Supplementary Materials for

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

Marina Veil[1], Lev Yampolsky[2,3], Björn Grüning[4,5], Daria Onichtchouk[1,6*].
Correspondence to: daria.onichtchouk@biologie.uni-freiburg.de

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S3 to S9
Captions for Data Tables S1 and S2

Other Supplementary Materials for this manuscript include the following:

Data Tables S1 and S2

Supplementary Methods

Library preparation and sequencing

Libraries were prepared using the Illumina sequencing library preparation protocol and single-end sequenced on an Illumina HiSeq 2500 by Eurofins company (Germany).

Mapping

All sequenced reads were mapped back to the zebrafish Zv9 assembly using CLC bio workbench (Qiagen) with default parameters, sequences which matched more than one position were mapped randomly. Numbers of mapped reads are listed in Table S3. To create the files for scoring and visualization of the nucleosome profiles from single end sequencing data, we used the strategy previously described (Zhang et al. 2008). Original BAM mapping files from CLC workbench were imported to Galaxy server (Afgan et al. 2016) in Freiburg (https://galaxy.uni-freiburg.de), and converted to BED format using BEDTools (Quinlan and Hall 2010), wrapped into the Galaxy by Bjoern A. Gruening (2014) Galaxy wrapper (https://github.com/bgruening/galaxytools). Then all mapped reads were extended to 147 bp in their 3' direction and truncated to the middle 61 bp. Both original mapping BAM files and processed BED files are available under GEO accession number GSE109410.

Selection of TFBS and control groups for analysis

Lists of 6,670 post-ZGA Pou5f3, 7,747 pre-ZGA Pou5f3 and 5,924 SoxB1 ChIP-seq peak genomic coordinates were taken from (Leichsenring et al. 2013). Mapped ChIP-seq reads for Pou5f3 pre- and post-ZGA and SoxB1 post-ZGA were from GSE39780 series in GEO NCBI**.** 14,010 dome stage Nanog ChIP-seq peak coordinates in Zv8 genome assembly were taken from (Xu et al. 2012), and converted into 13,775 Zv9 peaks using LiftOver utility from the UC Santa Cruz Genome Browser. Raw Nanog ChIP-seq data were uploaded from GSE34683 series in GEO NCBI and mapped to Zv9 genome assembly. To create the group of control genomic regions, we first calculated the genomic distances from 7,500 randomly taken Pou5f3, SoxB1 and Nanog peaks to the transcription start site (TSS) of the closest ENSEMBL transcript and then took the genomic regions at the same distances from randomly picked ENSEMBL transcripts. Previous analysis of the zebrafish promoters demonstrated the genome-wide formation of prominent nucleosome-free regions upstream of the promoters at ZGA (Haberle et al. 2014; Zhang et al. 2014). To exclude the overlap with these nucleosome-free regions we removed the peaks +/-1 kb from the annotated promoters of ENSEMBL transcripts. ChIP-seq and MNAse-seq data sets were uploaded to seqMINER (Ye et al. 2011), which allows to simultaneously visualize and score nucleosome and ChIP-seq signal distribution around the set of genomic regions. To reduce the heterogeneity of the analyzed ChIP-seq lists, we processed them in a standard way described below. First, we removed the ChIP-seq peaks, which were falling to the regions of very high MNase-seq occupancy (top 1% of the reads/bp in each MNase sample within any 20 bp in +/- 1 kb from the peak center), with the purpose to exclude the heterochromatin and repetitive regions. To assign the single (P, S, N), double (PS, SN, PN) or triple (PSN) occupancy to the post-ZGA TFBS, we used ChIP-seq BAM files to calculate TF occupancy for P, S and N separately in the central 320 bp (mean peak width) and average from two 320 bp flanks (background). For each genomic region in TFBS list, the region was scored as negative for TF binding, if peak/background signal ratio was less or equal to the arbitrary cutoff 1.3, and positive, if the peak/background ratio was more than 1.3. After the filtering, one quarter of post-ZGA TF peaks (6,139 regions) were removed from the original list of 26,369 regions. To the remaining list, 6,248 Pou5f3, 3,301 Sox2 and 2,437 Nanog binding events were assigned as positive in combinations. We did not assign SoxB1-only group, as SoxB1 binding overlapped with some levels of Pou5f3

or Nanog on all regions, due to deliberately low cutoff 1.3 we used. Nanog only bound and Pou5f3 only bound groups (N and P) were significantly large in spite of the low cutoff 1.3. To assign Ppre (pre-ZGA Pou5f3 – only) group, overlaps between Pou5f3 pre- with any of post-ZGA TFBS were removed, leaving 5170 genomic regions. Genomic coordinates of the resulting 8 groups of genomic regions (PSN, PS, PN, SN, P, N, Ppre and control) are listed in the Table S1 and were used for further analysis.

Motif finding and analysis

To find all the motifs specifically enriched in our data, we used MEME suite at http://meme-suite.org (Bailey et al. 2009). We sorted the TFBS by descending ChIP-seq signal for each TF separately. Top 1,000 60 bp wide sequences and top 270 220 bp wide sequences for post-ZGA Pou5f3, SoxB1 and Nanog, and for Pou5f3 pre-ZGA were used as an input for MEME for de-novo motif finding program (Bailey et al. 2009) with parameters:

Motif Site Distribution ZOOPS: Zero or one site per sequence
Maximum Number of Motifs 10
Minimum Motif Width 6
Maximum Motif Width 25

All the derived motifs were combined in one list. Motifs which were too similar to the others were removed using MAST (Gupta et al. 2007) with matrix correlation threshold 0.9. The remaining 14 motifs (PWMs listed in Table S2) were compared to JASPAR database using Tomtom (Gupta et al. 2007), to the known Nanog, Pou5f1/Pou5f3, Pou5f1/SoxB1 and SoxB1 motifs (Remenyi et al. 2003; Loh et al. 2006; Chen et al. 2008; Salmon-Divon et al. 2010; Xu et al. 2012; Leichsenring et al. 2013) and enhancer-specific motifs (Yanez-Cuna et al. 2014) from the literature and subdivided into 5 groups: *pou:sox* (3 motifs), *nanog1* (1 motif), *nanog2* (5 motifs), *sox* (3 motifs), enhancer-associated dinucleotide repeats (*gaga* repeat and *tgtg* repeat) and *atgg* repeat (Fig. S2). Genomic coordinates of the individual occurrences of each motif in +/- 1.5 kb from the center of all TFBS regions were obtained using FIMO (Grant et al. 2011) with p-value threshold 10^{-4}. The genomic coordinates of motifs were saved as a BED file and used for plotting the heat maps in seqMINER (Ye et al. 2011, Fig. 1B, Fig. S2), and for motif-centered nucleosome occupancy and nucleosome prediction plots in Galaxy (Afgan et al. 2016). BED files converted to bedgraph format with BEDTools (Quinlan and Hall 2010) were used for drawing motif occupancy plots. To score *pou:sox* and *nanog1&2* motifs per genomic region, overlapping motifs within *pou:sox* group and *nanog1* and *nanog2* groups, respectively, were merged together and trimmed to 20 bp. The numbers of non-redundant matches to *pou:sox* and *nanog* per each 320 bp long TFBS was scored in Galaxy and listed in Table S1.

Comparative nucleosome occupancy plots

Analysis was done on Galaxy instance in Freiburg (now Galaxy europe). 6 BED files with MNase-seq data for two stages and three genotypes were converted to Bedgraph format using BEDTools. Regions with top and bottom 1% genomic coverage values were removed from the analysis. For practical reasons of memory usage, only the regions matching +/- 5 kb from the TFBS and control groups were kept. Mean coverage in reads/bp in the genomic regions of interest was calculated in 1 bp or 10 bp steps (for 900 bp wide plots) or in 20 bp steps (for 3,000 bp wide plots), using custom Galaxy workflows available on request. Each value in the results table was normalized to RPKM (reads per million reads per one kb), using scaling factors calculated as 10^9/ total mapped reads (Table S3). The background values (average value for control group, Table S3) were subtracted from each value in the result table. The plots were smoothened with 20 bp or 80 bp moving average or un-smoothened, as indicated on the figures. ChIP-seq plots on (Fig. S1) were plotted in a similar way, using

ChIP-seq bedgraph coverage files, but without RPKM normalization or background subtraction.

Nucleosome predictions

Nucleosome prediction program from Kaplan et al.(Kaplan et al. 2009) 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 Galaxy ToolShed (ref. https://www.ncbi.nlm.nih.gov/pubmed/25001293) and is available at the Galaxy instance in Freiburg. The sequences around TFBS and controls were extended to 10 kb to account for the edge effects; the nucleosome prediction for each base in the middle 1-3 kb were taken for analysis. The maximal and minimal nucleosome prediction values within 320 bp around TFBS and control regions and their genomic positions are listed in Table S1. Using smaller windows around TFBS (45, 75 and 200 bp) for max. search did not change the 300 bp periodic shape around [nucmax], subsequently revealed by PT. To orient the genomic regions aligned on [nucmax] along ascending nucleosome prediction values, we searched for the min. nucleosome prediction at +/- 160 bp around [nucmax]. If the min. prediction was downstream of [nucmax], we reversed the strand from + to -. The strand for oriented plots is listed in Table S1.

Propeller Twist shape

Propeller twist values for individual sequences or aligned groups of sequences were calculated on TFBS shape server at http://rohslab.cmb.usc.edu/TFBSshape/(Yang et al. 2014)

Data normalization and statistical analysis

The sequencing coverage between the samples was normalized as RKPM (reads per million reads per one kilobase). Scaling factors were calculated as 10^9/ total mapped reads and are listed in Table S3. Normalized difference between the mutant and wild-type (Δmut) were calculated as Δmut=((rpkm(mut)-rpkm(wt))/(rkpm(mut)+rpkm (wt)); normalized difference between the stages (ΔWTpost-pre) was calculated as ΔWTpost-pre=(rpkm(WTpost)-rpkm(WTpre))/(rkpm(WTpost)+rpkm(WTpre)). Average RPKM values per 320 bp were taken. Data were analysed using JMP (SAS Institute 2012 version 10) using one-way ANOVA with transcription factor binding group as the factor, followed by Tukey-Kramer test for pair-wise differences with p-value set to 0.01, as indicated in the figure legends and tables.

Supplementary references

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.

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.

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.

