Pou5f3, SoxB1 and Nanog remodel chromatin on High Nucleosome Affinity Regions at Zygotic Genome Activation

Authors: Marina Veil[1], Lev Yampolsky[2,3], Björn Grüning[4,5], Daria Onichtchouk[1,6*].

Affiliations:

[1]Department of Developmental Biology, Institute of Biology I, Faculty of Biology, Albert Ludwigs University of Freiburg, Hauptstrasse 1, 79104, Freiburg, Germany

[2]Department of Biological Sciences, East Tennessee State University, Johnson City, TN, USA

[3]Zoological Institute, Basel University, Basel, Switzerland

[4]Department of Computer Science, Albert Ludwigs University of Freiburg, 79110, Freiburg, Germany

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

[6]BIOSS Center for Biological Signalling Studies, Albert Ludwigs University of Freiburg, Freiburg, Germany

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

Abstract:

The zebrafish embryo remains transcriptionally quiescent during the first 10 cell cycles. Only then Zygotic Genome Activation (ZGA) occurs and is accompanied by fast chromatin remodeling. At ZGA, homologs of mammalian stem cell transcription factors (TFs) Pou5f3/Oct4, Nanog and Sox19b bind to thousands of developmental enhancers to initiate transcription. So far, how these three ZGA TFs influence chromatin dynamics at ZGA has remained unresolved. To address this question, we analyzed before and after ZGA nucleosome positions in wild-type and Maternal-Zygotic (MZ) mutants for *pou5f3* and *nanog*. We show that Nanog, Sox19b and Pou5f3 bind to the High Nucleosome Affinity Regions (HNARs). HNARs are spanning over 600 bp, featuring high *in vivo* and predicted *in vitro* nucleosome occupancy and high propeller twist DNA shape value. Just prior to ZGA, Pou5f3 and Nanog non-specifically compete with histones and reduce nucleosome occupancy on strong nucleosome positioning sequences genome-wide. After ZGA, specific binding of Nanog and Pou5f3/SoxB1 complex is necessary to maintain open chromatin state on more than 6,000 HNARs. Nanog binds to the strongest nucleosome positioning sequence within HNAR, while the Pou5f3/SoxB1 complex binds to the flanks of it. We suggest a model, where the same intrinsic DNA properties of HNAR promote both high nucleosome occupancy and differential binding of TFs. We hope that our model will provide a useful framework for genome regulatory studies in the variety of biological systems.

Key words:

Pluripotency Transcription Factors, Zygotic Genome Activation, Nucleosome, Chromatin, High Nucleosome Affinity Region, Enhancer, Pou5f3, Oct4, SoxB1, Nanog, zebrafish, propeller twist, midblastula transition, maternal to zygotic transition, ZGA, MBT, MZT

Introduction

The beginning of development of multicellular organisms is driven by maternal products and occurs at the absence of transcription (Tadros and Lipshitz 2009). The main wave of the embryo's own transcription, Zygotic Genome Activation (ZGA), occurs several hours after the egg was fertilized. The mechanistic reasons for this transcriptional delay are not completely understood. In Drosophila and zebrafish, transcription after ZGA depends on a small number of maternal enhancer-binding Transcription Factors (TFs). Zygotic genome activators include Zelda in Drosophila (Liang et al. 2008), and three homologs of pluripotency TFs, Pou5f3/Oct4, Nanog and Sox19b in zebrafish (Lee et al. 2013; Leichsenring et al. 2013). Recent analysis of chromatin accessibility and gene expression implicated OCT4 also as ZGA activator in humans (Gao et al. 2018) and the transcription factor Nfya as ZGA activator in mice (Lu et al. 2016). The interactions of ZGA activators with chromatin are only starting to be addressed, and it is unclear if the same ZGA mechanisms are implemented in different animals.

The basic units of chromatin structure are the nucleosomes, which restrict the *in vivo* access of most TFs (Beato and Eisfeld 1997; Luo et al. 2014). While many TFs cannot bind their target site in the context of nucleosome DNA *in vitro*, cooperative interactions among multiple factors allow binding, even in the absence of defined orientation of their recognition motifs in DNA (Adams and Workman 1993, Zinzen et al. 2009). In steady-state cell culture systems, most of the TF binding sites fall within nucleosome-free DNA regions (Thurman et al. 2012), and it is difficult to address if the chromatin accessibility is a cause or consequence of TF binding. Studies of hematopoietic cell fate transitions suggested a hierarchical model, where a relatively small set of "pioneer" TFs collaboratively compete with nucleosomes to bind DNA in a cell-

type-specific manner. The binding of pioneer TFs, also called lineage-determining factors and master regulators, is proposed to prime DNA by moving nucleosomes and inducing the deposition of epigenetic enhancer marks. This enables concurrent or subsequent binding of signal-dependent transcription factors that direct regulated gene expression (Heinz et al. 2010; Trompouki et al. 2011; Li et al. 2018). Two alternative scenarios of TF-guided chromatin opening at enhancers during cell-fate transitions were suggested: a nucleosome-mediated cooperativity model, or specialized properties of pioneer transcription factors (Calo and Wysocka 2013; Slattery et al. 2014). The nucleosome-mediated cooperativity model postulates that if the recognition motifs for multiple cooperating TFs are located within the length of one nucleosome, the nucleosome will be removed (Mirny 2010; Moyle-Heyrman et al. 2011). The alternative model evokes special properties of pioneer factors to bind their DNA sites directly on nucleosomes (Iwafuchi-Doi and Zaret 2016). The pioneer factor model predicts that pioneer TFs will engage into the silent chromatin, while the majority of TFs will occupy nucleosome-depleted regions.

In *Drosophila*, the ZGA activator Zelda binds to its recognition motifs on thousands of enhancers already during nuclear cycle 8, four cell cycles ahead of the main wave of zygotic transcription (Harrison et al. 2011; Nien et al. 2011). Zelda binds to regions of high predicted nucleosome occupancy, and creates competency for other factors to bind the DNA by locally reducing nucleosome barriers, thus providing an example of a pioneer factor (Schulz et al. 2015; Sun et al. 2015). Genomic binding of the zebrafish ZGA activator Pou5f3 is detectable at 512-cell stage, just prior to the main ZGA wave (Leichsenring et al. 2013). Binding of Pou5f3 to the enhancers of its early regulated target genes may be significantly outcompeted by increasing

histone concentration at ZGA (3 hpf, 10th cell cycle) leading to the delayed transcriptional activation of the target genes. This histone-competition effect is not specific for Pou5f3 and Sox19b, and has also been detected for the heterologous factor tTA-VP16 (Joseph et al. 2017), suggesting that excess of maternally provided histones would not permit pre-binding of any TF before ZGA. Considering the data of Joseph et al. (2017), the specificity and dynamics of binding to the genomic sites before ZGA is likely different for Zelda in *Drosophila* and three ZGA activators in zebrafish. In zebrafish, nucleosome positioning signals guide the transcription start site (TSS) selection after ZGA, but not before, suggesting that the rules of nucleosome positioning may change over ZGA (Haberle et al., 2014).

Genetic ablation of maternal and zygotic expression of Pou5f3 and Nanog results in severe pleiotropic phenotypes, global deregulation of transcription and gastrulation arrest (Lunde et al. 2004; Onichtchouk et al 2010; Gagnon et al. 2018; Veil et al. 2018), suggesting that these factors act at genome-wide level. However, it is not clear, whether Pou5f3, Sox19b and Nanog preferentially target to the regions occupied by nucleosomes prior to ZGA, in the manner their mammalian homologs do in ES cells and during reprogramming of human fibroblasts (Teif et al. 2012; Soufi et al. 2015; Voong et al. 2016). It is also unclear, if Pou5f3, Sox19b or Nanog are able to remove nucleosomes from the regions they bind.

To address these questions, we analyzed the distribution of nucleosomes in *pou5f3* and *nanog* Maternal-Zygotic (MZ) mutants MZ*spg* and MZ*nanog*, and compared them to the wild-type, using MNase-seq. We show that the binding sites for Pou5f3, SoxB1 and Nanog are closed by nucleosomes before ZGA. During the 9th cell cycle immediately preceding ZGA, Nanog and

Pou5f3 non-specifically compete with histones genome-wide and reduce nucleosome occupancy on the strong nucleosome positioning sequences. 1.3 hours after ZGA, nucleosome positions are stabilized, so that sequence-specific binding of single TFs (Pou5f3 or Nanog) or double combinations of TFs (e.g. Pou5f3 and SoxB1, Nanog and SoxB1, Pou5f3 and Nanog) cannot efficiently outcompete nucleosomes. The nucleosome clearance post-ZGA occurs as a result of synergistic interaction of Pou5f3/SoxB1 complex and Nanog, which bind to High Nucleosome Affinity Regions (HNARs).

Results

Transcription Factor Binding Sites (TFBS) and control groups taken for analysis.

