chip1, MZ*sox19b* K27ac Chip2, WT K4*me3* Chip, MZ*spg* K4me3 Chip, MZ*sox19b* K4me3 Chip. The 7 corresponding input libraries were sequenced to 30 mln reads. Sequencing was performed by the Novogene company (China).

SoxB1 and Pou5f3 TF-binding and control peaks

SoxB1 and Pou5f3 binding peaks were derived from the published ChIP-seq data[10]. The FASTQ files were mapped to danrer11/GRCz11 assembly. Alternate loci and unassembled scaffolds were excluded. Peak calling was performed using MACS algorithm[86] with default parameters. The peaks overlapping Major Sattelite Repeats (MSRs) were removed. The peaks were extended to 300 bp (+/- 150 bp from the peak summit). The peaks were split into to three groups: SoxB1 peaks, overlapping with Pou5f3 peaks for at least 1 bp (SP), SoxB1-only peaks (S) and Pou5f3-only peaks (P). Random control peaks file was generated as follows: we calculated the distance from Pou5f3 and SoxB1 peaks to the nearest ENSEMBL transcript start, and took the control peaks at the same distance from the random ENSEMBL transcripts. The MSRs and all overlaps with Pou5f3 and SoxB1 peaks were removed. The peak coordinates in BED format are provided in the supplementary Table S3.

H3K27ac and H3K4me3 ChIP-seq data analysis and visualization.

H3K27ac and H3K4me3 ChIP-seq data processing was done using european Galaxy server usegalaxy.eu. Sequenced reads were mapped to the zebrafish danrer11 assembly using Bowtie2. Alternate loci and unassembled scaffolds were excluded. To create H3K4me3 Bigwig files for data visualization, the log2 ratio between each ChIP and merged inputs (in rpkm) was obtained using BAM compare program in deepTools, bin size =10. To create H3K27ac Bigwig files, the log2 ratio between each ChIP and merged inputs (in rpkm) was obtained using BAM compare program in deepTools, bin size =10. We first used two biological replicates of H3K27ac MZ*sox19b* to address, if there is a difference in H3K27ac mark between the genotypes – in this case, the difference between the genotype should be higher than the difference between the biological replicates within one genotype. We used deepTools2 mutliBigwigcompare tool to divide the whole zebrafish genome into 1680169 1 kb bins, and to compute a similarity matrix between four Bigwig files: H3K27ac MZ*sox19b*-rep-1, H3K27acMZ*sox19b*-rep-2, H3K27ac MZ*spg* and H3K27ac WT. Pairwise Pearson correlations and Principle Component Analysis (PCA) were performed using plotCorrelation and plotPCA tools. In both analyses the difference between the biological replicates was negligible when compared to the difference between the genotypes (Fig. S8a,b). Close correlation between the H3K27acMZ*sox19b* biological replicates demonstrated the reproducibility of our ChIP-seq protocol and validated the use of single replicates for our purposes. For the subsequent analysis we used H3K27acMZsox19b rep-1. The heatmaps or profiles of H3K27ac levels on selected genomic were plotted using plotheatmap or

plotprofile programs in deepTools. Histone modification profiles in single genes were visualized using UCSC browser.

Selection and analysis of putative enhancers

We considered TF- binding peak as putative enhancer, if H3K27- acetylation of the 1 kb region around the peak exceeded arbitrary threshold in at least one genotype. The Pou5f3- only, SoxB1- only, and double binding peaks (P,S and SP) were considered. Technically, we used the following procedure. TF- binding peaks BED files were extended to 1 kb (+/- 500 bp from the peak summits). H3K27ac MZ*sox19b*-rep-1, H3K27ac MZ*spg* and H3K27ac WT Bigwig files were converted to bedgraph format, all the regions where log2 ChIP/Input was less than 0.5 (arbitrary value) were filtered out. The resulting bedgraph files were joined with TF-peak BED files. The peak was considered as putative enhancer, If at least 500 bp around the peak were H3K27-acetylated over the arbitrary threshold (0.5 log2ChIP/Input). 70% of SP, 43% of S and 30% of P peaks were selected as enhancers using this criteria (Table S4). To assign four enhancer types, k-means algorithm (plotheatmap program in deepTools) was applied to the list of the putative enhancers, with cluster number 4. The CD, PD and SD groups were robustly recovered also when using the higher number of clusters: higher k- numbers resulted in splitting “unchanged” to several groups, by H3K27ac distribution or abundance. For statistical analysis and violine plots, H3K27acetylation per 1 kb was calculated for each enhancer in each genotype. H3K27 acetylation change in the mutant was calculated as log2 mut/WT ratio (data available as Source data file). Enhancers was linked to the closest transcription start sites of ENSEMBL genes within +/-20 kb, and to corresponding ENSEMBL transcripts.

Motif finding, scoring and visualization.