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, Dunkel I, Heise F, Linke C, Krobitsch S, Ehrenhofer-Murray AE, Sperling SR, Vingron M. 2010. The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS One* **5**: e15754.
Chung HR, Vingron M. 2009. Sequence-dependent nucleosome positioning. *J Mol Biol* **386**: 1411-1422.
Drew HR, Travers AA. 1985. DNA bending and its relation to nucleosome positioning. *J Mol Biol* **186**: 773-790.
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.
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.
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.
Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253-310.
Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D. 2013. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**: 1005-1009.
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.
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.
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.
Perez-Camps M, Tian J, Chng SC, Sem KP, Sudhaharan T, Teh C, Wachsmuth M, Korzh V, Ahmed S, Reversade B. 2016. Quantitative imaging reveals real-time Pou5f3-Nanog complexes driving dorsoventral mesendoderm patterning in zebrafish. *eLife* **5**.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
Reim G, Brand M. 2006. Maternal control of vertebrate dorsoventral axis formation and epiboly by the POU domain protein Spg/Pou2/Oct4. *Development* **133**: 2757-2770.
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.
Richmond TJ, Davey CA. 2003. The structure of DNA in the nucleosome core. *Nature* **423**: 145-150.
Salmon-Divon M, Dvinge H, Tammoja K, Bertone P. 2010. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**: 415.
Satchwell SC, Drew HR, Travers AA. 1986. Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* **191**: 659-675.
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.

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.

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**.

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.

Supplementary figures

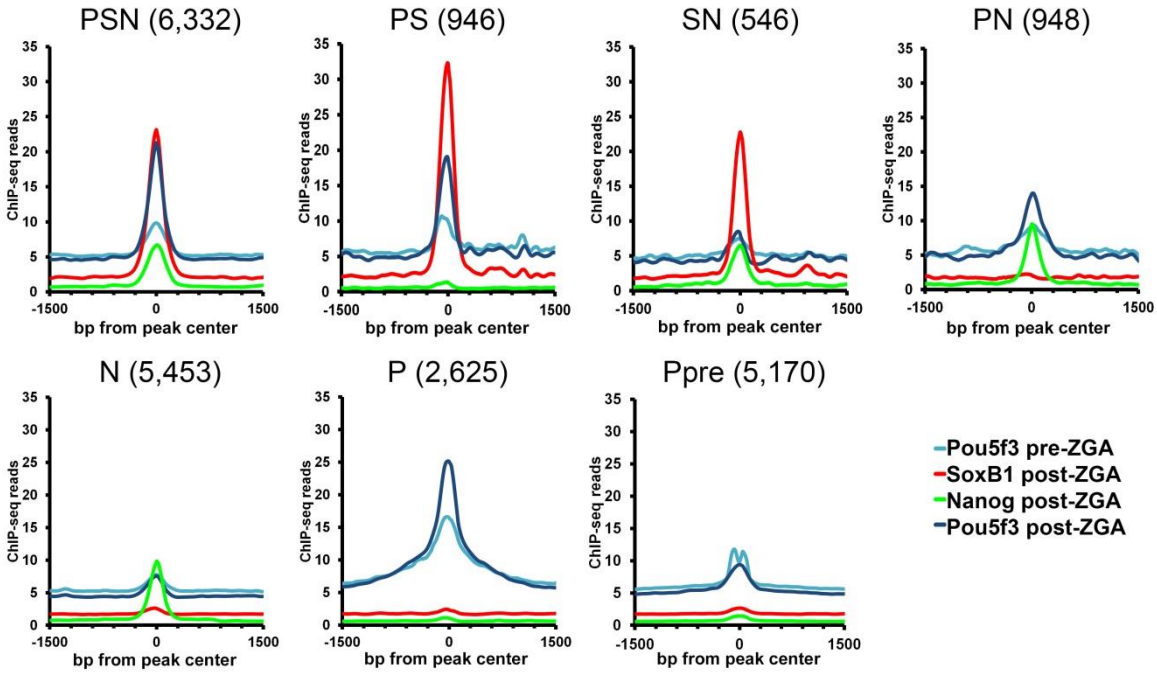

Fig. S1.

Summary TF binding plots for 7 TFBS groups defined in this study. The graphs show summary ChIP-seq signal profiles for each TF in all groups, smoothened with 20 bp moving average.

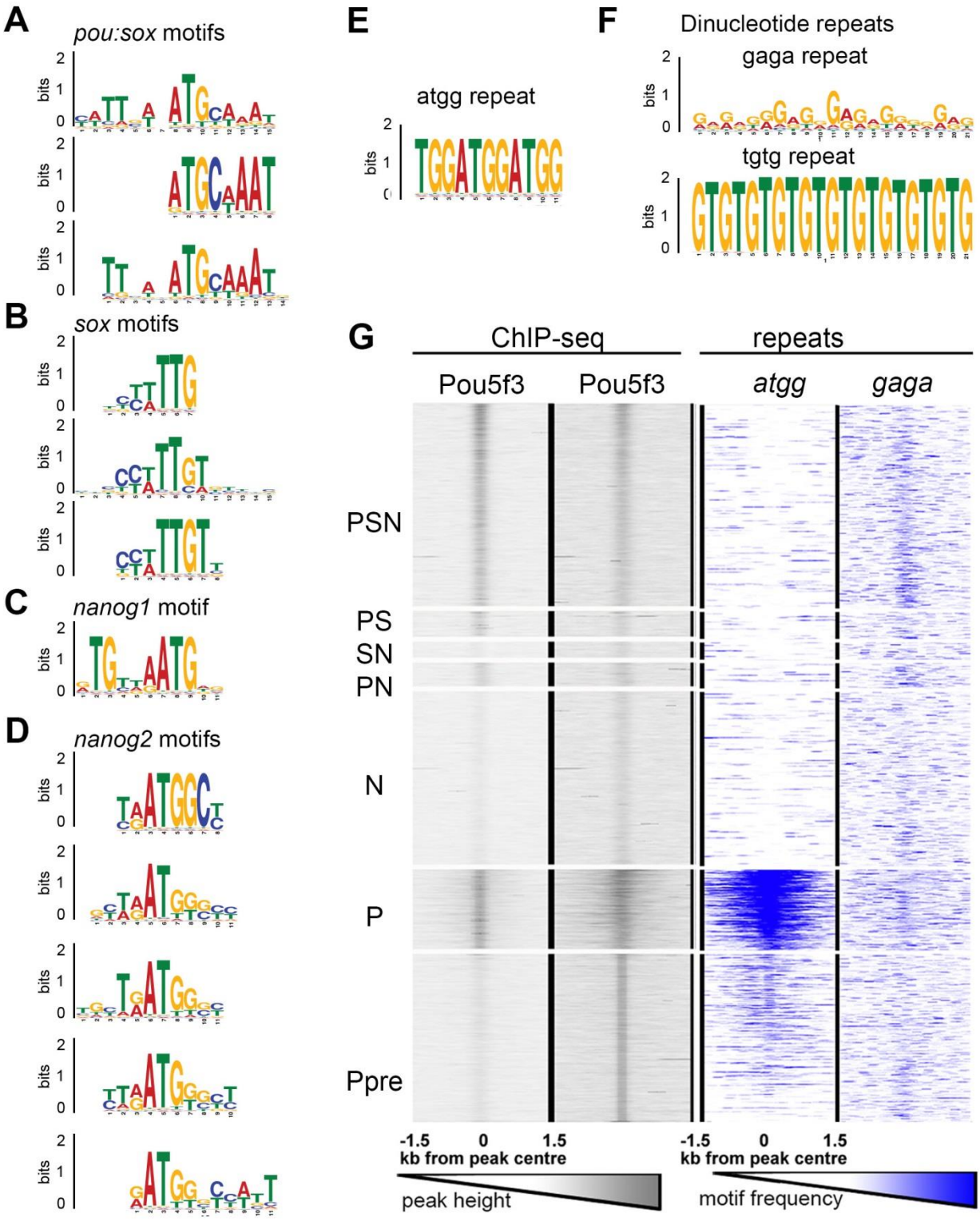

Fig. S2.

Motifs enriched in TFBS from this study. (*A-D*) Sequence logos for specific TF-binding motifs as indicated (see Table S2 for motif composition). (*E*) ATGG repeat. (*F*) Dinucleotide repeats, characterized as a general enhancer feature in Drosophila and human (Yanez-Cuna et al. 2014). (*G*) Occurrence of repetitive sequences in 7 TFBS groups defined in this study. ATGG repeat is enriched in P group of Pou5f3-only post-ZGA binding regions, GAGA repeat is enriched in all groups.

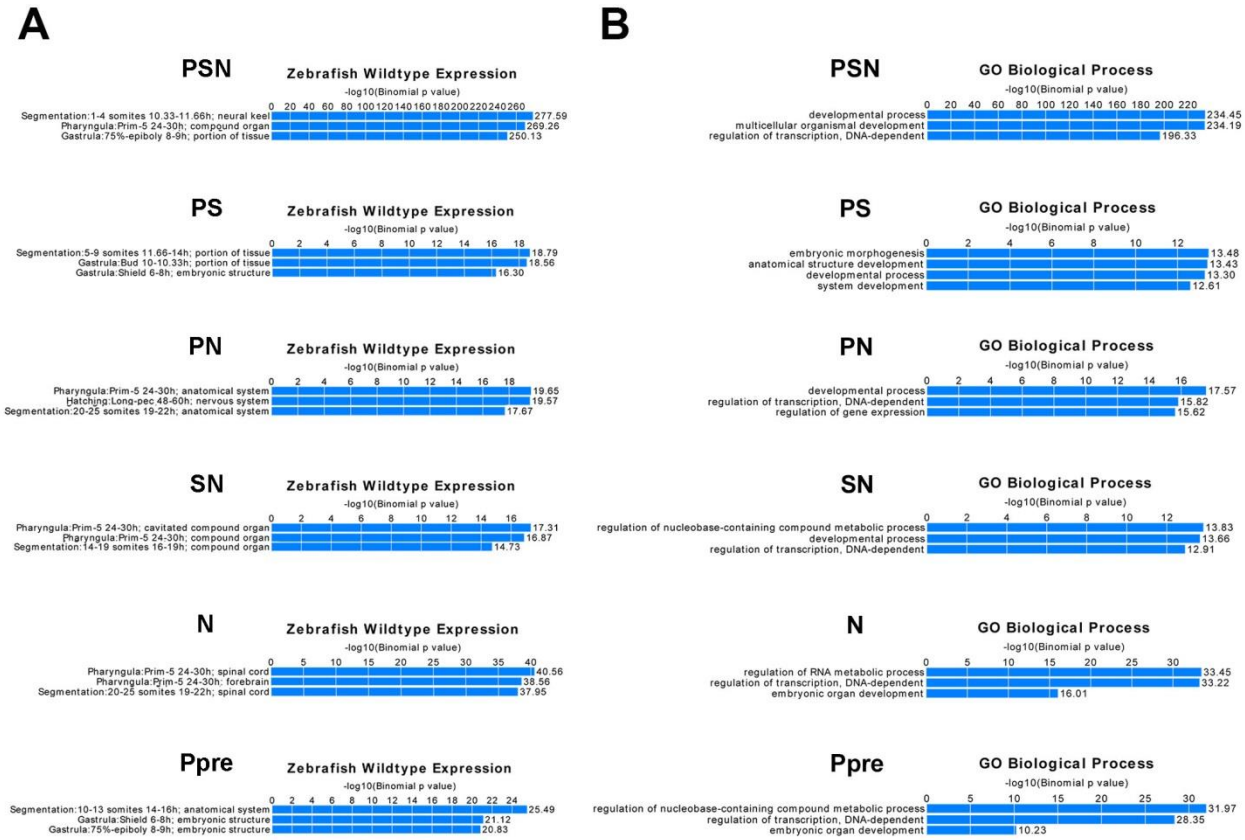

Fig. S3.