Our analysis strategy was to categorize the genomic regions by their binding of Pou5f3 (P), SoxB1 (S), and Nanog (N), separately or in combination, using standard criteria. We set to compare the relative changes in nucleosome occupancy for each of these groups between MZ*spg*, MZ*nanog* mutants and the wild-type, using random genomic regions for control. In cases in which we found statistically significant differences between the groups, we evaluated to which extent they can be explained by sequence-specific binding of the TFs to their consensus recognition motifs. We use the terms “consensus binding” and “nonconsensus binding” suggested by Afek et al. (2015), to distinguish between TF binding to its cognate consensus motifs on DNA (“consensus binding”), and TF binding at the absence of motifs (“nonconsensus binding”). For this analysis, we focused on the transcription factor binding sites (TFBS) defined by position of ChIP-seq peaks from our and others previous work (Leichsenring et al. 2013, Xu et al. 2012), and on the group of randomly taken control regions of similar size. We sorted the TFBS according to the occupancy by each single TF separately (P and N groups), by each

combination of two TFs (PN, PS and SN groups), and by all three TFs (PSN group) post-ZGA. Further, we identified an additional group of genomic regions occupied by Pou5f3 only pre-ZGA, which we termed Ppre (Methods, Fig. 1A, Fig. S1, Table S1). We next performed *de-novo* motif finding to estimate, score and compare the relative abundance of the sequence-specific recognition consensus motifs for each TF in each group. In addition to the sequences closely matching known Pou5f3, SoxB1 and Nanog consensus motifs (*pou:sox*, *sox*, *nanog1* and *nanog2*), we detected enrichment for di- and tri-nucleotide repeats, which were previously shown to coincide with enhancers (Yanez-Cuna et al. 2014, Fig. 1B, Fig. S2, Table S2). As expected, consensus motifs for Nanog (Loh et al., 2006; Xu et al. 2012,) and SoxB1 (Salmon-Divon et al. 2010) were enriched in all regions bound by the respective factors. The bipartite *pou:sox* motif, which promotes assembly of Pou5f3/SoxB1 complexes on DNA (Remenyi et al. 2002; Chen et al. 2008; Leichsenring et al. 2013) was only enriched at the regions co-bound by Pou5f3 and SoxB1 (in PSN and PS), but not at the regions bound by Pou5f3 alone (P or Ppre groups). This indicated that Pou5f3 binds to its consensus motifs only at the presence of SoxB1. For the large group of genomic regions bound by Pou5f3 pre-ZGA only (Ppre group), no motif could be recovered. Regions bound by Pou5f3 alone post-ZGA (P group) were highly enriched in ATGG repeats (Fig. 1B, Fig. S2).

Pou5f3, SoxB1 and Nanog preferentially bind to the enhancers of developmental genes with patterned expression, as reported previously (Xu et 2012; Leichsenring et al. 2013). We compared the enrichment in developmental enhancers in the groups defined above, using GREAT analysis (McLean et al. 2010). The frequency of enhancers within PSN group was an order of magnitude higher than in N, PS, PN, SN, Ppre groups ($p = e^{-234}$ in binomial test for GO

"developmental process" in PSN , $p=e^{-13-33}$ in binomial test for GO categories related to development, transcriptional regulation and patterned gene expression for N, PS, PN, SN, Ppre groups, Fig. S3). We found no enrichment for P group, indicating that nonconsensus binding of Pou5f3 alone post-ZGA does not mark enhancers. Surprisingly, the regions of Pou5f3-only pre-ZGA nonconsensus binding were moderately enriched in enhancers (e^{-28} in binomial test for GO "regulation of transcription").

Pou5f3 and Nanog reduce nucleosome density on the regions of high *in-vivo* nucleosome occupancy before ZGA.

We isolated chromatin from the wild-type (WT), MZ*spg* and MZ*nanog* mutants at pre-ZGA (512-cell stage, last cell cycle before ZGA) and post-ZGA (dome stage, 1.3 hours after ZGA), and performed micrococcal nuclease treatment followed by deep sequencing of the resulting fragments (MNase-seq) as described (Zhang et al. 2014, Methods, Fig. S4, Table S3). At pre-ZGA, all TFBS had higher nucleosome occupancy than surrounding sequences (Fig. 1C). The nucleosome occupancy in both mutants was significantly higher than in the wild-type within all TFBS, except for P group in MZ*nanog* (Fig. 1D-F, H-J, Fig. S5A-C, Fig. S6A, Table S4 for 1-way ANOVA). To evaluate the possible contribution of TF consensus binding to nucleosome clearance, we compared normalized nucleosome occupancy difference between each mutant and WT (defined as Δmut MZ*spg* and Δmut MZ*nanog*, Fig. S5) within PSN, N or PS groups with 0,1,2,3, and >=4 motifs. Normalized nucleosome occupancy difference did not increase with the number of cognate motifs for the missing TF per genomic region in any of the analyzed groups (PSN, N or PS, Fig. 2G,K, Table S5 for 1-way ANOVA), suggesting that consensus binding of Nanog or Pou5f3 is not involved in nucleosome clearance pre-ZGA. Δmut MZ*spg* and Δmut

MZ*nanog* were still significantly higher on nonconsensus TFBS (DNA sequences that lack known consensus motifs, Fig. 2G,K, "0" point), compared to random control genomic regions (Fig. 2G,K, "Co" point, Table S6 for Student's T-test).

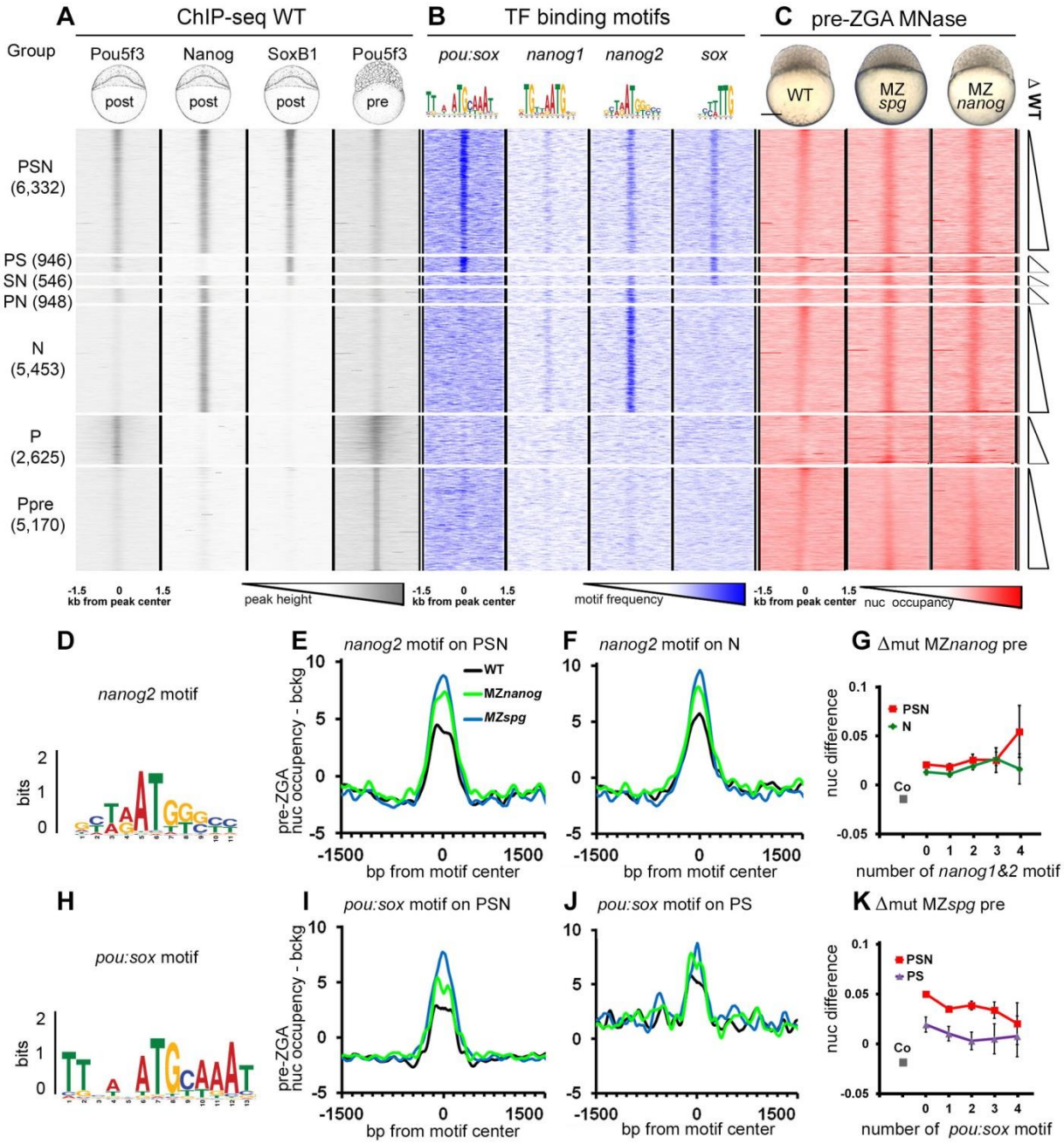

Fig. 1. Pou5f3, SoxB1 and Nanog bind to closed chromatin regions before ZGA and reduce nucleosome occupancy non-specifically. (*A-C*) Heat map of 6 groups of genomic regions defined by combinatorial binding of Pou5f3, Nanog and SoxB1 post-ZGA, as indicated at the left, and Ppre group

bound by Pou5f3 only pre-ZGA. 3 kb genomic regions were aligned at the ChIP-seq peak centers. Within each group, the data were sorted by ascending difference between WT pre-ZGA and post-ZGA nucleosome occupancy (ΔWT post-pre). Scale bar indicates 200 µm. (*A*) TF binding (*B*) Occurrence of consensus motifs. (*C*) Pre-ZGA nucleosome occupancy in the embryos of indicated genotypes. (*D,H*) *nanog2* and *pou:sox* consensus motifs (PWMs) (*E,F, I,J*) Mean nucleosome occupancy plots centered on *nanog2* or *pou:sox* motifs, X-axis: bp from the motif center. Y-axis: nucleosome occupancy in RPKM minus background. (*G,K*) Nucleosome occupancy difference between the WT and indicated mutants (Δmut, Y-axis) with 0, 1, 2, 3 and ≥4 motifs and control.