De-novo motif finding on SoxB1 and Pou5f3 peaks was performed using MEME suite[87] on Galaxy server. SP, P or S peaks were sorted by descending ChIP-seq signal, sequences +/- 30bp from peak summit were taken for motif finding. Top 1,000 sequences were used as an input for MEME with the following parameters Motif Site Distribution- Zero or one site per sequence, Maximum Number of Motifs-15, Minimum Motif Width-6, Maximum Motif Width-25. To sort out the consensus motifs for Pou5f3 and SoxB1, the de-novo found motifs were compared to JASPAR vertebrate non-redundant database[61] using Tomtom[88]. Three groups of motifs were kept for further analysis: *pou:sox*, *sox* and *pou* (Table S3). For *Pou:sox* motif group, the top matching profile was MA0142.1 (Pou5f1::Sox2), for *sox* motif group MA0143.3 (Sox2), and for *pou* motif group MA0792.1 (POU5F1B), MA0627.1 (Pou2f3), MA0789.1 (POU3F4), or MA0787.1 (POU3F2). To score the motif occurrences on TF binding and control regions, the ChIP-seq or control peaks were extended to 3 kb around the summit and genomic coordinates of the individual motif occurrences were obtained using FIMO[89] with p-value threshold 10^{-4}. The genomic coordinates of each motif were saved as a BED file. BED files for all motifs in the group were concatenated and redundant hits were removed. BED files were converted to Bigwig format with BEDtools, and used for visualization of the motif density on the selected genomic regions using deepTools2. The numbers of non-redundant matches for *pou:sox, sox* and *pou* motifs per 300 bp peak was scored in Galaxy, listed in TableS3 and provided as source data files.

Gene Ontology analysis.

Enhancer clusters and transcript groups were tested for enrichment in GO: Biological Proccess and GO: Molecular Function. For the enhancer groups we used GREAT analysis[76] at http://great.stanford.edu/great/public-3.0.0/html/. Genomic regions were converted to zv9/danrer9 genomic assembly, which is the only assembly available for this server, using Liftover Utility from UCSC and associated with genes using 20 kb single nearest gene association rule. The categories were ranked by ascending FDR value, 5 top categories were listed. For the transcript groups A-K we used DAVID[90]. The enriched GO: categories were ranked by ascending p-value (cutoff 0.05). 4 top categories per group were listed.

RNA-seq time curves: material collection, processing and sequencing.

The freshly laid eggs were obtained from natural crossings in mass-crossing cages (4 males + 4 females). 5-10 cages were set up per genotype, the eggs from different cages were pooled. At least 600 freshly laid eggs per each genotype collected within 10-15 minutes were taken for single experiment. To match the developmental curves as precisely as possible, the material was simultaneously collected for two genotypes in parallel. In each of five experimental days, the material was collected for two genotypes, as specified below: MZ*sox19b*-rep1 and MZ*spg*-rep1, WT-rep1 and MZ*spg*-rep2, WT-rep2 and MZ*sox19b*-rep2, WT-rep3 and MZ*spg*-rep2, WT-rep1 and MZ*sox19spg*-rep1, WT-rep4 and MZ*sox19spg*-rep2. 45-60 minutes after the egg collection, we ensured that the

embryos of both genotypes are at 2-4 cell stage, removed non-fertilized eggs and distributed the embryos to 8 dishes per genotype, 40-45 embryos per dish, to obtain 8 time points per genotype. The temperature was kept at 28.5°C throughout the experiment. The embryos from one dish per genotype were snap-frozen in liquid nitrogen every 30 minutes, starting from 2.5 hpf (pre-ZGA, 256 cell stage, 8th cell cycle) till midgastrula (6 hpf). In total, we collected 4 biological replicates of the time curve for the wild-type, and two biological replicates for each of MZ*sox19b*, MZ*spg* and MZ*sox19bspg* mutants. Total RNA was isolated with QIAGEN RNeasy kit following the user's manual. We checked RNA quantity and quality by using Agilent RNA 6000 nano Kit on Agilent Bioanalyzer, according to manufacturer instructions. The 6 hpf time point from MZ*sox19bspg* -rep1 had to be excluded, due to non-sufficient RNA quality; the corresponding WT-rep1 time point was also removed. Poly-A enriched library preparation and single-end 50bp sequencing (35M reads per sample) on Illumina platform was performed by NOVOGENE company (China) for each sample. To estimate the non-biological variation between the samples, we prepared and sequenced two technical libraries for each time point in the following time curves WT-*rep1*, MZ*sox19b*-rep1 and MZ*spg*-rep1.

RNA-seq time curves: data processing and visualization.