6 out of 7 TFBS groups defined in this study are significantly associated with transcriptional regulators and developmental genes. Selected top categories for Zebrafish WT expression (*A*) and Gene Ontology Biological Process (*B*) for each group from GREAT analysis (McLean et al. 2010). Note that enrichment for PSN group (regions bound by Pou5f3, SoxB1 and Nanog post-ZGA) is the order of magnitude higher than others in both categories. P group of Pou5f3-only post-ZGA binding regions did not show significant enrichments.

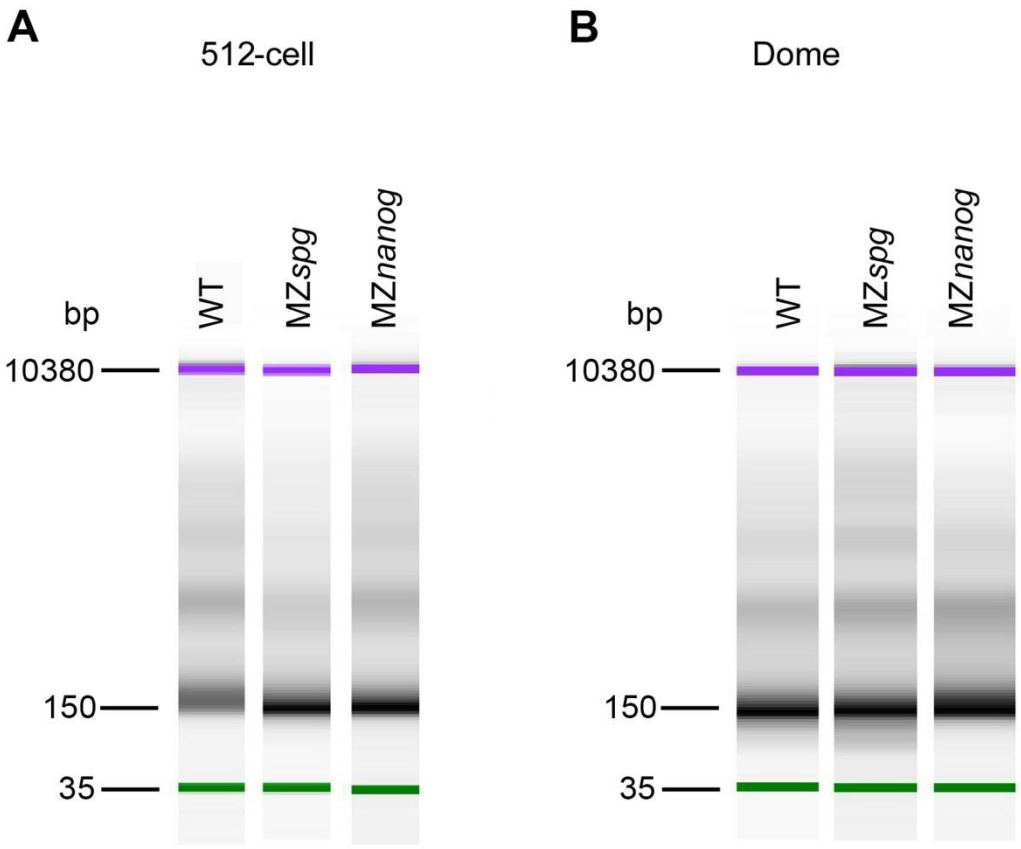

Fig. S4.

Nucleosome fragment length estimation after MNase digestion by using a bioanalyzer (Agilent). Mononucleosomes are displayed at the size of 150 bp for 512-cell stage (pre-ZGA) (*A*) and dome stage (post-ZGA) (*B*).

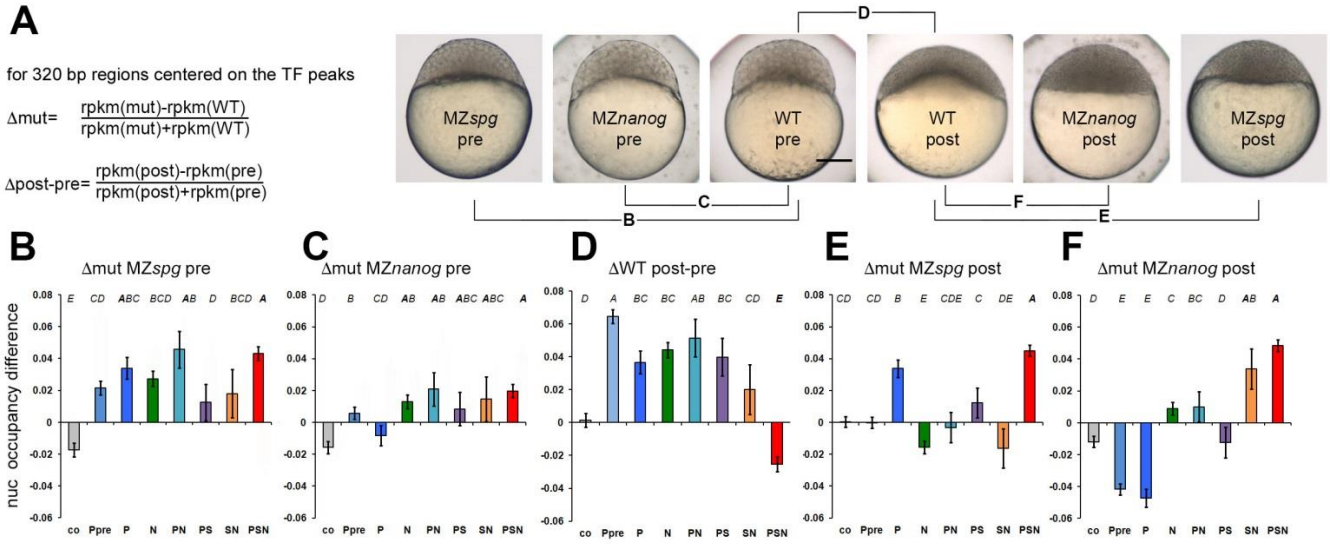

Fig. S5.

Statistics of the nucleosome occupancy change between stages and genotypes. (*A-F*) RPKM values were taken for the 320 bp (mean TF peak width). Error bars are 95% confidence intervals. n(Co) = 6886, n values for the other groups are indicated in (Fig.1 main text). Statistical evaluation was done using 1-way ANOVA (Table S4) and Tukey-Kramer test: groups not sharing a letter on the top of the graph are significantly different (P=0.01).

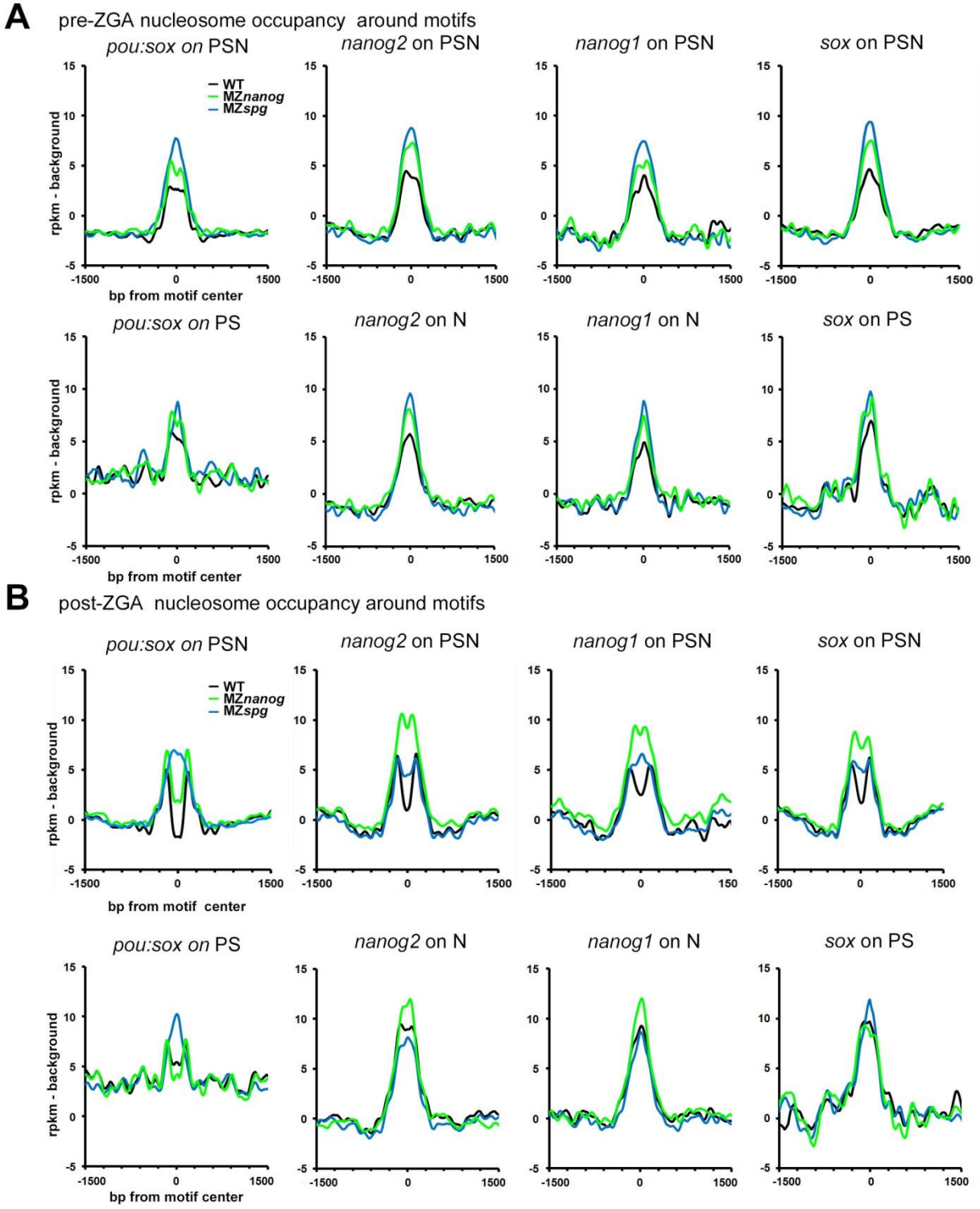

Fig. S6.