Taken together, our data suggested that pre-ZGA, Pou5f3 and Nanog compete with nucleosomes on the regions of high *in-vivo* nucleosome occupancy. Pre-ZGA competition with nucleosomes does not depend on consensus binding of TFs to their cognate motifs.

Pou5f3/SoxB1 and Nanog synergistically deplete nucleosomes on their consensus motifs post-ZGA.

Post-ZGA, a central nucleosome-depleted region became visible in the regions simultaneously bound by Pou5f3, Nanog and SoxB1 (PSN) in the WT but not in the MZ*nanog* and MZ*spg* mutants. For all other groups (single or dual P, S or N occupation sites), TFBS remained occupied by nucleosomes in all three genotypes (Fig. 2A). Post-ZGA nucleosome occupancy in WT was reduced in the PSN group when compared to pre-ZGA, while it was increased in all other TFBS groups, and unchanged in the random control group (ΔWT post-pre, Fig. S5A,D, Table S4 for 1-way ANOVA). The nucleosome occupancy in both MZ*spg* and MZ*nanog* mutants was strongly increased in the PSN group compared to WT, indicating that both Pou5f3 and Nanog are involved in nucleosome clearance (Fig. S5A,E,F, Table S4). To evaluate the contribution of Nanog consensus binding to post-ZGA nucleosome clearance, we compared nucleosome occupancy profiles centered on *nanog2* and *nanog1* motifs for regions of triple TF

binding (PSN) versus Nanog binding alone (N). A 320 bp wide nucleosome gap appeared on both *nanog* motifs in WT on PSN group regions, which was strongly reduced in MZ*spg* as well as in MZ*nanog* mutants; in contrast, *nanog* motifs in the N group remained covered with nucleosomes in all three genotypes (Fig. 2B,C, Fig. S6B). The normalized nucleosome occupancy difference between MZ*nanog* and WT (Δmut MZ*nanog*) increased with the number of *nanog* consensus motifs in PSN, but not in N group regions (Fig. 2D, Fig.S7, Table S5). Thus, consensus Nanog binding induced nucleosome clearance only when Pou5f3 and SoxB1 binding sites were present nearby. Similarly, the Pou5f3/SoxB1 consensus binding to *pou:sox* motifs resulted in nucleosome clearance when Nanog was bound nearby (compare Fig. 2E (PSN) and 2F (PS)).

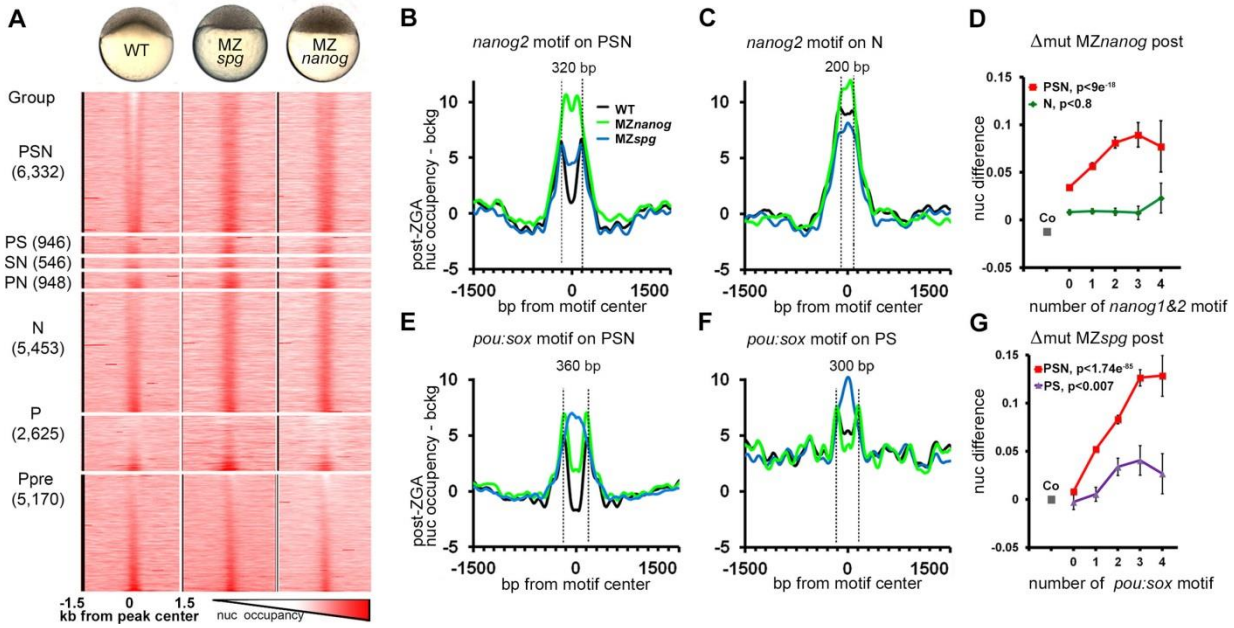

Fig. 2. Post-ZGA nucleosome clearance by Pou5f3 and Nanog occurs only in PSN regions. (*A-C*) Heat map of post-ZGA nucleosome occupancy, 7 groups of genomic regions as indicated at the left, and Ppre group bound by Pou5f3 only pre-ZGA. 3 kb genomic regions were aligned at the ChIP-seq peak centers. Within each group, the data were sorted by ascending difference between WT pre-ZGA and post-ZGA nucleosome occupancy (ΔWT post-pre). (*B,C,E,F*) Post-ZGA mean nucleosome occupancy plots centered on *nanog2* or *pou:sox* motifs, X-axis: bp from the motif center. Y-axis: nucleosome occupancy

in RPKM minus background. Gray dashed lines indicate the width of the region around the motif, where the nucleosome occupancy changes between the WT and mutants are evident. (*D,G*) Nucleosome occupancy difference between the WT and indicated mutants (Δmut, Y-axis) with 0, 1, 2, 3 and ≥4 motifs and control.

The normalized nucleosome occupancy difference between MZ*spg* and WT (Δmut MZ*spg*) strongly increased with the number of *pou:sox* motifs in PSN, but only marginally in PS group regions (Fig. 2G, Fig.S7, Table S5). Finally, on *sox* motifs, nucleosome clearance occurred only on PSN, but not on PS regions (Fig. S6B). Taken together, our data show that strong post-ZGA nucleosome depletion occurs shortly after ZGA on the regions bound by Pou5f3, Nanog and SoxB1 (PSN) and requires synergistic action of all three TFs. Post-ZGA nucleosome clearance on PSN regions depends on consensus binding of Pou5f3/SoxB1 and Nanog to their cognate motifs.

Nonconsensus binding of Pou5f3, SoxB1 and Nanog to the 600 bp long High Nucleosome Affinity Regions (HNARs).

The reasons for high nucleosome occupancy of pre-ZGA TFBS could be intrinsic DNA features, favoring nucleosome occupancy. Nucleosomes cover 75-90% of the genome (Kornberg 1974) and can form virtually anywhere, but the properties of the underlying DNA may change nucleosome affinity up to several orders of magnitude (Field et al. 2008). Increased GC content and 10 bp periodic fluctuations of dinucleotides underlie the high ability of DNA to wrap around histone octamers, while AT-rich sequences strongly disfavor nucleosome formation (Satchwell et al. 1986; Field et al. 2008; Chung and Vingron 2009; Kaplan et al. 2009; Tillo and Hughes 2009). To evaluate if Pou5f3, SoxB1 and Nanog preferentially recognize nucleosome-positioning signals, we used a nucleosome prediction program based on *in vitro* nucleosome DNA sequence

preferences mentioned above (Kaplan et al. 2009). This program predicts nucleosome positioning scores by a probability (from 0 to 1) for each nucleotide position to be located in the nucleosome center (or dyad). We selected the maximal [nucmax] score (for the most likely predicted dyad position) around 320 bp for each TFBS (ChIP-seq peaks) and control regions (Table S1) and compared them across the groups. [Nucmax] predicted scores of all (individual, double- and triple-bound) TFBS were higher than respective scores for random control regions (Fig. S8A). Notably, the [nucmax] scores of TFBS groups where Nanog binds (PSN, SN, PN and N) were significantly higher than of those of the groups where Nanog does not bind (P, Ppre and PS, Fig. S8B, Table S7).