FASTQ files were further processed on european Galaxy server usegalaxy.eu. All sequenced clean reads data were trimmed on 5' end for 5 bp by Trim Galore! program according to the sequencing quality control. Trimmed reads were mapped to danRer 11 using RNA STAR[91] and ENSEMBL gene models. Number of reads for each gene was counted in Feature count[92]. Feature counts were cross-normalized using DEseq2[93]. This processed RNA-seq data table is deposited in GEO with the number GSE137424. To estimate biological and techinical variance across the samples, we excluded non-expressed ENSEMBL gene models ("0" reads throughout all experiments), and calculated pairwise Pearson correlations between biological replicates (the same time point and the same genotype, material from different experimental days) and also between technical replicates (two libraries prepared from the same material). Pearson correlation coefficients for biological replicates were at the range 0.93- 0.98, which indicated very strong linear relationship and was close to the technical variance range (Pearson correlation coefficients 0.96-0-99 for technical replicate pairs). To account only for biological variance, we excluded technical replicates from the further analysis. The normalized expression data used for analysis (biological replicates only) is available as Source file. This data was used for gene expression visualization, statistics, and as an input for RNA-sense (see sub-chapter "RNA-sense program" below for explanations). The data visualization was done in R with custom R scripts (boxplots, violin plots, heatmaps) or in Excel (individual genes, averaged time curves). For the heatmaps and summary graphs, the biological replicates for each genotype were averaged. For the heatmap view of gene expression change, we calculated the maximal expression value across the time curve for each transcript (max) in all four genotypes. Expression/max ratio for each timepoint was plotted. To compare the zygotic expression changes between the transcript groups, we calculated the log2 ratio of the expression in each time point from 3 to 6 hpf to the expression at 2.5 hpf for each transcript. The averaged results are shown as box plots with 1-way Anova and Tukey-Kramer test statistics, or as mean time curves (error bars show Standard Error of the Mean). For the time curves of the individual transcripts, the line graphs were drawn in Excel in linear or logarithmic scale (error bars show SEM of biological replicates).

RNA-sense program.

In order to facilitate the biological interpretation of time-resolved RNA-seq data, we developed a 3-step procedure called RNA-sense (Movie S2). In principle, the usage of RNA-sense is not only restricted to RNA time series. RNA-sense can be applied to compare two groups of data series to capture the differences in the dynamic changes between the groups. The series data could be temporal, spatial, or any other continuous condition like series concentration of drug treatment. The data itself could be any sequencing data, including DNA, RNA and protein, or any other comparable large datasets.

In step one, time-resolved RNA-seq data in one of two conditions, e.g. the wild-type (user-defined parameter ExperimentStepDetection="WT"), are analyzed with respect to their temporal profile. The transcripts expressed below a user-defined threshold are excluded from the analysis. First, for each gene and for each measurement time point t, dynamic data is split into two groups before and after (after and equal to) time point t. The data is fitted by both a one-step model (two different means before and after time point t) and by a constant model (mean over all data points) and models are compared pairwise by means of likelihood ratio tests for each time point t. If the one-step model is significantly better (with user-defined p-value cutoff, pVal switch) than the constant model, a switch is detected for this time point. The difference of the means before and after the time point defines the

direction of the switch “up” or “down”. If switches were detected at different time points for each gene, the first possible time point is chosen.

In step two, fold changes between wild-type and mutant data are analyzed. For each gene and for each time point, Robinson and Smyth exact negative binomial test (with user-defined p-value cutoff, pVal FC) is performed to determine whether genes are significantly up- or downregulated in the mutant with respect to wild-type. The function exact.nb.test from the R package NBPSeq is used for analysis (https://cran.rstudio.com/web/).

In step three, results of step one and two are combined. Genes are grouped in a matrix form with respect to switch time (y-coordinate) and mutant fold change (x-coordinate). Genes for which fold change was detected at several time points appear several times in the matrix (Fig.S6A, B). For each tile of the matrix, Fisher’s exact test for non-randomness is performed to analyze the correlation between the two properties switch time and fold change detection. Tiles with a low p-value in the Fisher test show a high correlation between switch time and fold change. This can be interpreted as a high number of genes for which the switch point is shifted in time in the mutant condition.
RNA-sense is a flexible tool with several user-defined parameters:
-ExperimentStepDetection: tells which of two conditions (i.e. WT or mutant) should be used for switch detection in Step 1
-threshold – the transcript is included in the analysis, if the expression value in at least one data point reaches the threshold.
-pVal switch – p-value threshold for switch detection in Step 1
-FC – fold change value threshold (optional) for Step 2
-pVal FC – p-value threshold for fold change analysis at Step 2
In our time series data, single 6 hpf data point was missing in one of the two biological replicates for MZ*sox19bspg* and in parallel WT control sample. To get an input for RNA-sense program, these 6 hpf data points were replaced by the mean of the other replicates. RNA-sense analysis was run with the following user-defined parameters: pVal switch=0.15, pVal FC=0.01, FC=2, threshold=100.
The code and example files for automatically performing the 3-step procedure are available in the R-package RNAsense that was developed jointly with the paper is available on Bioconductor https://bioconductor.org/, DOI: 10.18129/B9.bioc.RNAsense

Analysis of RNA-seq time curves using RNA-sense program

To identify maternal and zygotic transcripts in the wild-type, Step 1 of RNA-sense was performed for the WT condition, with the threshold 100 and pVal switch 0.15. Switching UP transcripts were considered zygotic, switching DOWN transcripts were considered maternal. To validate the assignment of zygotic and maternal transcripts identified in our study, we linked our transcript list to the previously defined zygotic and maternal groups from three studies[11,52,53] via ENSEMBL gene name. To assign the regulation status for each transcript, RNA-sense was performed for the WT against each mutant, with 2 fold change and p-value <0.05 cutoff (FC=2, pVal FC=0.05 at RNA-sense step 2). Zygotic transcript was considered to be downregulated in the mutant, if it was downregulated at one or more time points at or after the switch UP. Maternal transcript was considered to be upregulated in the mutant, if it was upregulated at one or more time points after the switch DOWN. The zygotic and maternal transcripts were grouped by regulation in the mutants to 8 non-overlapping groups each. The list of zygotic genes, downregulated or unchanged in the mutants, the list of maternal genes upregulated or unchanged in the mutants, and all accompanying information is provided in the Table S1.