Nucleosome occupancy profiles on specific motifs in PSN, N and PS groups. Mean nucleosome occupancy plots in the WT (black), MZ*nanog* (green) and MZ*spg* (blue) centered on *pou:sox*, *nanog1*, *nanog2* or *sox* motifs, in PSN, N and PS groups as indicated. X-axis: bp from the motif center. Y-axis: nucleosome occupancy (RPKM minus background). (*A*) Pre-ZGA. Note, that at pre-ZGA stage, no nucleosome phasing is seen on the motifs, and nucleosome occupancy is the highest in MZ*spg*, lowest in the wild-type and intermediate in MZ*nanog* in all groups. (*B*) post-ZGA. At post-ZGA, nucleosome clearance is seen as a deep

trough on all the motifs in PSN group (first row), nucleosome occupancy is increased in both mutants. No comparable nucleosome clearance is seen in the wild-type PS and N groups on the same motifs (second row).

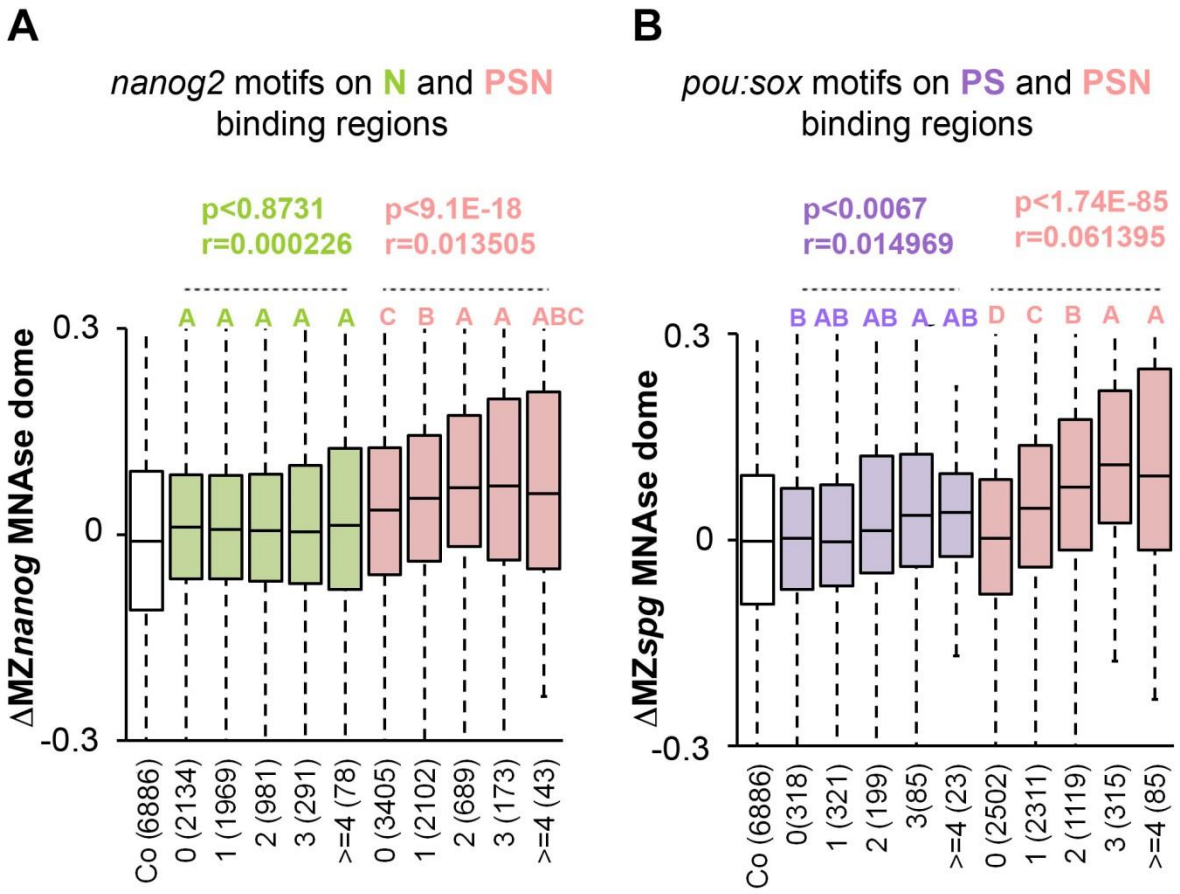

Fig. S7.

Nucleosome clearance depends on specific motifs only in PSN group. Related to Fig. 2J,L of the main text. (*A*) Δmut MZ*nanog* or (*B*) Δmut MZ*spg* was calculated for 320 bp regions of the indicated groups, for the genomic regions overlapping 0, 1, 2, 3 and ≥4 motifs and control (numbers of genomic regions in parentheses). Tukey-Kramer test: groups sharing a letter of the same color on the top of the graph are not significantly different (P=0.01). Note that Δmut increases with the number of both *nanog* and *pou:sox* motifs in the PSN group, weakly increases with the number of *pou:sox* motifs in PS group and does not depend on *nanog* motifs in N group. ANOVA details in Table S5. At pre-ZGA stage Δmut did not depend on the number of motifs in any group or genotype (p-values in 1-way ANOVA > 0.05) except for the decrease of Δmut at PSN regions with the number of *pou:sox* motifs (P<0.003). See Table S2 for the data, and Table S5 for 1-way ANOVA results.

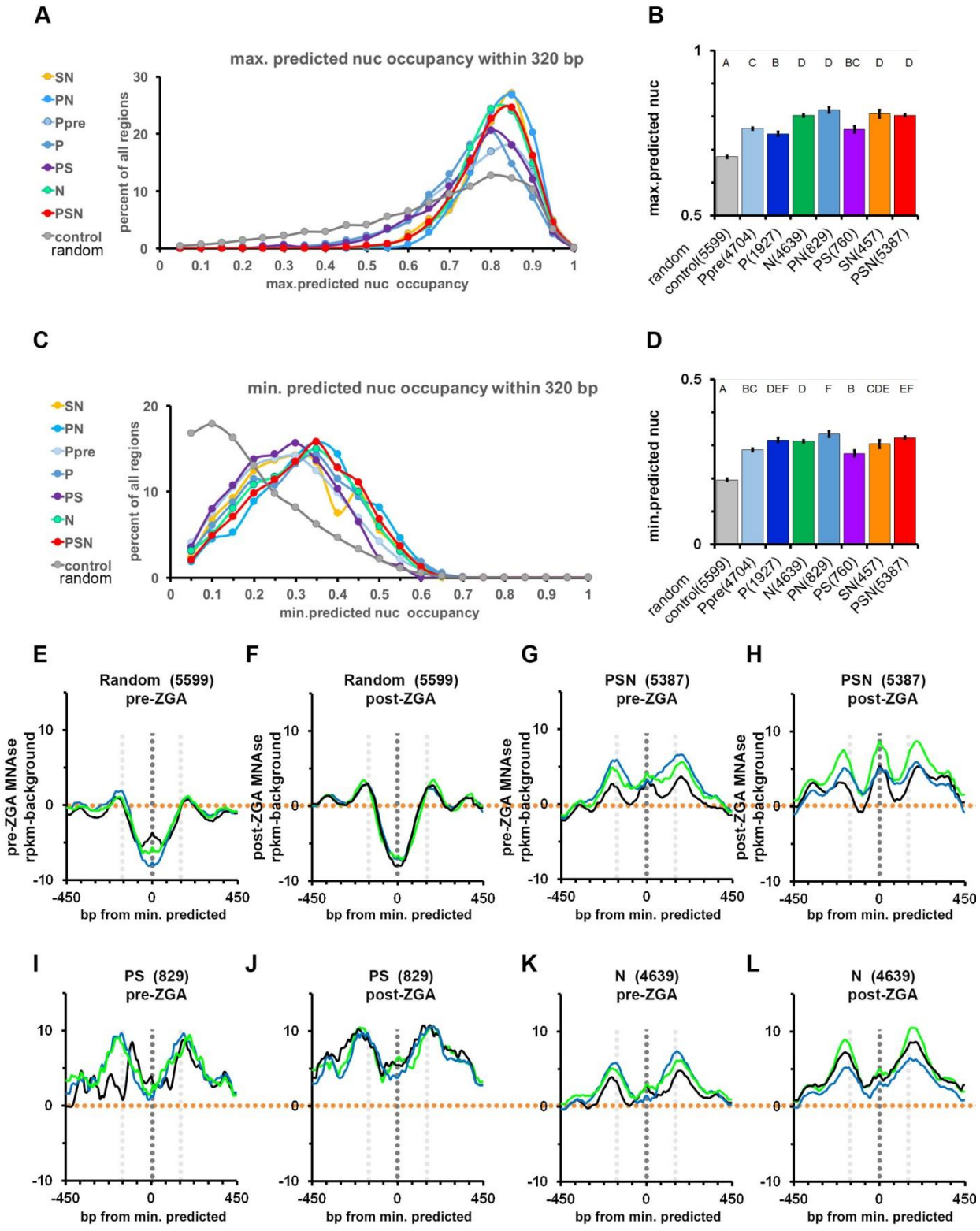

Fig. S8.

Pou5f3, SoxB1 and Nanog bind to the regions of high predicted *in-vitro* nucleosome occupancy which are at least 320 bp long. (*A-D*) Maximum and minimum predicted nucleosome occupancy value [nucmax] and [nucmin] within each 320 bp long TFBS and control regions were calculated using the program of Kaplan et al (Kaplan et al. 2009). (*A*,*C*)

The distribution of [nucmax] (*A*) and [nucmin] (*C*) values for 8 groups: Y-axis shows percent of all regions within a group. (*B,D*) The differences of [nucmax] (*B*) and [nucmin] (*D*) between the groups (see Tables S7 and S8 for the 1-way ANOVA details). Mean values are shown, error bars are 95% confidence intervals. Letters above the graph represent the results of Tukey-Kramer test: categories not significantly different from each other pair-wise ($p \geq 0.01$) share a letter. (*E-L*) Experimental nucleosome occupancy centered on the predicted [nucmin] inter-nucleosomal regions, in the indicated genotypes in control (*E,F*) and indicated TFBS (*G-L*), pre-ZGA or post-ZGA, as indicated. In control, nucleosomes are absent from [nucmin], as predicted; while in PSN, N and PS [nucmin] is occupied by nucleosomes in all genotypes and stages. Orange dotted lines show background level, black dotted lines show 0 and gray lines and +/- 150 bp from [nucmin].