Minimum nucleosome prediction scores selected within 320 bp [nucmin], corresponded most likely to predicted inter-nucleosomal sequences within the region (Segal et al. 2006). In all TFBS [nucmin] scores were also significantly higher than respective scores for random control regions (Fig. S8C,D, Table S8). The difference in [nucmin] predictions agreed with experimental data: experimental nucleosome occupancy of random control regions aligned on [nucmin] positions was below genomic average, as predicted, but it was above the genomic average in [nucmin] positions of all TFBS (Fig. S8E-L and data not shown). These data implied that high nucleosome affinity regions underlying TFBS span at least 320 bp – more than one nucleosome length.

TF recognition motif sequences may themselves contain high predicted nucleosome occupancy features, or, alternatively, the high predicted nucleosome occupancy may be encoded by sequences neighboring the recognition motifs. To distinguish between these possibilities, we compared predicted nucleosome occupancy of the motifs which are actually bound by Nanog,

SoxB1 or Pou5f3/SoxB1 (motifs within 320 bp of respective TFBS), with predicted nucleosome occupancy of randomly occurring unbound motifs (motifs in a distance from 1 to 1.5 kb from TFBS center). All bound motifs could be discriminated from unbound by higher predicted nucleosome occupancy of neighboring sequences (Fig. 3A-C, Fig.S9). Intriguingly, nucleosome prediction profiles around *nanog1* and *nanog2* motifs (Fig. 3A,B) were higher and different in shape from the profiles on *pou:sox* motifs (Fig. 3C), while *pou:sox* motif profiles were similar to the profiles on the regions bound by Pou5f3 pre-ZGA (Ppre Fig. 3D).

There is no agreement in the literature related to intrinsic nucleosome occupancy features within regulatory sequences. Under-representation of nucleosome positioning sequences in enhancers (Daenen et al. 2008; Papatsenko et al. 2009; Khoueiry et al. 2010; He et al. 2010), but also their over-representation (Tillo et al. 2010; Barozzi et al. 2014; Sun et al. 2015) have been reported. Our data strongly suggest that intrinsic nucleosome occupancy in zebrafish early enhancers is high, but in the view of the contradictory literature we had to confirm this conclusion by using an independent approach. Propeller twist (PT) is the angle between the heterocycles in the two complementary bases in a base pair, a DNA shape parameter, which is not directly encoded by primary sequence. PT values have been related to the flexibility of DNA bending around proteins (el Hassan and Calladine 1996), and strongly correlate with the ability of DNA to wrap around nucleosomes in yeast (Lee et al. 2007, Gan et al. 2012). Intrinsic nucleosome-forming preferences of the DNA on the known PSN-bound enhancers of *pou5f3* (Parvin et al. 2008), *vox* and *vent* genes (Belting et al. 2011), were independently captured by the primary sequence–derived model of Kaplan et al. (2009) and high values of propeller twist (Fig. 3E, Fig. S10).

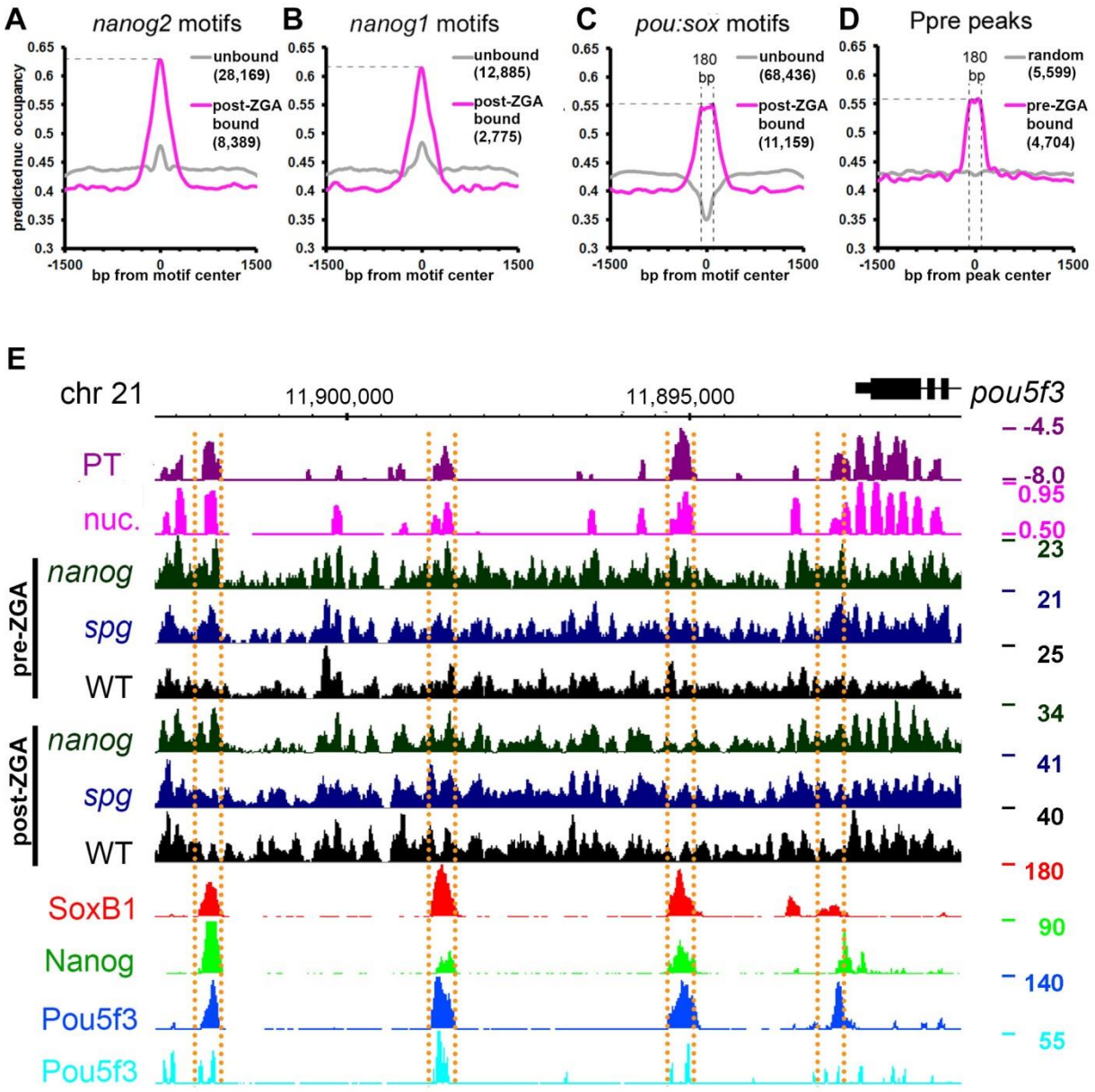

Fig. 3. Pou5f3, SoxB1 and Nanog recognize the regions of high predicted nucleosome occupancy and propeller twist. (*A-D*) Nucleosome prediction plots, numbers of genomic regions are indicated on the legend. (*A-C*) Mean predicted nucleosome occupancy on the indicated motifs bound by respective TFs (magenta, motifs +/-160 bp from TF binding peak center) was compared with the unbound motifs with the same sequence composition (gray, motifs +/- 1-1.5 kb from TF binding peak center). (*D*) Mean predicted nucleosome occupancy centered on Pou5f3 pre-MBT TF binding peaks (magenta) was compared with random sequences (gray). (*D*) 20 kb upstream genomic region of *pou5f3* gene. From top to bottom: PT- propeller twist (°, smoothened with 80 bp moving average), nuc - *in-vitro* nucleosome occupancy prediction, nucleosome occupancy plots as indicated, post-ZGA TF peaks, pre-ZGA Pou5f3 TF peaks. Orange lines mark three enhancers and promoter of *pou5f3*.

We aimed to further describe the length and structure of putative enhancers, which underlie P, S and N TFBSs, by combining PT and the nucleosome prediction model of Kaplan et al. (2009). Since it is unknown, how the features captured by sequence-based nucleosome positioning programs are reflected in PT values, we started from the control sequences. We randomly selected 5,000-6,000 genomic sequences from human and yeast genomes as 320 bp control sequences and found [nucmax] predicted nucleosome scores (dyad) within each region by the model of Kaplan et al. (2009). Then we plotted the average PT values with 1 bp resolution for zebrafish, human and yeast aligned on the [nucmax] position. As shown in Fig. 4A, a sharp increase in PT marked the central 150 bp nucleosome footprint, which was embedded into a 300 bp periodic shape. The differences in the height of PT graphs roughly corresponded to the median GC content in zebrafish, yeast and human genomes (36.8%, 38.4% and 40.9%, respectively) as provided by NCBI.