To identify the genes zygotically upregulated in the mutants, we needed to consider the transcripts which are absent in the wild-type and appear in the mutants only. For this purpose, Step 1 of RNA-sense was performed for each of the mutant conditions against the wild-type, using the same parameters as above. The transcript was considered as zygotically upregulated in the mutant, if 1) the transcript was switching UP in this mutant 2) the transcript was upregulated at least 2 fold in the mutant compared to the wild-type, at at least one time point after the switch. This procedure resulted in the list of 1062 transcripts. Out of them, 27 transcripts were upregulated in all three and 224 in two mutant genotypes. These transcripts were assigned as upregulated to the genotype where maximal expression over the time curve was the highest (e.g. using this criteria, *nog1* was assigned as upregulated in MZ*sox19bspg*). Three non-overlapping groups were assigned: MZ*spg*>WT, MZ*sox19b*>WTand MZ*sox19bspg*>WT. The list of zygotic genes, upregulated in the mutants, and accompanying information is provided in the Table S2.

Grouping the zygotic transcripts by enhancer type.

Enhancers were linked to the closest transcription start sites of ENSEMBL genes within +/-20 kb and to all corresponding ENSEMBL transcripts. The number and type of enhancers linked to each expressed transcript is listed in the Tables S1 and S2. To group the transcripts by enhancer type, zygotic transcripts from Tables S1 and Table S2 were merged and redundancies were removed. The resulting list of 5264 transcripts was classified into 5 groups 225 CD(codependent), 284 PD (Pou5f3-dependent), 269 SD (Sox19b-dependent), 1610 U-only (unchanged only) and 2957 with no enhancers in +/- 20 kb. Transcript linked to at least one regulated enhancer (CD, PD and SD) was included into the respective group independent on the other enhancers linked to the same transcript; 72 transcripts linked to more than one type of regulated enhancers (CD, PD and SD) contribute to more than one group. U – only group included the transcripts linked only to unchanged, but not to regulated enhancers. The list of all zygotic genes classified by enhancer type, and accompanying information is provided in the Table S5

In-vitro nucleosome and Propeller Twist predictions

Nucleosome prediction program from Kaplan et al. 2009[94] was integrated into the Galaxy platform using the Galaxy tool SDK planemo (https://github.com/galaxyproject/planemo) and following the best practices for Galaxy tool development (http://galaxy-iuc-standards.readthedocs.io/en/latest/best_practices.html). The tool was uploaded into the european Galaxy ToolShed (ref. https://www.ncbi.nlm.nih.gov/pubmed/25001293) and is available at the european Galaxy instance. The genomic coordinates of Pou5f3, SoxB1 and control ChIP-seq peaks were extended to +/-5 kb to account for the edge effects, and *in-vitro* nucleosome prediction value was derived for every bp. For visualization, nucleosome predictions were converted to BigWig files in DeepTools2 and used for plotting of the central 3 kb. Propeller Twist° values for 1 kb Pou5f3, SoxB1 and control ChIP-seq peaks were derived using TFBS shape server at http://rohslab.cmb.usc.edu/TFBSshape[95]. The values +/- 150 bp around the summit were averaged and used for statistics.

Quantification and statistical analysis

Statistic comparisons of three or more samples were performed using one-way ANOVA and Tukey-Kramer Test in R. Two samples were compared using Student 2-tailed t-test. To estimate the over- or underrepresentation of enhancers linked to transcript groups, 2-tailed Fisher exact test was performed using GraphPad software. Enhancer types CD, PD and SD were tested for significant over- or underrepresentation in the transcript groups A-K. Groups A to H compared to all zygotic genes, downregulated or unchanged in the mutants (Table S1, Fig.3b). Groups IJK were compared to all zygotic genes, upregulated in the mutants (Table S2, Fig.4a).

Data and Software Availability

Software

RNA-sense package (version: version 0.99.27) developed during this study is available at https://bioconductor.org/packages/release/bioc/html/RNAsense.html
R: Version 3.6.1

Source data.

The source data underlying main Fig. 3b,d,g, Fig.4 c,d, Fig.5 g-j, Fig.6b,f,g, Fig. 7d-f, and Supplementary Fig. S5,S6, S7a-e, S8 i,j are provided as a Source Data file.