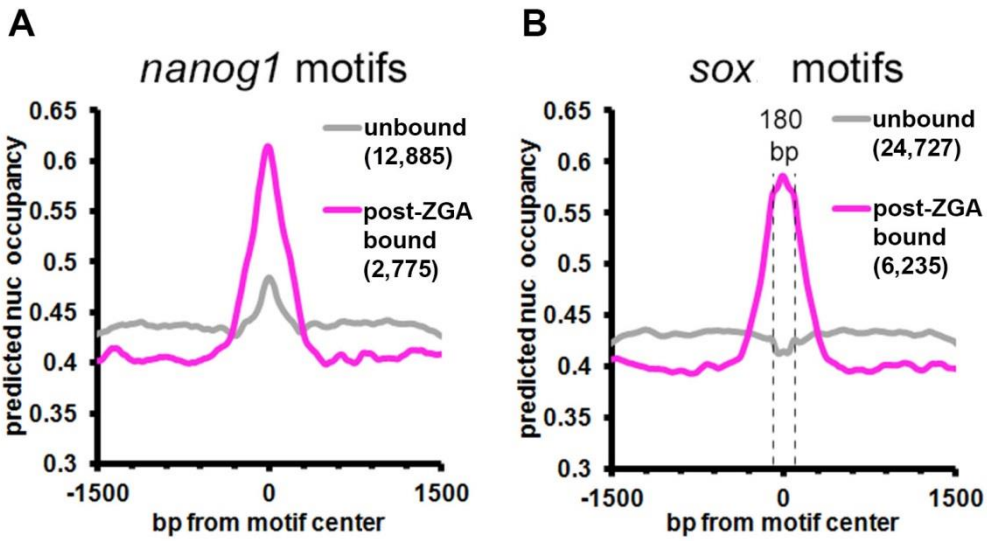

Fig. S9.

Motifs bound and unbound by TFs differ in predicted nucleosome occupancy. Indicated motifs bound (magenta, motifs +/-160 bp from TF binding peak center) versus unbound (gray, motifs +/- 1-1,5 kb from TF binding peak center) by (*A*) Nanog or (*B*) SoxB1. Note that nucleosome prediction peak shapes and max. values are similar between two Nanog motifs *nanog1* (*A*) and *nanog2* (Fig. 3B of the main text), and different from *pou:sox* (Fig. 3C of the main text) motifs. The shape and max. value on *sox* motifs (*B*) is intermediate between *pou:sox* (180 bp shallow region) and *nanog 1&2* (pointed center).

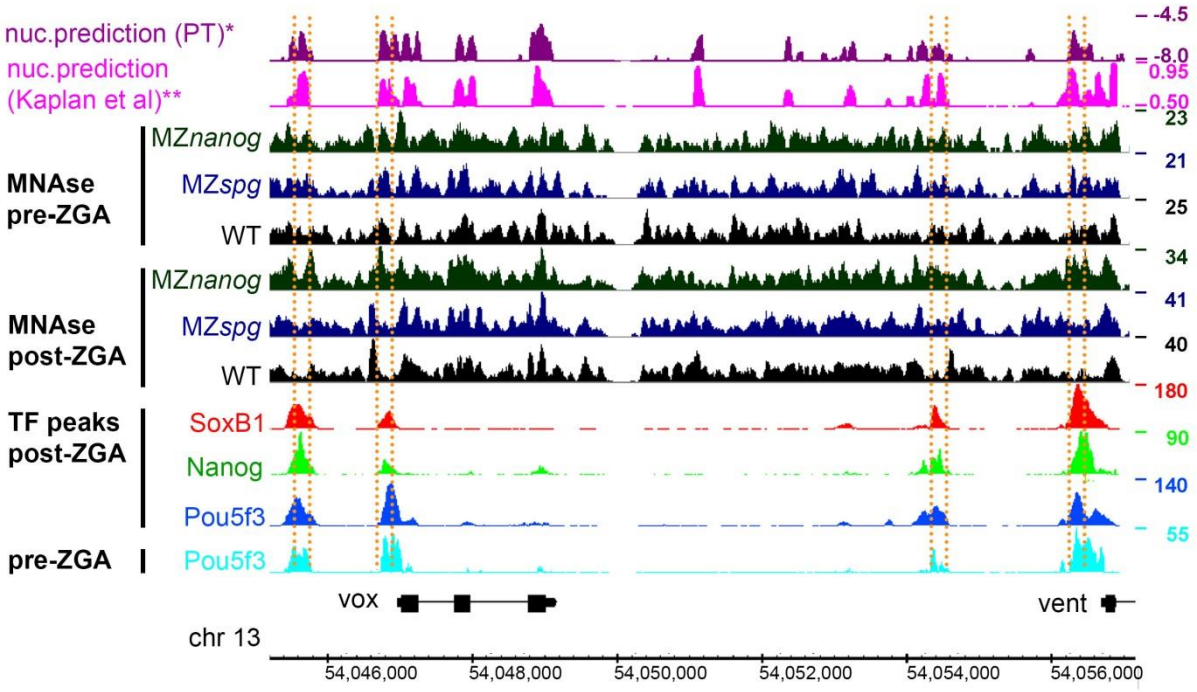

Fig. S10.

Pou5f3, SoxB1 and Nanog bind to the regions of high propeller twist on the *vox* and *vent* enhancers and promoters. 10 kb genomic region with *vox* and *vent* genes (Zv9). *vox* and *vent* are known transcriptional targets of both Pou5f3 (Reim and Brand 2006; Belting et al. 2011) and Nanog (Perez-Camps et al. 2016; Veil et al. 2018). From top to bottom: PT-propeller twist DNA shape (°), values smoothened with 80 bp moving average, nuc –*in-vitro* nucleosome occupancy prediction (Kaplan et al. 2009), 512-cell, dome stage – pre-ZGA and post-ZGA experimental nucleosome occupancy in the WT (black), MZ*nanog* (green) and MZ*spg* (blue), post-ZGA TF peaks, as indicated, pre-ZGA Pou5f3 TF peaks. Orange lines mark enhancers and promoters of *vox* and *vent*.

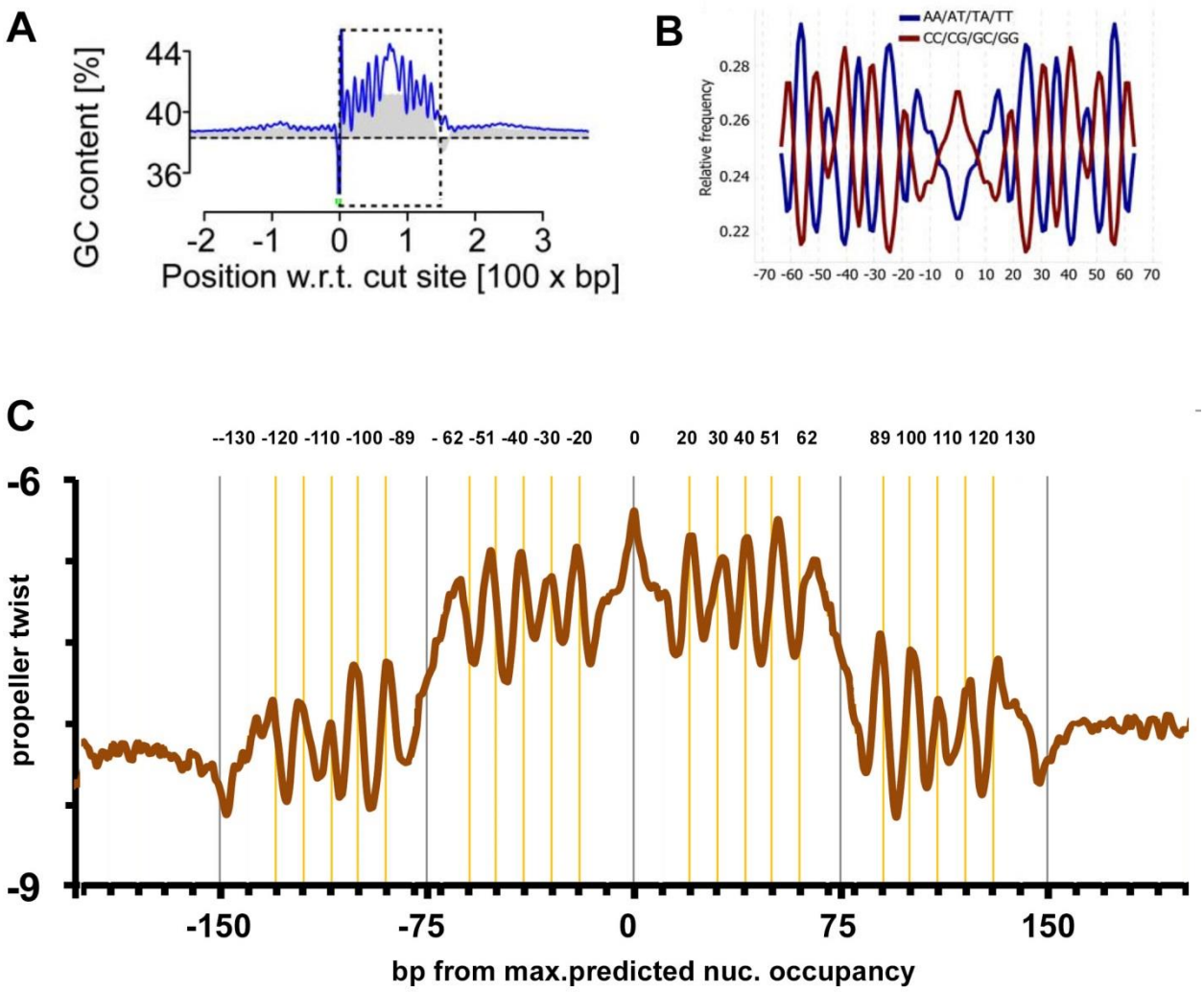

Fig. S11.

Comparison of 10 bp periodicity in 300 bp PT footprint found in this study with the previously characterized GC content periodicity of 150 bp yeast *in-vitro* nucleosome footprints. (*A,B*) Images from (Chung et al. 2010) (*A*) and (Field et al. 2008) (*B*) show the GC content peaks at positions ± 62, ± 51, ± 40, ± 30, ± 20 from the center (dyad) of the nucleosome-bound yeast DNA. These locations correspond to 10 out of 14 positions where the major groove of the DNA faces inwards towards the histone octamer (Chung and Vingron 2009). Images are reproduced under permission of Creative Commons Attribution License. 10 bp periodic oscillations of AT/GC content is a reproducible feature of nucleosomal DNA (Field et al. 2008; Chung and Vingron 2009) reflecting the DNA positioning on the nucleosome core (Satchwell et al. 1986; Richmond and Davey 2003). (*C*) Un-smoothened PT plot of 5,387 PSN group genomic regions, aligned on predicted maximal nucleosome occupancy point [nucmax] (Kaplan et al. 2009). Yellow lines show 10 PT peaks at ± 62, ± 51, ± 40, ± 30, ± 20 which exactly match the AC/GC content fluctuations seen above, and 10 additional PT peaks at ± 130, ± 120, ± 110, ± 100 and ± 89, within the same 10 bp period, which were not described before. PT values were derived using TFBSshape server http://rohslab.cmb.usc.edu/TFBSshape/ (Yang et al. 2014). Average values per 1 bp were plotted. Minor tick marks in the X-axis are 10 bp apart.