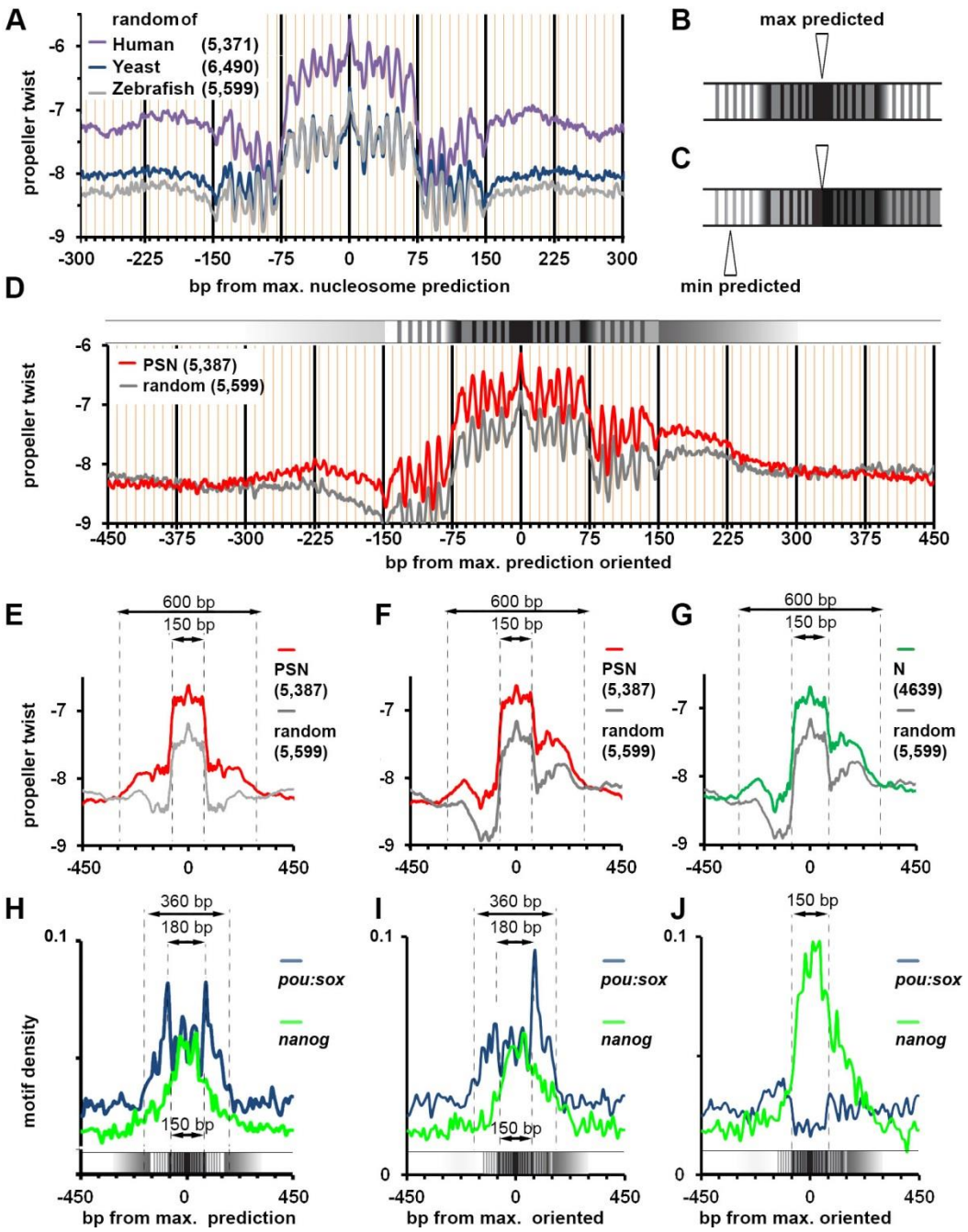

Fig. 4. Nucleosome positioning features and TF motifs on High Nucleosome Affinity Regions. (*A*) Un-smoothened propeller twist (PT) summary plot of random zebrafish, yeast and human 320 bp genomic regions aligned at [nucmax] prediction by Kaplan et al. (2009; average values per bp, yellow X-axis gridlines 10 bp apart). (*B*) Scheme of 300 central bp in symmetric PT plot. (*C*) Scheme of oriented PT plot: black shading indicated higher PT values in (*B*,*C*). (*D*) Un-smoothened oriented PT plot of PSN and control genomic regions aligned at [nucmax]. (*E*-*G*) Smoothened PT plots (80 bp moving average) of PSN and control genomic regions, aligned at [nucmax], symmetric (*E*) or oriented (*F*,*G*). (*H*-*J*) Distribution of Nanog and Pou5f3/SoxB1 recognition motifs for the groups shown above. Motif density in bp motif per bp sequence. Black shading indicates higher PT values.

The ~10 bp periodicity for PT prediction within the central 300 bp of DNA represents a single rotational frame, which was reported to stabilize nucleosome positioning (see Fig. S11 and Discussion). We next used the [nucmax] and [nucmin] positions as viewpoints to estimate the differences between TFBS and random nucleosome footprints using PT plots. To uncover possible hidden asymmetry of the nucleosome positioning signals, we also analyzed "oriented plots", where each genomic region centered on [nucmax] was oriented according to the increase in predicted nucleosome occupancy (Fig. 4C, Fig.12 A-D, Methods, Table S1). Most of the single or combinatorial TFBSs differ from the control group by higher propeller twist values over ~600 bp around [nucmax]; with the exception of P group (Fig. 4D-F, Fig. S12-S13). Higher PT values on the regions surrounding TFBSs correlated with relative increase of GC content over the average genomic value (Fig. S14), and with shifts in two other DNA shape parameters (Fig. S15). We define here the 600 bp DNA region with high predicted nucleosome occupancy/propeller twist/GC content, which is centered on 300 bp rotational periodicity frame, as a High Nucleosome Affinity Region (HNAR). *Pou:sox* motifs were distributed over the whole 300 bp central periodic shape in PSN (Fig. 4H-J, Fig.S16A) and PS (Fig. S16B) genomic regions. In the symmetric projection of PSN regions the highest density of *pou:sox* motifs was observed as two peaks 180 bp apart from each other (Fig. 4H). In the oriented projection of PSN group (Fig. 4F), *pou:sox* motifs added up in one peak approximately 90 bp from the [nucmax] within the higher PT flank around the footprint (Fig. 4I, Fig. S16A). Most of *nanog* motifs were located within the central nucleosome footprints in both PSN and N genomic regions (Fig. 4G,I-J, Fig. S16A,C).

In summary, our observations indicate that Pou5f3 pre-ZGA, Pou5f3/SoxB1 complex, SoxB1 and Nanog recognize and bind exclusively high nucleosome affinity regions (HNARs), in a

manner characteristic for pioneer factors. Ppre TFBS did not show enrichment for any specific motif (Fig. 1B), thus nonconsensus binding of Pou5f3 pre-ZGA to HNARs must employ some features not directly encoded by DNA primary sequence, e.g. DNA shape. Pou5f3 pre-ZGA, Pou5f3/SoxB1 complex and Nanog recognize distinct positions within HNARs: Nanog binds to the central nucleosome footprint (Fig. 3A,B, Fig4 J), while pre-ZGA Pou5f3 and post-ZGA Pou5f3/SoxB1 complex preferentially binds to the putative linker region outside of the central nucleosome footprint (Fig. 3C,D, Fig. 4H,I).

Zygotic nucleosome stabilization and switch from non-specific pre-ZGA to specific post-ZGA TF-dependent nucleosome clearance.

We analyzed next, how the TFs affect redistribution of real nucleosomes relative to the strongest predicted central nucleosome positioning sequence of the HNARs and control random nucleosome footprints. Pre-ZGA, Pou5f3 and Nanog-dependent nucleosome density reduction of the central nucleosome footprint was detectable in all HNARs and also in random control nucleosome footprints (Fig. 5A-C, Fig. S17A-D,I-L). Post-ZGA, nucleosome occupancy of random control nucleosome footprint in the wild-type was increased compared to pre-ZGA (Fig.5E, compare to Fig.5A), and not affected in the mutants. The single peak of max. real nucleosome density corresponded to prediction (red arrow in Fig. 5E). In all TFBS groups, except PSN, nucleosome occupancy of the strongest predicted footprint increased compared to pre-ZGA (Fig. 5G, Fig. S17E-H,M-P). In contrast, in the wild-type PSN group post-ZGA, peaks of the real nucleosome density shifted further away from the central footprint, so that the real and

predicted nucleosome occupancy were reversed (Fig. 5F, Fig. S17E, see also Fig.S18 and Fig. S19 for the oriented plots centered on [nucmax] and [nucmin], respectively).