Data Resources (published and generated during this work)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Histone H3 (acetyl K27) rabbit, 1/100 dilution	Abcam plc., Cambridge, UK	ab 4729
Anti-Histone H3 (tri-methyl K4) rabbit, 1/100 dilution	Millipore Co., Temecula, California, USA	07-449
Bacterial and Virus Strains		
One Shot™ TOP10 chemically competent *E. coli*	Invitrogen™	C404003

Chemicals, Peptides, and Recombinant Proteins		
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Sigma-Aldrich Chemie GmbH, Germany	5056489001
Micrococcal Nuclease	Sigma-Aldrich Chemie GmbH, Germany	N3755-200UN
SYTOX Green	ThermoFisher SCIENTIFIC	S7020
Critical Commercial Assays		
Agencourt® AMPure® XP Beads	Beckmann Coulter, Krefeld, Germany	A63880
Agilent High Sensitivity DNA Kit	Agilent Technologies, Santa Clara, California, USA	5067-4626
Agilent RNA 6000 Nano Kit	Agilent Technologies, Santa Clara, California, USA	5067-1511
Dynabeads® Protein G	invitrogen Dynal AS, Oslo, Norway	10003D
E.Z.N.A® Cycle Pure Kit	Omega Biotek, Norcross, Georgia, USA	D6493-02
Microcon®-30 Centrifugal Filters	Merck Millipore, Darmstadt, Germany	MRCF0R030
mMESSAGE mMACHINE® SP6 transcription Kit	Ambion	10086184
NEBNext Ultra DNA Library Prep Kit for Illumina	New England Biolabs, Inc., Frankfurt a.M., Germany	E7370S
NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)	New England Biolabs, Inc., Frankfurt a.M., Germany	E7335S
RNeasy® Mini Kit	QIAGEN, Hilden, Germany	74104
Quant-iT™ PicoGreen® dsDNA Assay Kit	invitrogen™ Molecular Probes® Waltham, Massachusetts, USA	Q33120
Deposited Data		
ATAC-seq of the WT 4.3 hpf zebrafish embryos	Liu et al. 2018	GEO: GSE101779
ChIP-seq for H3K27ac and H3K4me3 in three genotypes, 4.3 hpf	this work	GEO: GSE143306
ChIP-seq for Pou5f3 and SoxB1, 5.3 hpf	Leichsenring et al., 2013	GEO: GSE39780
Consensus *pou, pou:sox* and *sox* motifs	This work	Supplementary Table S3
List of Pou5f3 and /or SoxB1 ChIP-seq peaks and control genomic regions	Data from Leichsenring et al., 2013, remapped to GRCz11 in this work	Supplementary Table S3
MNase-seq of MZ*sox19b* embryos, 4.3 hpf	this work	GEO: GSE125945
MNase-seq of WT and MZ*spg* embryos, 4.3 hpf	Veil et al., 2019	GEO: GSE109410
Time-resolved RNA-seq at 8 time points (2.5 hpf to 6 hpf, 30 min intervals) in the wild-type, MZ*spg*, MZ*sox19b* and MZ*sox19bspg* embryos	this work	GEO: GSE137424, Supplementary Tables S1 and S2
Zebrafish reference genome assembly danrer11/ GRCz11	Genome reference Consortium	https://www.ncbi.nlm.nih.gov/grc/zebrafish
Experimental Models: Organisms/Strains		

Wild-type zebrafish strain AB/TL	ZIRC	ZL1/ZL86
MZ*spg* zebrafish	Lunde et al., 2004	m793
MZ*sox19b* zebrafish	this work	m1434
Oligonucleotides		
MO3-Sox2 Sox2 Morpholino 5' - GAAAGTCTACCCCACCAGCCGTAAA - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-080329-1
MO4-Sox2 Sox2 Morpholino 5' - GAGAGGCTGCTGAAGTTACCTTAGC - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-080329-2
MO3-Sox3 Sox3 Morpholino 5' - TACATTCTTAAAAGTGGTGCCAAGC - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-3
MO4-Sox3 Sox3 Morpholino 5' - GAAGTCAGTCAAAAGTTCAGAGAGC - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-4
MO1-Sox19a Sox19a Morpholino 5' - GTACATGGCTGCCAACAGAAGTTAG - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-5
MO2-Sox19a Sox19a Morpholino 5' - AAAACGAGAGCGAGCCGTCTGTAAC - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-6
MO1-Sox19b Sox19b Morpholino 5' - GTACATCATGCCACTTCTCGCTTTG - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-7
MO2-Sox19b Sox19b Morpholino 5' - ACGAGCGAGCCTAATCAGGTCAAAC - 3'	Okuda et al., 2010	ZFIN ID: ZDB-MRPHLNO-100527-8
MOa-nog1 Noggin1 Morpholino 5' - GCGGGAAATCCATCCTTTTGAAATC - 3'	Dal-Pra et al., 2006	ZFIN ID: ZDB-MRPHLNO-080212-1
MO1-chrd Chordin Morpholino 5' - ATCCACAGCAGCCCCTCCATCATCC - 3'	Dal-Pra et al., 2006	ZFIN ID: ZDB-MRPHLNO-050221-6
PCR primer for genotyping *sox19b* mutants Sox19b-f1 5'-ATTTGGGGTGCTTTCTTCAGC-3'	this work	no
PCR primer for genotyping *sox19b* mutants Sox19b-r1 5'-GTTCTCCTGGGCCATCTTCC-3'	this work	no
sox19bF1: Sox19b PCR cloning primer sox19Bf1 with BamHI: 5'-GGGGGATCCATGGAGCACGAGCTGAAGAC-3',	this work	no
Sox19b PCR cloning primer Sox19bR1 with Xho1: 5'-GGGGCTCGAGTCAGATGTGAGTGAGGGGAAC-3'	this work	no
PCR primer for genotyping *spg* mutants spg-f1 GTCGTCTGACTGAACATTTTGC	this work	no
PCR primer for genotyping *spg* mutants spg-r1 GCAGTGATTCTGAGGAAGAGGT	this work	no
PCR primer for ChIP-seq control, positive reference tiparp_f_1 5' CGCTCCCAACTCCATGTATC-3'	this work	no
PCR primer for ChIP-seq control, positive reference tiparp_r_1 5'-AACGCAAGCCAAACGATCTC-3'	this work	no
PCR primer for ChIP-seq control, negative reference igsf2_f_2 5'-GAACTGCATTAGAGACCCAC-3'	this work	no
PCR primer for ChIP-seq control, negative reference igsf2_r_2 5'-CAATCAACTGGGAAAGCATGA-3'	this work	no