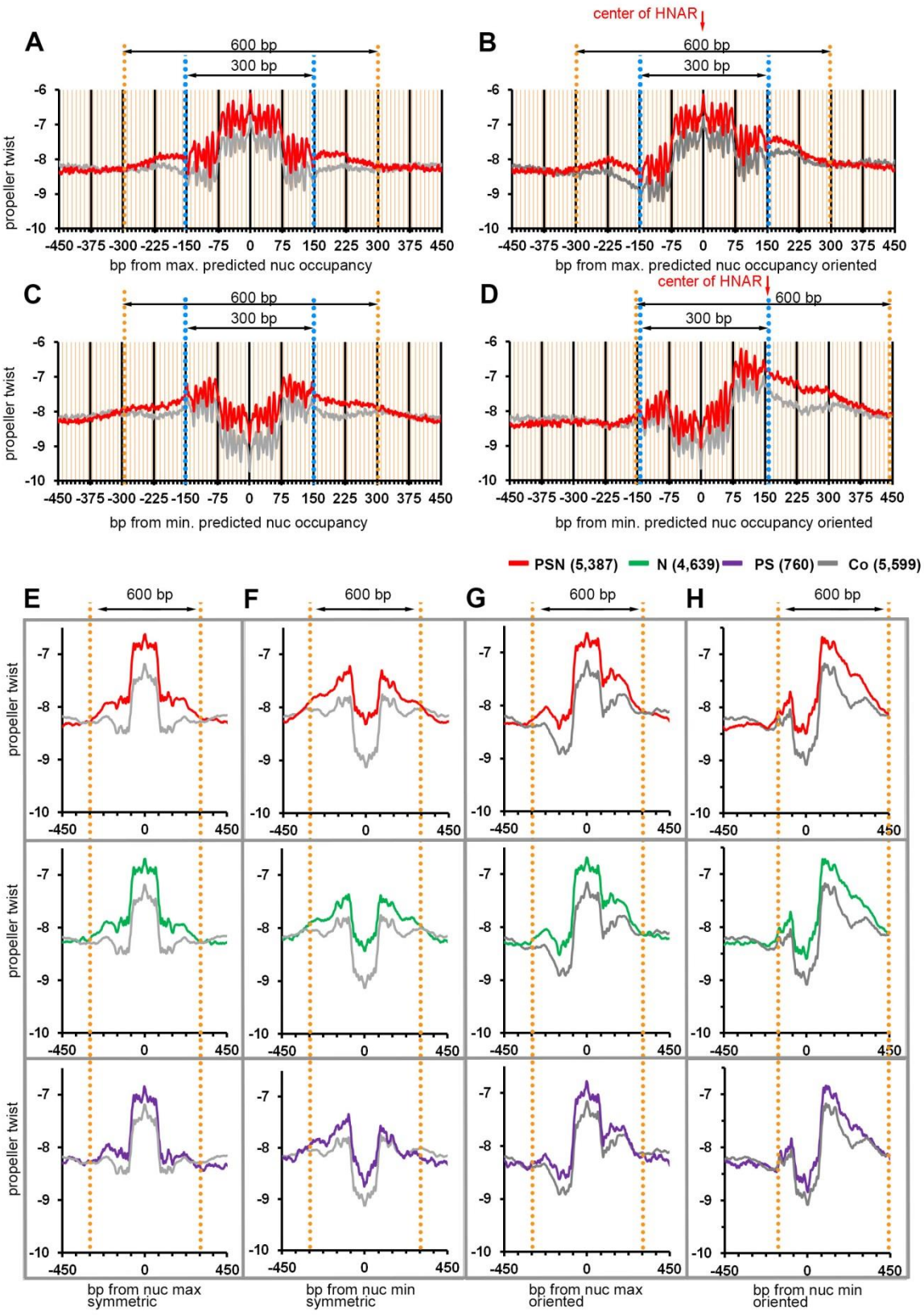

Fig. S12.

Estimation of the length of HNARs PSN, N and PS. We estimated the length of high PT region by the difference with the control and TFBS in symmetric and oriented plots, as 600 bp (PSN, N) and 500 bp (PS). (*A-D*) Un-smoothened PT plots and (*E-H*) smoothened PT plots

(80 bp moving average) of TFBS genomic regions compared to control, centered on predicted dyad [nucmax] or predicted intra-nucleosomal region [nucmin] (Kaplan et al. 2009) within 320 bp around the TFBS, oriented or symmetric as indicated. Average PT value per 1 bp was calculated for – 450 to + 450 bp from central position and plotted in Excel. Blue dotted lines indicate the borders of 300 bp periodic frame. Orange lines indicate the 600 bp when most of TFBS are different from control by higher PT (PSN compared to control). (*B,G*) Genomic regions centered on [nucmax] were oriented so that minimum nucleosome prediction value within +/-160 bp around [nucmax] is upstream (at the left side) from [nucmax] position (Table S1). (*D,H*) Genomic regions centered on [nucmin] were oriented the same way. The center of PSN HNAR is shown by red arrow. Note that [nucmax] is in the center of the HNAR, while [nucmin] is close to the edge of the HNAR in oriented PT plots.

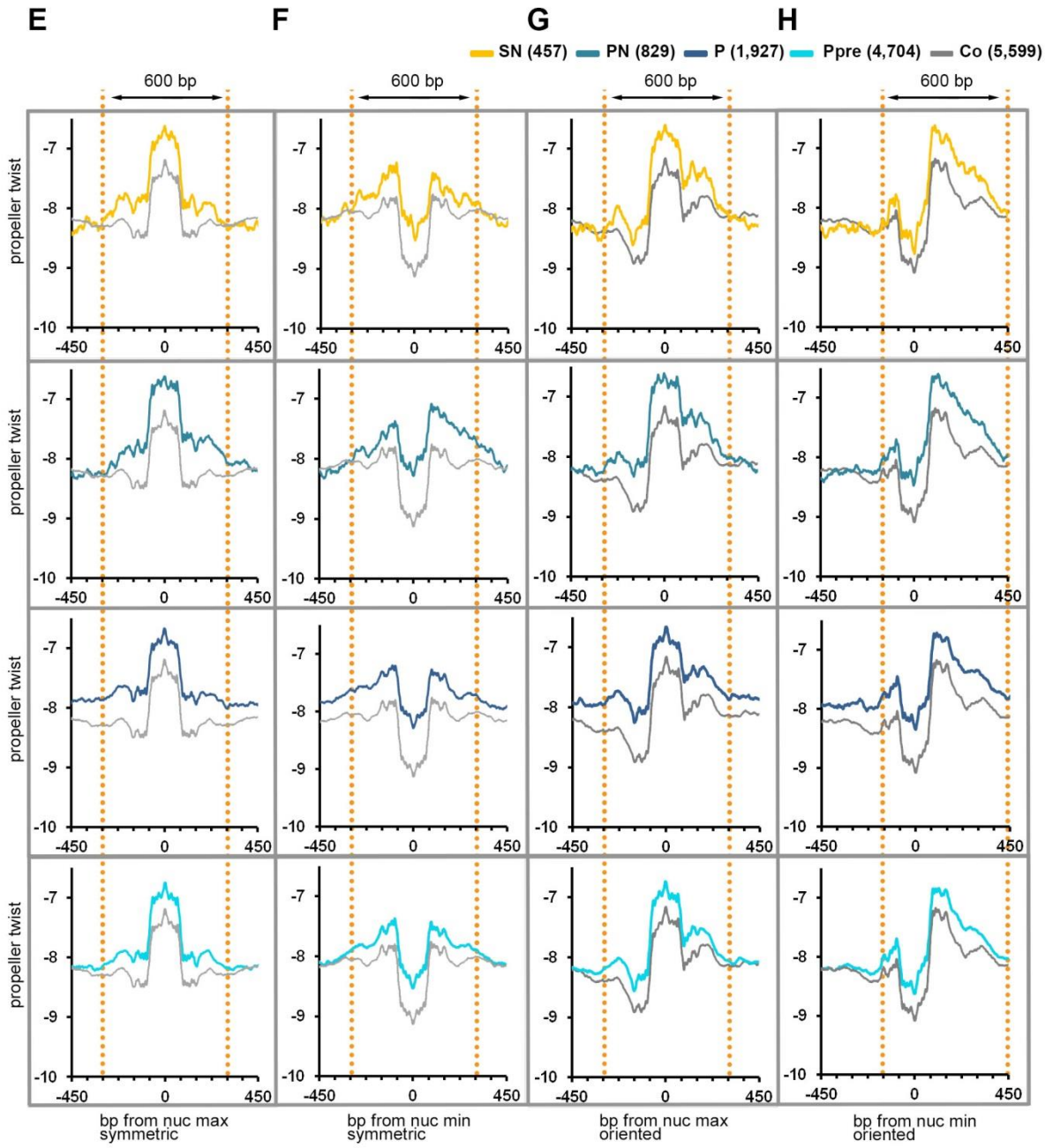

Fig. S13.

Estimation of the length of the HNAR SN, PN, P and Ppre (Fig. S12 continued). We estimated the length of high PT regions by the difference with the control and TFBS in symmetric and oriented plots as 600-700 bp (Ppre, SN, PN) and >900 bp (P) (*E-H*) see the legend in Fig. S12. Note that in the group of genomic regions, bound by Pou5f3 alone post-ZGA (P), higher PT compared to the control extends beyond +/-450 bp from [nucmin] or [nucmax]. P group was the only TFBS group which did not show enrichment in developmental enhancers, and dinucleotide repeats (see Fig. S2, S3). We hypothesize that 600 bp bounds of high PT/predicted nucleosome occupancy reflect the potential of DNA sequence with high PT to become an enhancer.