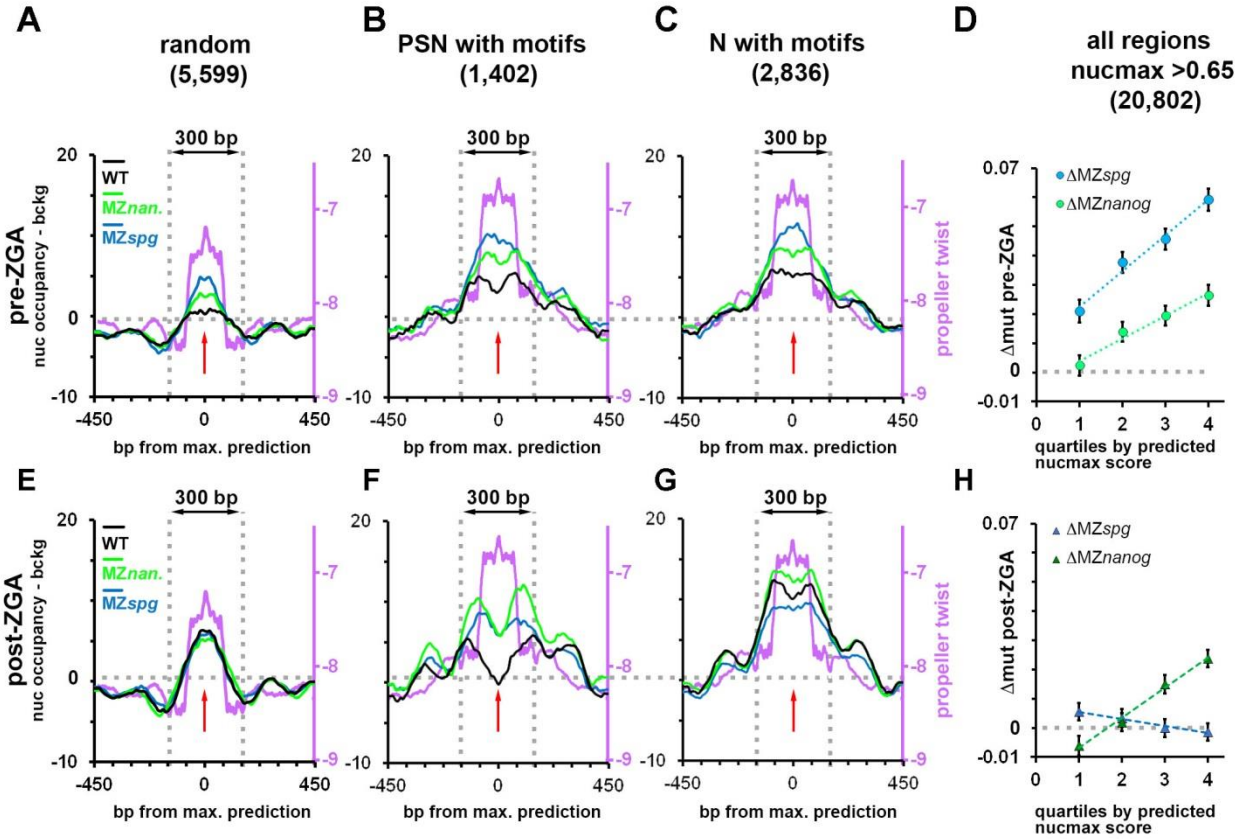

Fig. 5. Pre-ZGA and post-ZGA changes of real vs. predicted maximal nucleosome density on HNARs vs. control in three genotypes. (*A-C, E-G*) Genomic regions indicated on top aligned at strongest predicted dyad ([nucmax]), red arrows**.** Violet lines show smoothened PT values over 80 bp. Vertical gray dotted lines 300 bp apart show the borders of central rotational frame, horizontal dotted lines show the background. Right axis- PT, left axis – nucleosome occupancy pre-ZGA (A-C) and post-ZGA (E-G), unsmoothened. (*D,H*) Dependencies of Δmut from central nucleosome footprint strength pre-ZGA (D) and post-ZGA (H). Quartiles by ascending nucleosome prediction score.

The nucleosome-mediated cooperativity model predicts that stabilization of nucleosomes by strong nucleosome-positioning sequence will increase the nucleosome displacement effect of

multiple cooperating TFs, binding within one nucleosome footprint (150 bp, Mirny 2010). We tested next, if the specific or non-specific effects of Pou5f3 and Nanog on nucleosome density fit to this scenario. We divided the genomic regions with [nucmax] 0.65-0.99 (the range of [nucmax] for more than 90% of TFBS, Fig.S8A) to quartiles by [nucmax] (see Fig. S20 for PT graphs). Then we compared nucleosome occupancy increase in the mutants over the wild type (Δmut MZ*spg* and Δmut MZ*nanog*) per quartile. Pre-ZGA, the nucleosome displacement by Pou5f3 and Nanog increased with predicted nucleosome footprint strength in the whole data set (Fig. 5D). This tendency was significant and the same for the random control group and TFBS groups (Fig. S21, Table S9). Taken together with the nonconsensus action of Pou5f3 and Nanog at pre-ZGA (Fig. 1G,K), and their ability to displace a nucleosome from random nucleosome footprints (Fig. 5A), this may mean that both TFs compete with nucleosomes on strong nucleosome-positioning sequences genome-wide. At post-ZGA, Δmut MZ*nanog* increased, but Δmut MZ*spg* rather decreased with predicted nucleosome strength in random control group and TFBS groups (Fig. 5H, Fig. S21, Table S9). Therefore, post-ZGA synergy of Pou5f3 and Nanog is unlikely due to the nucleosome-mediated cooperativity scenario.

Discussion.

Here, we used two independent methods to show that the affinity of Pou5f3, Nanog and nucleosomes to the same locations within the genome is supported by the intrinsic DNA features favoring high *in vitro* nucleosome occupancy. We define the regions, which span 600 bp of sequence length of high predicted nucleosome occupancy/PT/GC content, as "High Nucleosome Affinity Regions" (HNARs) to distinguish them from strong nucleosome positioning sequences of one nucleosome length. We further show that Pou5f3 and Nanog binding is the cause, but not consequence of chromatin opening on HNARs, which ranks them as pioneer factors. ZGA TFs are involved in two processes: non-specific competition with histones on a broad range of strong nucleosome footprints before ZGA, and maintenance of open chromatin state on by synergistic consensus binding to HNARs after ZGA.

HNAR is an area of competition between pioneer transcription factors and nucleosomes.

The length of HNARs well matches the average size of developmental enhancers (Levine and Tjian 2003). The PT plots (Fig. 4) visualized two main sequence features, which define the strength of *in vitro* nucleosome positioning: 10-bp rotational periodicity and increased GC content (Satchwell et al. 1986; Field et al. 2008; Chung and Vingron 2009; Kaplan et al. 2009; Tillo and Hughes 2009). A theoretical model suggests that the same dinucleotide (i.e. GC or GA) repeated at 10 bp intervals leads to intrinsic curvature that favors the wrapping of the DNA around the histone octamer with the same orientation relative to the histone-octamer surface, for instance having the major groove facing inwards (Albert et al. 2007). Rotational positioning, e.g. tendency of nucleosomes to occupy ''in-phase'' distances that are multiples of 10 bp, was traced

in large *in vivo* bulk MNase-seq data up to 300 bp (Gaffney et al. 2012) or less (Zhang et al. 2008; Valouev et al. 2011).