Recombinant DNA		
CS2+Sox19b plasmid for mRNA synthesis	this work	no
Software and Algorithms		
Bed Tools	Quinlan and Hall,2010	BED Tools in usegalaxy.eu
Bowtie2	Langmead and Salzberg, 2012	Bowtie2 in usegalaxy.eu
DeepTools2	Ramirez et al., 2016	deepTools in usegalaxy.eu
DESeq2	Love et al., 2014	DESeq2 in usegalaxy.eu
FeatureCounts	Liao et al., 2014	featureCounts in usegalaxy.eu
Galaxy server	Afgan et al., 2018	https://usegalaxy.eu/
GREAT: Genomic Regions Enrichment of Annotations Tool, version 3.0.0	Hiller et al., 2013	http://great.stanford.edu/great/public-3.0.0/html/
In-vitro nucleosome prediction program	Kaplan et al., 2009 and https://github.com/bgruening/galaxytools	Nucleosome Predictions in usegalaxy.eu
k-means clustering algorithm	Ramirez et al., 2016	Available option in plotheatmap in deepTools2 in usegalaxy.eu
MACS2	Ferg et al., 2007	MACS2 callpeak and MACS2 bdgpeakcall in usegalaxy.eu
R programming packages	see the complete list in the Supplementary data	see the complete list in the Supplementary data
RNA Star	Dobin et al., 2013	RNA Star in usegalaxy.eu
RNA-sense	this work	https://bioconductor.org/packages/release/bioc/html/RNAsense.html
TFBSshape	Yang et al., 2013	https://rohslab.usc.edu/TFBSshape/

References

1 Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146**, doi:10.1242/dev.161471 (2019).

2 Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proc Natl Acad Sci U S A* **112**, E1086-1095, doi:10.1073/pnas.1413990112 (2015).

3 Joseph, S. R. *et al.* Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife* **6**, doi:10.7554/eLife.23326 (2017).

4 Wilky, H., Chari, S., Govindan, J. & Amodeo, A. A. Histone concentration regulates the cell cycle and transcription in early development. *Development*, doi:10.1242/dev.177402 (2019).

5 Ferg, M. *et al.* The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J* **26**, 3945-3956, doi:10.1038/sj.emboj.7601821 (2007).

6 Chan, S. H. *et al.* Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Dev Cell* **49**, 867-881 e868, doi:10.1016/j.devcel.2019.05.037 (2019).

7 Liang, H. L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400-403, doi:10.1038/nature07388 (2008).

8 Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925-934, doi:10.1038/ng.3844 (2017).

9 Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*, 360-364., doi:10.1038/nature12632 (2013).

10 Leichsenring, M., Maes, J., Mossner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005-1009, doi:10.1126/science.1242527 (2013).

11 Gentsch, G. E., Spruce, T., Owens, N. D. L. & Smith, J. C. Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat Commun* **10**, 4269, doi:10.1038/s41467-019-12263-w (2019).

12 Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*, doi:10.1038/s41576-018-0087-x (2018).

13 Palfy, M., Schulze, G., Valen, E. & Vastenhouw, N. L. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS Genet* **16**, e1008546, doi:10.1371/journal.pgen.1008546 (2020).

14 Schulz, K. N. *et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**, 1715-1726, doi:10.1101/gr.192682.115 (2015).

15 Sun, Y. *et al.* Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res*, 1703–1714, doi:10.1101/gr.192542.115 (2015).

16 Veil, M., Yampolsky, L., Gruening, B. & Onichtchouk, D. Pou5f3, SoxB1, and Nanog remodel chromatin on High Nucleosome Affinity Regions at Zygotic Genome Activation. *Genome Res*, doi:10.1101/gr.240572.118 (2019).