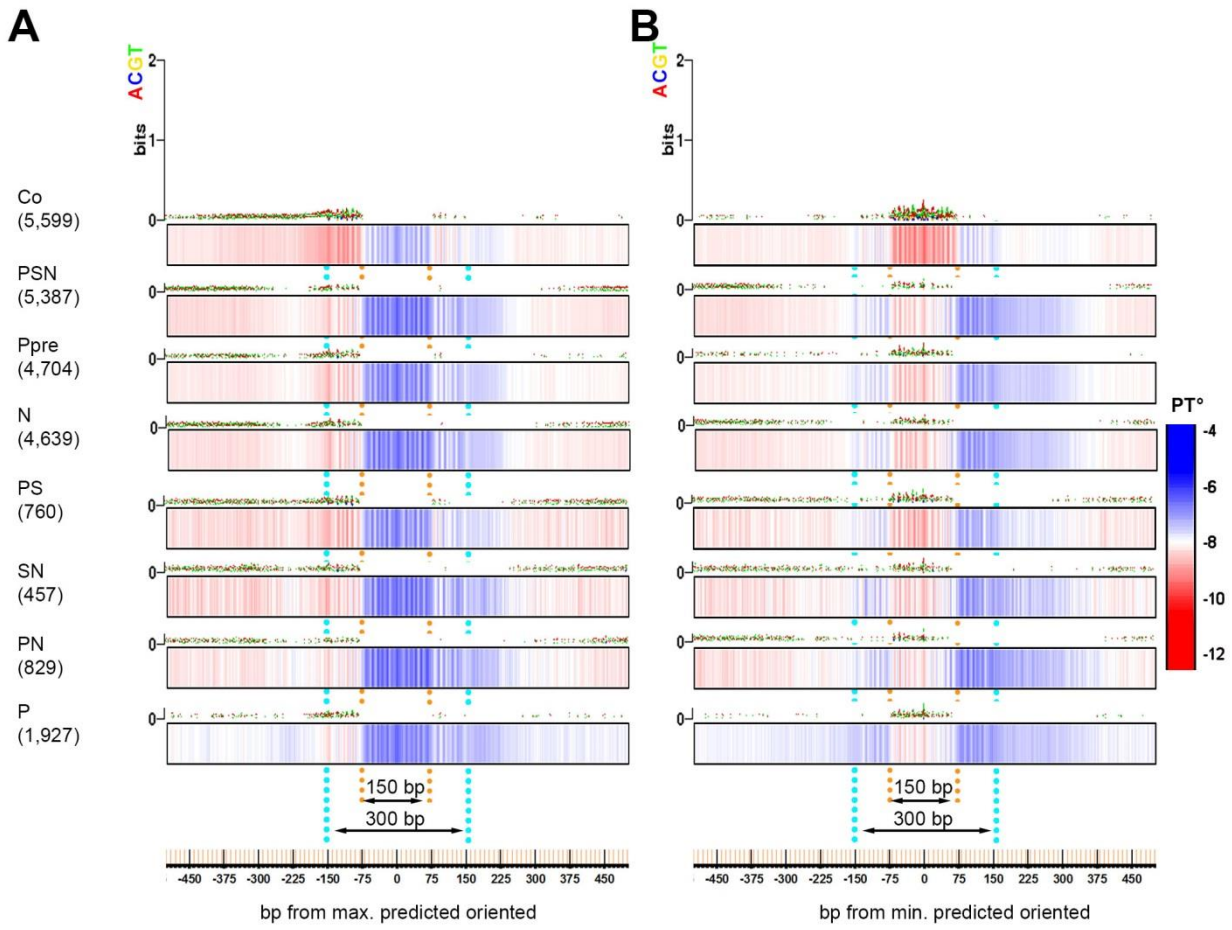

Fig. S14.

PT value sensitively reflects two attributes of DNA affinity to nucleosomes: GC content and dinucleotide rotational periodicity. Two sequence-based features: increased GC content and AT/GC periodicity, were reported to define high affinity of DNA to nucleosomes. The preference for GC pairs was related to a lower energetic cost required for deformation of the DNA to wrap around the histones (Drew and Travers 1985; Field et al. 2008; Chung and Vingron 2009; Tillo et al. 2010). The nucleosome prediction program of Kaplan et al. (Kaplan et al. 2009) is based to large extent on capturing periodic fluctuations and AT/GC content and does not account for PT or other DNA shape parameters. Here we compared PT and base composition plots around max. and min. nucleosome occupancy positions (Kaplan et al. 2009). We found that PT value and PT periodic oscillations visualize the sequence properties captured by nucleosome prediction program in a much more sensitive and perhaps direct way than sequence composition itself. Base composition and PT were determined using TFBSshape server (Yang et al. 2014) in 7 TFBS and control groups. The plots were downloaded from the server and contrasted together with the PT scale. (*A*) All genomic regions were centered on max. predicted nucleosome occupancy value [nucmax](Kaplan et al. 2009) within 320 bp around TFBS. The regions were oriented so that minimum nucleosome prediction value within +/-160 bp around [nucmax] is at the left (5') side. (*B*) All genomic regions were centered on min. predicted nucleosome occupancy value [nucmin] (Kaplan et al. 2009) within 320 bp around TFBS, and oriented so that [nucmin] is at the 5' side from [nucmax] position. Numbers of genomic regions is indicated in parentheses. Note that the Zebrafish genome is AT rich (median GC% = 36.8 (NCBI)), therefore absence of AT/CG bias i.e. around the max. positions indicates the increase in GC content over genomic average.

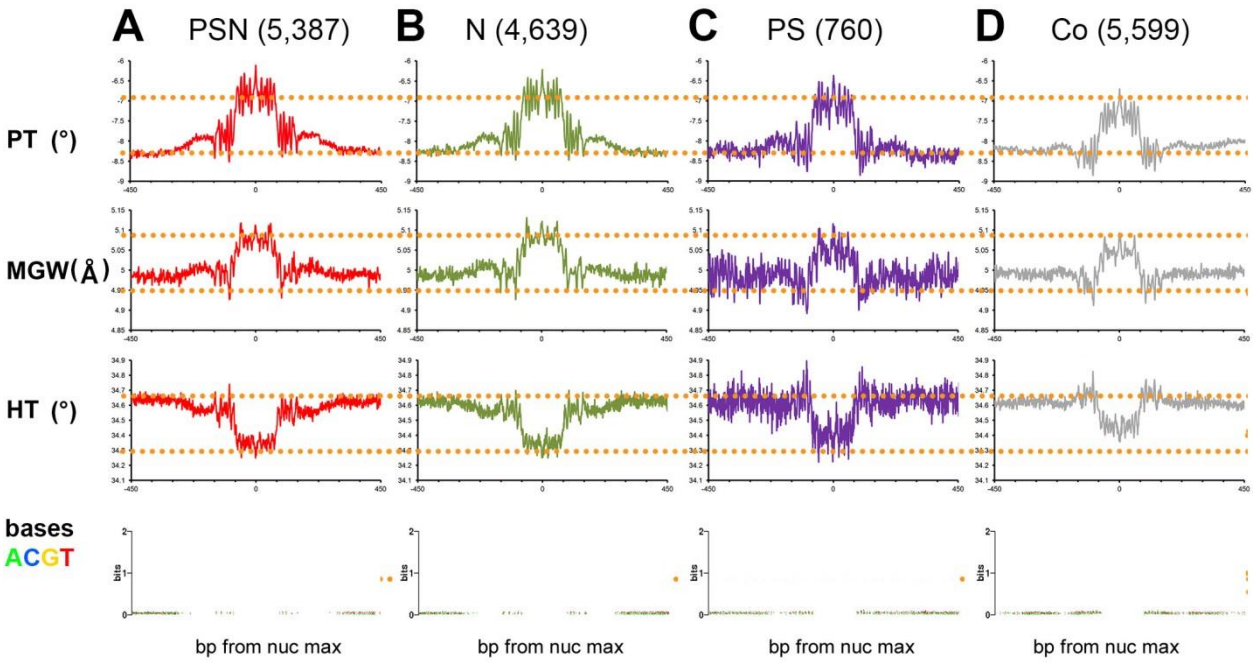

Fig. S15.

DNA shape and GC content on HNAR. Genomic regions aligned on maximum predicted nucleosome occupancy [nucmax] positions. (*A*) PSN, (*B*) N, (*C*) PS, (*D*) control, numbers of regions in parentheses. Propeller twist (PT), Minor Groove Width (MGW) and Helical Twist (HT) values were calculated using TFBSshape server http://rohslab.cmb.usc.edu/TFBSshape/ (Yang et al. 2014). Average value per bp for – 450 to + 450 bp from [nucmax] position are plotted. Average base composition plots were exported from TFBSshape server. Symmetric central nucleosome footprint patterns are similar between the groups, while parameter values are different in TFBS and control groups. Orange dotted lines across the graphs illustrate this difference: in TFBS groups, PT and MGW values were higher while HT was lower than in the control. Base pair plots show weak enrichment for A/T nucleotides outside of central nucleosome footprint (PS, co) or broader (in PSN and N). Zebrafish genome is AT rich (median GC% = 36.8 (NCBI)), the absence of AT/CG bias around the max. positions indicates the subtle increase in GC content over genomic average.

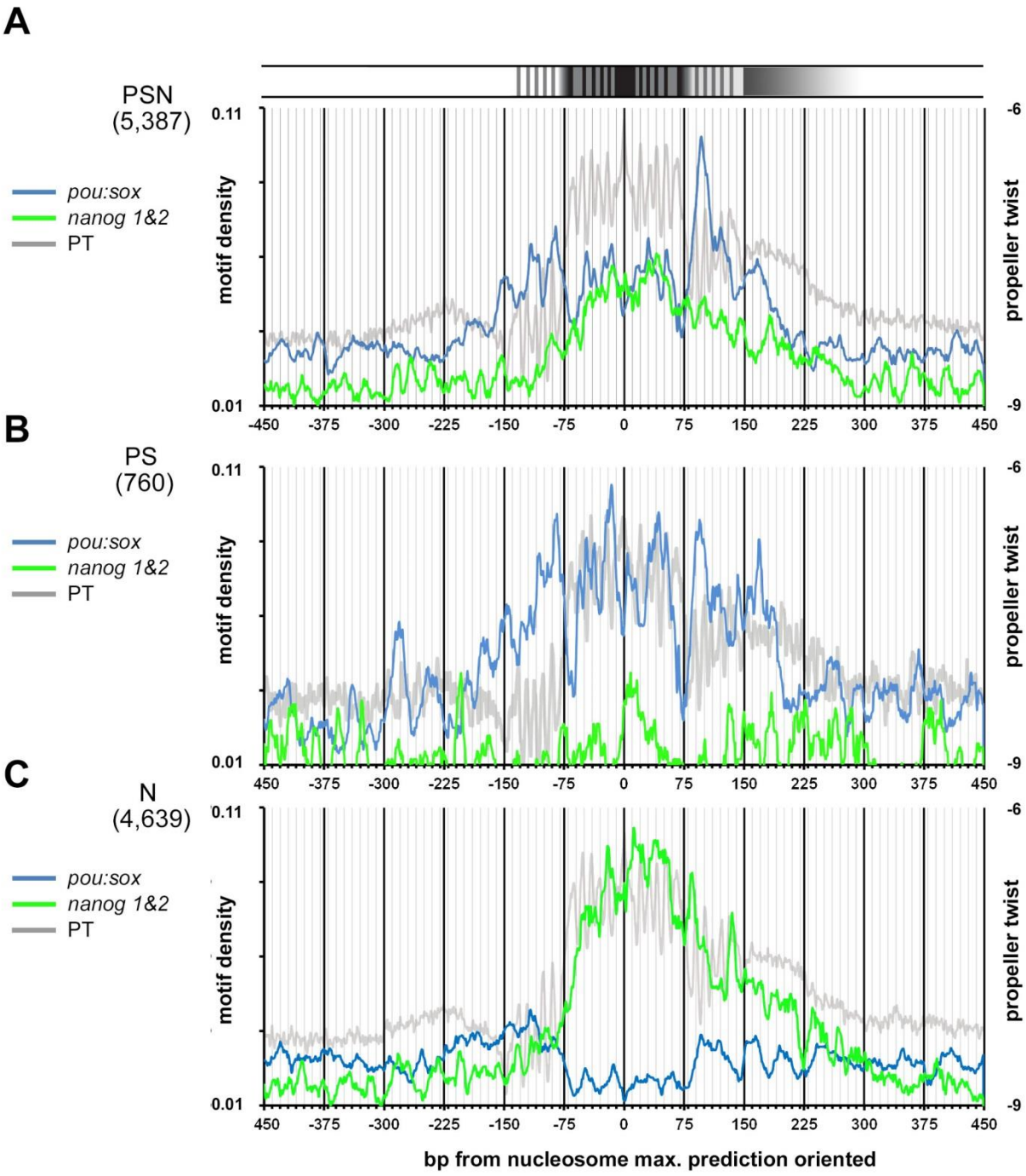

Fig. S16.