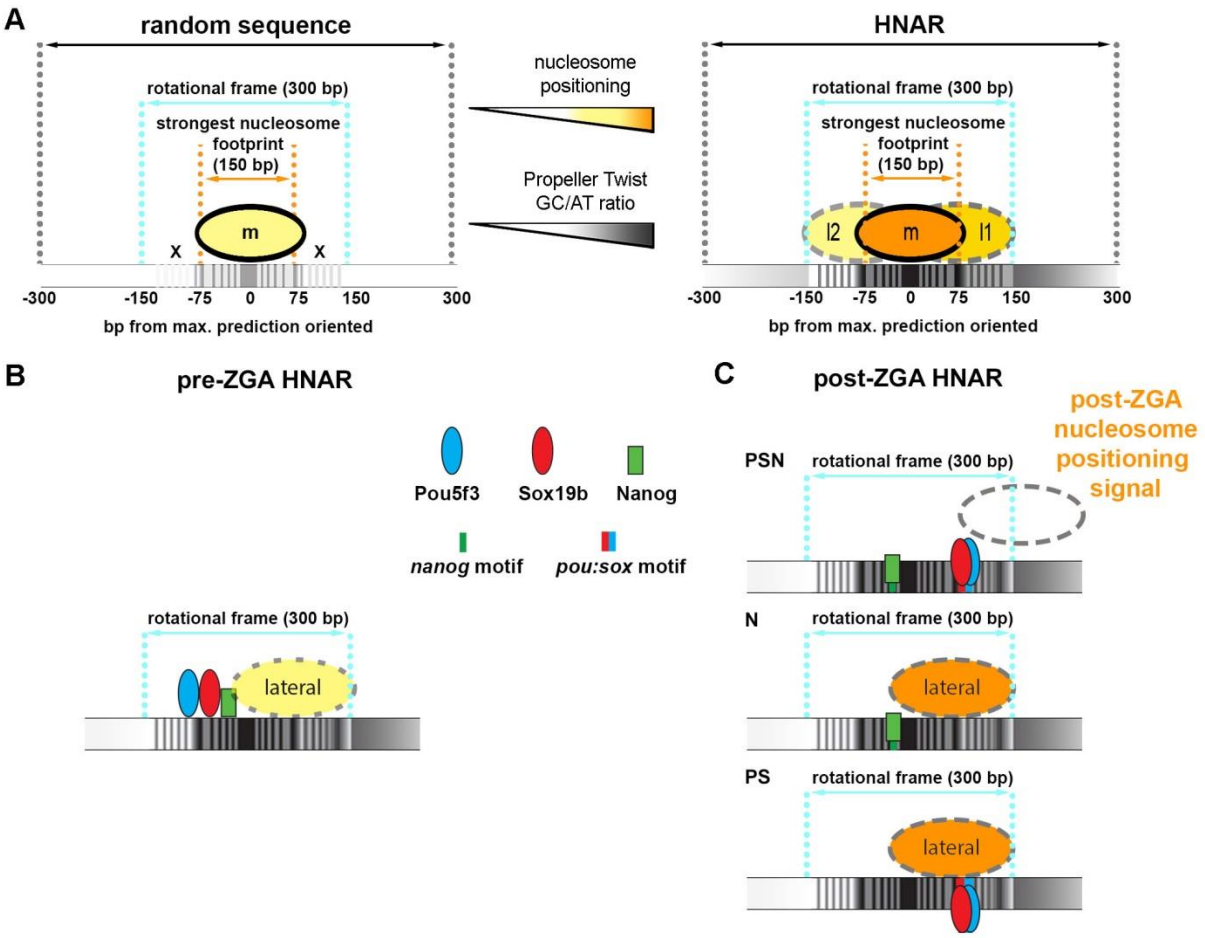

Fig. 6. Hypothetical scheme of pre- and post-ZGA Pou5f3, SoxB1 and Nanog TF-dependent nucleosome shifts and clearance of HNARs. (*A*) Schematic drawing of random nucleosome footprint (left) and HNAR (right). In HNAR, central 300 bp periodic frame supports alternative medial, lateral 1 and lateral 2 nucleosome positions with different strengths (m>l1>l2); in random footprint only medial position is supported. (*B*) pre-ZGA HNAR: Pou5f3, Sox19b and Nanog non-specifically compete with histones on the strongest nucleosome footprint; nucleosomes are shifted to the lateral positions of the rotational frame. (*C*) Post-ZGA, stabilization of nucleosomes by unknown nucleosome positioning signal. Nanog binding to the central footprint and binding Pou5f3/SoxB1 complex to the flanks maintains the open state of 300 bp periodic frame. Nanog binding alone shades nucleosome to lateral positions of the periodic frame; Pou5f3/SoxB1 binding has no effect.

Thus, 300 bp periodic oscillations in our PT summary plots matched the maximal extent of rotational periodicity documented *in vivo* (Gaffney et al. 2012, Fig.1C therein). The random

nucleosome footprint and HNAR, both oriented from low to high nucleosome prediction/PT, are shown in Fig. 6A: a central nucleosome positioning sequence of 150 bp is embedded into the 300 bp rotational periodicity region in both. PT values/relative GC content is higher in HNAR, which results in stronger nucleosome positioning. The graph above HNAR shows three rotational positions within the 300 bp frame, given that high PT and periodic PT signal represent two independent factors which stabilize nucleosome on DNA. The central (medial), lateral 1 and lateral 2 positions are supported by 10 periodic peaks each. The medial position is the most stable, the l1 (lateral) position is weaker, because of the overall lower PT, the l2 position is the weakest supported by even lower PT value compared to m and l1. In the random nucleosome footprint, only medial position is stabilized by underlying DNA features. The model on Fig. 6A is supported by *in-vivo* distribution of nucleosomes (data presented on Fig. 5A-C, E-G, and Fig. S17-S19).

Initial clearance of strong nucleosome footprints pre-ZGA by nonconsensus TF binding.

At the 9th cell cycle, before ZGA commences, Pou5f3 and Nanog preferentially bind to HNARs closed by nucleosomes, where they compete with histones and reduce nucleosome density. We propose that pre-ZGA competition of Pou5f3 and Nanog TFs with histones is driven by their ability to recognize nonconsensus cues of strong nucleosome positioning sequences. Both transcription factors compete with histones for the strongest nucleosome positioning signal; this results in shading of nucleosomes to the lateral position within the same rotational periodicity frame (Fig. 6B). This is in agreement with the results of Joseph et al. (2017), who demonstrated that at ZGA, excess of histones reduces genomic binding of transcription factors and *vice versa*. Most of the Pou5f3/SoxB1 earliest direct targets are tissue-specifically expressed (Onichtchouk

et al. 2010; Leichsenring et al. 2013), and regional transcriptional cofactors were identified in some cases. For example, early expression of *klf-17* in the superficial extraembryonic layer of the embryo (EVL) is directly regulated by ubiquitous Pou5f3 and maternal EVL-restricted IRF-6 (Kotkamp et al. 2014; Liu et al. 2016). We assume that pre-ZGA non-specific competition of Pou5f3 and Nanog with histones is enough to reduce nucleosome occupancy to the level, which allows additional TFs to bind and remove the nucleosome, acting cooperatively with ZGA TFs.

Establishment of open chromatin regions after ZGA requires consensus binding and synergistic action of Pou5f3/SoxB1 and Nanog within HNAR.

We detected widespread alignment of nucleosomes on their DNA-encoded nucleosome positioning signals from pre- to post-ZGA, in all groups except PSN (Fig. 5, Fig. S17). Increase of nucleosome positioning strength was previously documented for promoters: zygotic TSS selection grammar is characterized by nucleosome positioning signals, precisely aligned with the first downstream (+1) nucleosome (Haberle et al. 2014). Strengthening of nucleosome attachment to their DNA sites over ZGA is independent of transcription (Haberle et al. 2014, Zhang et al., 2014), but may depend on the elongation of the cell cycle, or on the rapid exchange of maternal nucleosome-associated proteins, such as of conserved maternal linker histones B4/H1M to their zygotic variants (Muller et al. 2002). We speculate that after an initial short phase during 9-10th cell cycle when histones may be non-specifically outcompeted with TFs and *vice versa*, most temporally “open” chromatin sites are closed again, when not maintained by PSN post-ZGA binding, early lineage-specific factors or both. Post-ZGA, Pou5f3/SoxB1 and Nanog maintain nucleosomes from invading the centers of HNARs, when all three factors are present within one HNAR (Fig. 6C).

The simplest explanation for the synergistic chromatin remodeling by Pou5f3, SoxB1 and Nanog, would be co-binding of all three TFs to the DNA, so that Nanog binding is facilitated by the presence of Pou5f3/SoxB1 complex nearby, or the other way around. We find this explanation unlikely, because of two reasons. First, we did not find significant differences in Nanog binding between N and PSN groups, nor in Pou5f3 or SoxB1 ChIP-seq peaks intensities between PS and PSN groups (Fig. 1A, Fig. S1 and data not shown). Second, the densities of *pou:sox* and *nanog* motifs in PSN regions did not exceed those in N or PS regions (Fig. S18-S19, Table S1), indicating that Pou5f3/SoxB1 and Nanog bind independently to their motifs. Post-ZGA synergistic nucleosome clearance by Pou5f3/SoxB1 and Nanog on PSN regions is unlikely due to nucleosome-mediated cooperativity within one nucleosome length, at least judged by the test we used: the effects of Pou5f3 on the nucleosome density did not increase with the predicted nucleosome strength (Fig. 5H). Post-ZGA, the combination of Nanog binding within the 150 bp central footprint of HNAR, and Pou5f3/SoxB1 binding within the 300 bp rotational periodicity frame, presumably to the regions where the linker histone binds, reverses real nucleosome occupancy against predicted (Fig. 6C). This provides a relatively open platform in chromatin, where time- and lineage-specific TFs expressed during gastrulation can bind and switch on transcription of their target genes.

It has been suggested that strong nucleosome positioning signals overlap with the enhancers (Tillo et al. 2010) and pioneer TFs may recognize these signals (Barozzi et al. 2014; Sun et al. 2015). However, it is currently unclear how ubiquitously occurring strong nucleosome positioning signals differ from those contributing to the enhancers. We propose that TF-specific DNA shape cues may be decoded as differential binding positions of the TFs within the HNAR

relatively to its center. 300 bp regions of ~10 bp rotational periodicity, remarkably similar in shape, can be recovered from zebrafish, yeast, and human genomes. This suggests that our model of differential binding to HNARs will be instrumental for future studies of genome regulatory biology of eukaryotes.

5

Materials and Methods

Zebrafish maintenance and embryo collection

Wild-type fish of AB/TL strain were raised, maintained and crossed under standard conditions as described by Westerfield (Westerfield 2000). The mutant *spg*^{m793} and *nanog*^{m1435} lines were maintained as described previously (Lunde et al. 2004; Veil et al. 2018). Embryos obtained from crosses were collected within 10 minutes and raised in egg water at 28.5°C. Staging was performed following the 1995 Kimmel staging series (Kimmel et al. 1995). Stages of MZ*spg*^{m793} and MZ*nanog*^{m1435} 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).