17 Fernandez Garcia, M. *et al.* Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell* **75**, 921-932 e926, doi:10.1016/j.molcel.2019.06.009 (2019).
18 Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).
19 Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci U S A.* **107**, 22534-22539. doi: 22510.21073/pnas.0913805107. Epub 0913802010 Dec 0913805113. (2010).
20 Montserrat, N. *et al.* Generation of induced pluripotent stem cells from human renal proximal tubular cells with only two transcription factors, OCT4 and SOX2. *J Biol Chem* **287**, 24131-24138, doi:10.1074/jbc.M112.350413 (2012).
21 Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* **122**, 947-956. (2005).
22 Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274-1285, doi:10.1016/j.cell.2014.01.062 (2014).
23 Biddle, J. W., Nguyen, M. & Gunawardena, J. Negative reciprocity, not ordered assembly, underlies the interaction of Sox2 and Oct4 on DNA. *Elife* **8**, doi:10.7554/eLife.41017 (2019).
24 Li, S., Zheng, E. B., Zhao, L. & Liu, S. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep* **28**, 2689-2703 e2684, doi:10.1016/j.celrep.2019.07.103 (2019).
25 Friman, E. T. *et al.* Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *Elife* **8**, doi:10.7554/eLife.50087 (2019).
26 Lunde, K., Belting, H. G. & Driever, W. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. *Current biology : CB* **14**, 48-55 (2004).
27 Okuda, Y., Ogura, E., Kondoh, H. & Kamachi, Y. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. *PLoS Genetics* **6**, e1000936, doi:10.1371/journal.pgen.1000936 (2010).
28 Shih, Y. H. *et al.* SoxB1 transcription factors restrict organizer gene expression by repressing multiple events downstream of Wnt signalling. *Development* **137**, 2671-2681, doi:10.1242/dev.054130 (2010).
29 Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* **39**, e82, doi:10.1093/nar/gkr218 (2011).
30 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310 (1995).
31 El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193-197, doi:10.1038/s41586-019-1064-z (2019).
32 Eckerle, S., Ringler, M., Lecaudey, V., Nitschke, R. & Driever, W. Progesterone modulates microtubule dynamics and epiboly progression during zebrafish gastrulation. *Dev Biol* **434**, 249-266, doi:10.1016/j.ydbio.2017.12.016 (2018).
33 Song, S. *et al.* Pou5f1-dependent EGF expression controls E-cadherin endocytosis, cell adhesion, and zebrafish epiboly movements. *Developmental cell* **24**, 486-501, doi:10.1016/j.devcel.2013.01.016 (2013).
34 Belting, H. G. *et al.* Pou5f1 contributes to dorsoventral patterning by positive regulation of vox and modulation of fgf8a expression. *Dev biol* **356(2):323-36**, 323-336, doi:10.1016/j.ydbio.2011.05.660 (2011).
35 Dal-Pra, S., Furthauer, M., Van-Celst, J., Thisse, B. & Thisse, C. Noggin1 and Follistatin-like2 function redundantly to Chordin to antagonize BMP activity. *Developmental biology* **298**, 514-526, doi:10.1016/j.ydbio.2006.07.002 (2006).

36 Kishimoto, Y., Lee, K. H., Zon, L., Hammerschmidt, M. & Schulte-Merker, S. The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. *Development* **124**, 4457-4466 (1997).
37 De Robertis, E. M. Spemann's organizer and the self-regulation of embryonic fields. *Mech Develop* **126**, 925-941, doi:10.1016/j.mod.2009.08.004 (2009).
38 Shimizu, T. *et al.* Cooperative roles of Bozozok/Dharma and Nodal-related proteins in the formation of the dorsal organizer in zebrafish. *Mech Dev* **91**, 293-303 (2000).
39 Harvey, S. A. *et al.* Identification of the zebrafish maternal and paternal transcriptomes. *Development* **140**, 2703-2710, doi:10.1242/dev.095091 (2013).
40 Haberle, V. *et al.* Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**, 381-385, doi:10.1038/nature12974 (2014).
41 Stickney, H. L., Imai, Y., Draper, B., Moens, C. & Talbot, W. S. Zebrafish bmp4 functions during late gastrulation to specify ventroposterior cell fates. *Dev Biol* **310**, 71-84, doi:10.1016/j.ydbio.2007.07.027 (2007).
42 Kawahara, A., Wilm, T., Solnica-Krezel, L. & Dawid, I. B. Antagonistic role of vega1 and bozozok/dharma homeobox genes in organizer formation. *Proc Natl Acad Sci U S A* **97**, 12121-12126 (2000).
43 Melby, A. E., Beach, C., Mullins, M. & Kimelman, D. Patterning the early zebrafish by the opposing actions of bozozok and vox/vent. *Dev Biol* **224**, 275-285 (2000).
44 Shimizu, T. *et al.* A novel repressor-type homeobox gene, ved, is involved in dharma/bozozok-mediated dorsal organizer formation in zebrafish. *Mech Dev* **118**, 125-138 (2002).
45 Ruzicka, L. *et al.* The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res* **47**, D867-D873, doi:10.1093/nar/gky1090 (2019).
46 Kotkamp, K. *et al.* Pou5f1/Oct4 Promotes Cell Survival via Direct Activation of mych Expression during Zebrafish Gastrulation. *PLoS One* **9**, e92356, doi:10.1371/journal.pone.0092356 (2014).
47 Kotkamp, K., Mossner, R., Allen, A., Onichtchouk, D. & Driever, W. A Pou5f1/Oct4 dependent Klf2a, Klf2b, and Klf17 regulatory sub-network contributes to EVL and ectoderm development during zebrafish embryogenesis. *Developmental biology* **385**, 433-447, doi:10.1016/j.ydbio.2013.10.025 (2014).
48 Dick, A. *et al.* Essential role of Bmp7 (snailhouse) and its prodomain in dorsoventral patterning of the zebrafish embryo. *Development* **127**, 343-354 (2000).
49 Kwon, H. J., Bhat, N., Sweet, E. M., Cornell, R. A. & Riley, B. B. Identification of early requirements for preplacodal ectoderm and sensory organ development. *PLoS Genet* **6**, e1001133, doi:10.1371/journal.pgen.1001133 (2010).
50 Kikuchi, Y. *et al.* The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors. *Genes Dev* **14**, 1279-1289 (2000).
51 Kikuchi, Y. *et al.* casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. *Genes Dev* **15**, 1493-1505, doi:10.1101/gad.892301 (2001).
52 Poulain, M. & Lepage, T. Mezzo, a paired-like homeobox protein is an immediate target of Nodal signalling and regulates endoderm specification in zebrafish. *Development* **129**, 4901-4914 (2002).
53 Brown, J. L. *et al.* Transcriptional profiling of endogenous germ layer precursor cells identifies dusp4 as an essential gene in zebrafish endoderm specification. *Proc Natl Acad Sci U S A* **105**, 12337-12342, doi:10.1073/pnas.0805589105 (2008).

54 Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B. & Nusslein-Volhard, C. no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009-1015 (1994).
55 Dal-Pra, S., Thisse, C. & Thisse, B. FoxA transcription factors are essential for the development of dorsal axial structures. *Dev Biol* **350**, 484-495, doi:10.1016/j.ydbio.2010.12.018 (2011).
56 Bruce, A. E. *et al.* The maternally expressed zebrafish T-box gene eomesodermin regulates organizer formation. *Development* **130**, 5503-5517, doi:10.1242/dev.00763 (2003).
57 Nelson, A. C. *et al.* Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. *BMC Biol* **12**, 81, doi:10.1186/s12915-014-0081-5 (2014).
58 Slagle, C. E., Aoki, T. & Burdine, R. D. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and Eomesodermin. *PLoS Genet* **7**, e1002072, doi:10.1371/journal.pgen.1002072 (2011).
59 Talbot, W. S. *et al.* A homeobox gene essential for zebrafish notochord development. *Nature* **378**, 150-157, doi:10.1038/378150a0 (1995).
60 Bischof, J. & Driever, W. Regulation of hhex expression in the yolk syncytial layer, the potential Nieuwkoop center homolog in zebrafish. *Developmental biology* **276**, 552-562, doi:10.1016/j.ydbio.2004.09.035 (2004).
61 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).
62 Afek, A., Cohen, H., Barber-Zucker, S., Gordan, R. & Lukatsky, D. B. Nonconsensus Protein Binding to Repetitive DNA Sequence Elements Significantly Affects Eukaryotic Genomes. *PLoS Comput Biol* **11**, e1004429, doi:10.1371/journal.pcbi.1004429 (2015).
63 Samee, M. A. H., Bruneau, B. G. & Pollard, K. S. A De Novo Shape Motif Discovery Algorithm Reveals Preferences of Transcription Factors for DNA Shape Beyond Sequence Motifs. *Cell Syst* **8**, 27-42 e26, doi:10.1016/j.cels.2018.12.001 (2019).
64 Sato, Y. *et al.* Histone H3K27 acetylation precedes active transcription during zebrafish zygotic genome activation as revealed by live-cell analysis. *Development* **146**, doi:10.1242/dev.179127 (2019).
65 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).
66 Tan, C. & Takada, S. Nucleosome allostery in pioneer transcription factor binding. *Proc Natl Acad Sci U S A* **117**, 20586-20596, doi:10.1073/pnas.2005500117 (2020).
67 Gagnon, J. A., Obbad, K. & Schier, A. F. The primary role of zebrafish nanog is in extra-embryonic tissue. *Development* **145**, doi:10.1242/dev.147793 (2018).
68 Veil, M. *et al.* Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* **145**, doi:10.1242/dev.155366 (2018).
69 Paraiso, K. D. *et al.* Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. *Cell Rep* **27**, 2962-2977 e2965, doi:10.1016/j.celrep.2019.05.013 (2019).
70 Tosic, J. *et al.* Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat Cell Biol* **21**, 1518-1531, doi:10.1038/s41556-019-0423-1 (2019).
71 Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76-81, doi:10.1038/s41586-018-0549-5 (2018).

72 Neyfakh, A. A. X-ray inactivation of nuclei as method for studying their function in the early development of fishes. *Journal of embryology and experimental morphology* **7**, 173-192 (1959).
73 Neyfakh, A. A. Radiation Investigation of Nucleo-Cytoplasmic Interrelations in Morphogenesis and Biochemical Differentiation. *Nature* **201**, 880-884 (1964).
74 Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459 e420, doi:10.1016/j.cell.2016.12.016 (2017).