MNase digestion and sequencing

Around 200-400 wild-type, MZ*spg* and MZ*nanog* embryos were staged to 512-cell (2.75 hours post-fertilization) or dome (4.3 hours post-fertilization) stage. Embryo fixation and the MNase digestion was performed as described (Zhang et al. 2014). 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% mononucleosomes (Supplemental Fig. S4). Libraries were prepared using the Illumina sequencing library preparation protocol and single-end sequenced on an Illumina HiSeq 2500 by Eurofins Company (Germany).

Mapping, data processing details and statistical analysis are in the Supplementary Material.

References and Notes:

Acknowledgments: We are grateful to Nadine Vastenhouw for sharing MNase-seq protocol, to Wolfgang Driever, Patrick Lemaire, Meijiang Gao and Marco Ell for commenting the earlier versions of the manuscript, to Sabine Goetter for excellent fish care and to Andrea Buderer for administrative support;

Funding: This work was supported by DFG-ON86/4-1 and DFG-EXC294 A7-2 for D.O. The Freiburg Galaxy Team is funded by DFG grant SFB 992/1 2012 and BMBF grant 031 A538A RBC;

Author contributions: D.O. conceived the project and wrote the manuscript, M.V. performed all the experiments, D.O, L.Y and B.G performed the data analysis. All authors edited the manuscript;

Competing interests: Authors declare no competing interests;

Data and materials availability:

MNase-seq data are deposited in GEO NCBI under accession number GSE109410. All processed data is available in the main text or the supplementary materials.

Supplementary Materials:

Supplementary Methods

Figures S1-S21

Tables S1-S9

Supplementary References

References

Adams CC, Workman JL. 1993. Nucleosome displacement in transcription. *Cell* **72**: 305-308.

Afek A, Cohen H, Barber-Zucker S, Gordân R, Lukatsky DB. 2015. Nonconsensus Protein Binding to Repetitive DNA Sequence Elements Significantly Affects Eukaryotic Genomes. *PLoS Comput Biol* **11(8)**: e1004429

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Eberhard C et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3-W10.

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. *Nature* **446**: 572-576.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: W202-208.

Barozzi I, Simonatto M, Bonifacio S, Yang L, Rohs R, Ghisletti S, Natoli G. 2014. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Molecular cell* **54**: 844-857.

Beato M, Eisfeld K. 1997. Transcription factor access to chromatin. *Nucleic Acids Res* **25**: 3559-3563.

Belting HG, Wendik B, Lunde K, Leichsenring M, Mossner R, Driever W, Onichtchouk D. 2011. Pou5f1 contributes to dorsoventral patterning by positive regulation of vox and modulation of fgf8a expression. *Dev biol* **356(2):323-36**: 323-336.

Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825-837.

Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E et al. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**: 1274-1285.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106-1117.

Chung HR, Vingron M. 2009. Sequence-dependent nucleosome positioning. *J Mol Biol* **386**: 1411-1422.

Daenen F, van Roy F, De Bleser PJ. 2008. Low nucleosome occupancy is encoded around functional human transcription factor binding sites. *BMC Genomics* **9**: 332.

el Hassan MA, Calladine CR. 1996. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J Mol Biol* **259**: 95-103.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology* **4**: e1000216.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genetics* **8**: e1003036.

Gan Y, Guan J, Zhou S, Zhang W. 2012. Structural features based genome-wide characterization and prediction of nucleosome organization. *BMC Bioinformatics* **13**: 49.

Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, Yi L, Yu G, Hou Z, Fan D et al. 2018. Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* **173**: 248-259 e215.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biol* **8**: R24.

Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, Gehrig J, Dong X, Akalin A, Suzuki AM et al. 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**: 381-385.

Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. 2011. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genetics* **7**: e1002266.

He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M et al. 2010. Nucleosome dynamics define transcriptional enhancers. *Nature genetics* **42**: 343-347.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**: 576-589.

Iwafuchi-Doi M, Zaret KS. 2016. Cell fate control by pioneer transcription factors. *Development* **143**: 1833-1837.

Joseph SR, Palfy M, Hilbert L, Kumar M, Karschau J, Zaburdaev V, Shevchenko A, Vastenhouw NL. 2017. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife* **6**.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362-366.

Khoueiry P, Rothbacher U, Ohtsuka Y, Daian F, Frangulian E, Roure A, Dubchak I, Lemaire P. 2010. A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites. *Current biology : CB* **20**: 792-802.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253-310.

Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**: 868-871.

Kotkamp K, Mossner R, Allen A, Onichtchouk D, Driever W. 2014. 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.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235-1244.

Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D. 2013. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**: 1005-1009.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147-151.

Li R, Cauchy P, Ramamoorthy S, Boller S, Chavez L, Grosschedl R. 2018. Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. *Genes Dev* **32**: 96-111.

Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. 2014. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *eLife* **3**.

Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**: 400-403.

Liu H, Leslie EJ, Jia Z, Smith T, Eshete M, Butali A, Dunnwald M, Murray J, Cornell RA. 2016. Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. *Human molecular genetics* **25**: 766-776.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**: 431-440.

Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y. 2016. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**: 1375-1388.

Lunde K, Belting HG, Driever W. 2004. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. *Current biology : CB* **14**: 48-55.

Luo Y, North JA, Rose SD, Poirier MG. 2014. Nucleosomes accelerate transcription factor dissociation. *Nucleic Acids Res* **42**: 3017-3027.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**: 495-501.

Mirny LA. 2010. 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.

Moyle-Heyrman G, Tims HS, Widom J. 2011. Structural constraints in collaborative competition of transcription factors against the nucleosome. *Journal of molecular biology* **412**: 634-646.

Muller K, Thisse C, Thisse B, Raz E. 2002. Expression of a linker histone-like gene in the primordial germ cells in zebrafish. *Mech Dev* **117**: 253-257.

Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C. 2011. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genetics* **7**: e1002339.

Onichtchouk D, Geier F, Polok B, Messerschmidt DM, Mossner R, Wendik B, Song S, Taylor V, Timmer J, Driever W. 2010. Zebrafish Pou5f1-dependent transcriptional networks in temporal control of early development. *Mol Syst Biol* **6**: 354.

Papatsenko D, Goltsev Y, Levine M. 2009. Organization of developmental enhancers in the Drosophila embryo. *Nucleic Acids Res* **37**: 5665-5677.

Parvin MS, Okuyama N, Inoue F, Islam ME, Kawakami A, Takeda H, Yamasu K. 2008. Autoregulatory loop and retinoic acid repression regulate pou2/pou5f1 gene expression in the zebrafish embryonic brain. *Dev Dyn* **237**: 1373-1388.

Remenyi A, Lins K, Nissen LJ, Reinbold R, Scholer HR, Wilmanns M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17**: 2048-2059.

Salmon-Divon M, Dvinge H, Tammoja K, Bertone P. 2010. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**: 415.

Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM. 2015. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**: 1715-1726.

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772-778.

Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordan R, Rohs R. 2014. Absence of a simple code: how transcription factors read the genome. *Trends in biochemical sciences* **39**: 381-399.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**: 555-568.

Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, Rushlow C. 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res* doi:10.1101/gr.192542.115.

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. *Development* **136**: 3033-3042.

Teif VB, Vainshtein Y, Caudron-Herger M, Mallm JP, Marth C, Hofer T, Rippe K. 2012. Genome-wide nucleosome positioning during embryonic stem cell development. *Nature structural & molecular biology* doi:10.1038/nsmb.2419.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75-82.

Tillo D, Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* **10**: 442.

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. 2010. High nucleosome occupancy is encoded at human regulatory sequences. *PLoS One* **5**: e9129.

Trompouki E, Bowman TV, Lawton LN, Fan ZP, Wu DC, DiBiase A, Martin CS, Cech JN, Sessa AK, Leblanc JL et al. 2011. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**: 577-589.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**: 516-520.

Veil M, Schaechtle MA, Gao M, Kirner V, Buryanova L, Grethen R, Onichtchouk D. 2018. Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. *Development* **145**.

Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. 2016. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* **167**: 1555-1570 e1515.

Westerfield M. 2000. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio).* Univ. of Oregon Press, Eugene.

Xu C, Fan ZP, Muller P, Fogley R, Dibiase A, Trompouki E, Unternaehrer J, Xiong F, Torregroza I, Evans T et al. 2012. Nanog-like Regulates Endoderm Formation through the Mxtx2-Nodal Pathway. *Developmental cell* **22**: 625-638.

Yanez-Cuna JO, Arnold CD, Stampfel G, Boryn LM, Gerlach D, Rath M, Stark A. 2014. Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res* **24**: 1147-1156.

Yang L, Zhou T, Dror I, Mathelier A, Wasserman WW, Gordan R, Rohs R. 2014. TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res* **42**: D148-155.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Research* **39**: e35.

Zhang Y, Shin H, Song JS, Lei Y, Liu XS. 2008. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. *BMC Genomics* **9**: 537.

Zhang Y, Vastenhouw NL, Feng J, Fu K, Wang C, Ge Y, Pauli A, van Hummelen P, Schier AF, Liu XS. 2014. Canonical nucleosome organization at promoters forms during genome activation. *Genome Res* **24**: 260-266.

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65-70.