Distribution of Pou5f3/SoxB1 and Nanog recognition motifs on HNAR for PSN, PS and N group, average per 1 bp. (*A*-*C*) Genomic regions were aligned at [nucmax], and oriented with minimum to the left. Mean density of *pou:sox* and *nanog1&2* motifs was plotted with 1bp steps on PSN (*A*), PS (*B*) and N (*C*) groups. Motif density in bp motif/bp sequence. Propeller Twist profile shown in gray (°scale at the right). Note that *pou:sox* motifs on PSN sharply peak between positions +90 and +100, outside of the central footprint, while *nanog* motifs in N and PSN are located in the middle footprint.

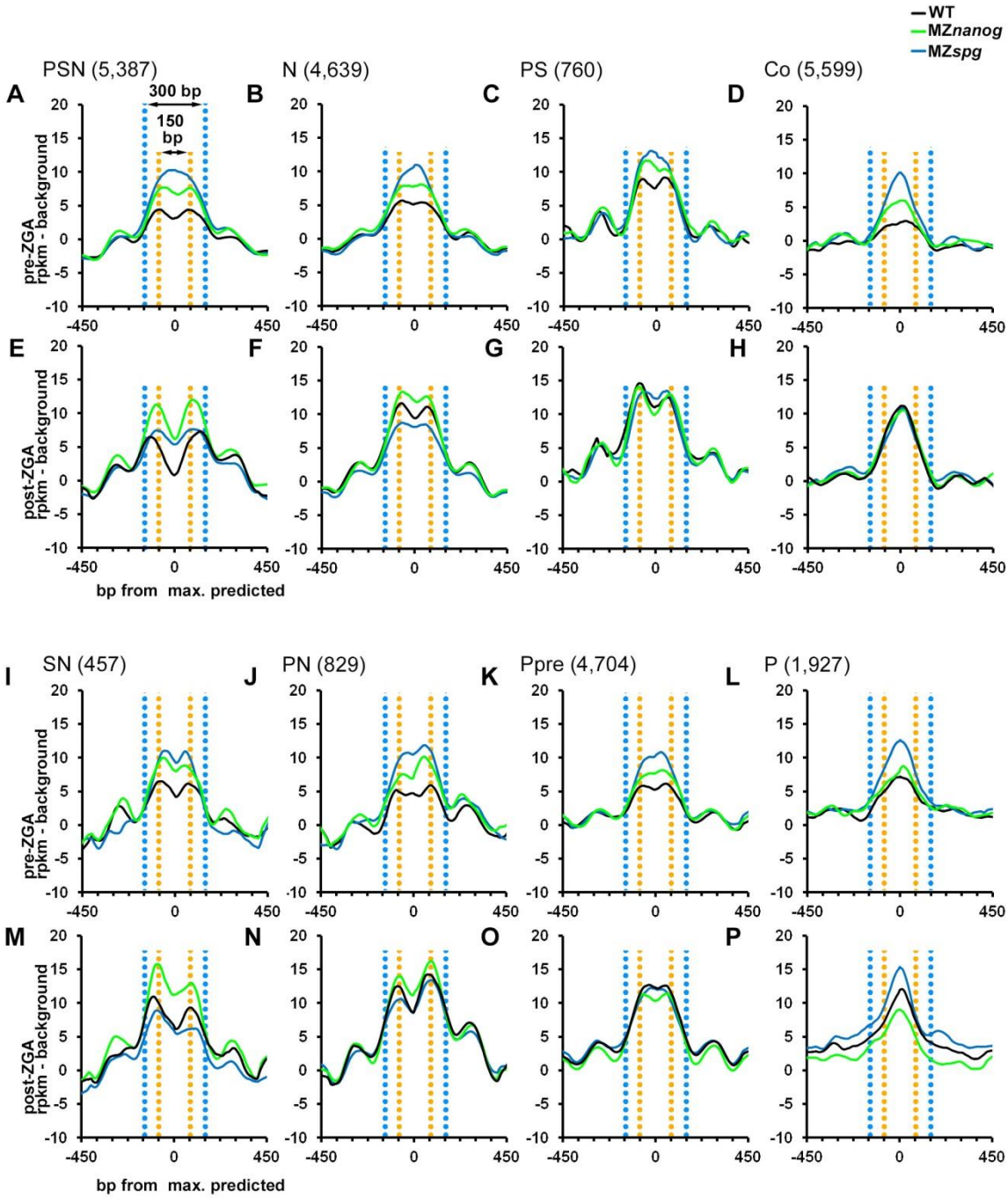

Fig. S17.

Experimental nucleosome occupancy centered on maximal predicted nucleosome occupancy value (predicted dyad), symmetric plots. Mean nucleosome occupancy plots in the WT (black), MZ*nanog* (green) and MZ*spg* (blue) aligned on the max. predicted nucleosome occupancy [nucmax] in TFBS and control groups as indicated. X-axis: bp from the [nucmax] position. Y-axis: nucleosome occupancy in RPKM minus background, smoothened plots (40 bp moving average), pre-ZGA (above) or post-ZGA (below). Orange and blue dotted lines mark +/-75bp from [nucmax] (borders of the medial nucleosome footprint) and +/-150 bp (borders of central periodic region) from [nucmax]. At Pre-ZGA, nucleosomes phasing in the WT to +/- 75 bp is seen in all groups except control (*D*) and P (*L*). Pre-ZGA, nucleosome occupancy is the highest in MZ*spg*, lowest in the WT and intermediate in MZ*nanog* in all groups. At post-ZGA, nucleosomes are cleared from the

center in WT PSN group (*A*, second row), stay at [nucmax] position in the center in control and P groups (*H*, *P* second rows) and stay on +/- 75 bp positions in all the other groups.

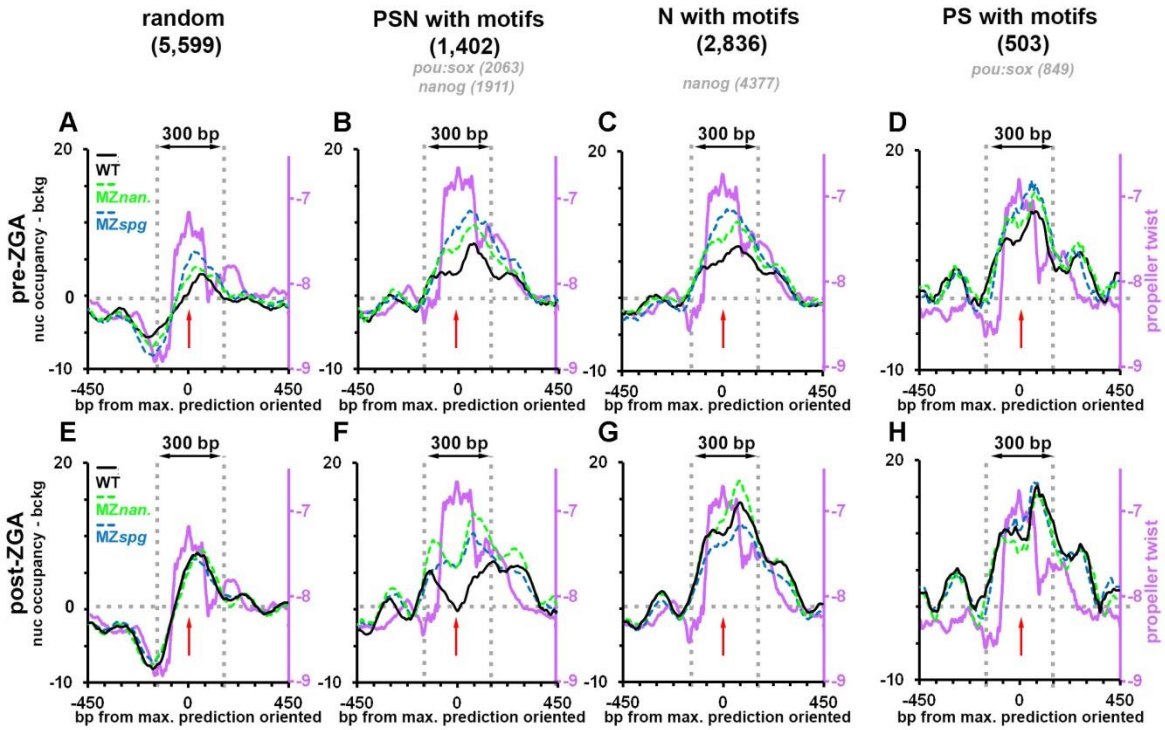

Fig. S18.

Pre-ZGA and post-ZGA changes of real vs. predicted maximal nucleosome density on PSN, N and PS HNARs compared to random control nucleosome footprint. (*A-H*) Genomic regions indicated on top aligned at [nucmax] and oriented by increasing predicted nucleosome occupancy (Table S1)**.** Gray dotted lines 300 bp apart and show the borders of central rotational frame, horisontal lines show the background. Right axis- PT, left axis – nucleosome occupancy pre-ZGA (*A-D*) and post-ZGA (*E-H*). Violet lines show smoothened PT values over 80 bp, red arrow shows the center of predicted nucleosome footprint. Note that in PSN region the [nucmax] is cleared from nucleosomes in the wild-type (red arrows), which is not the case in N and PS.

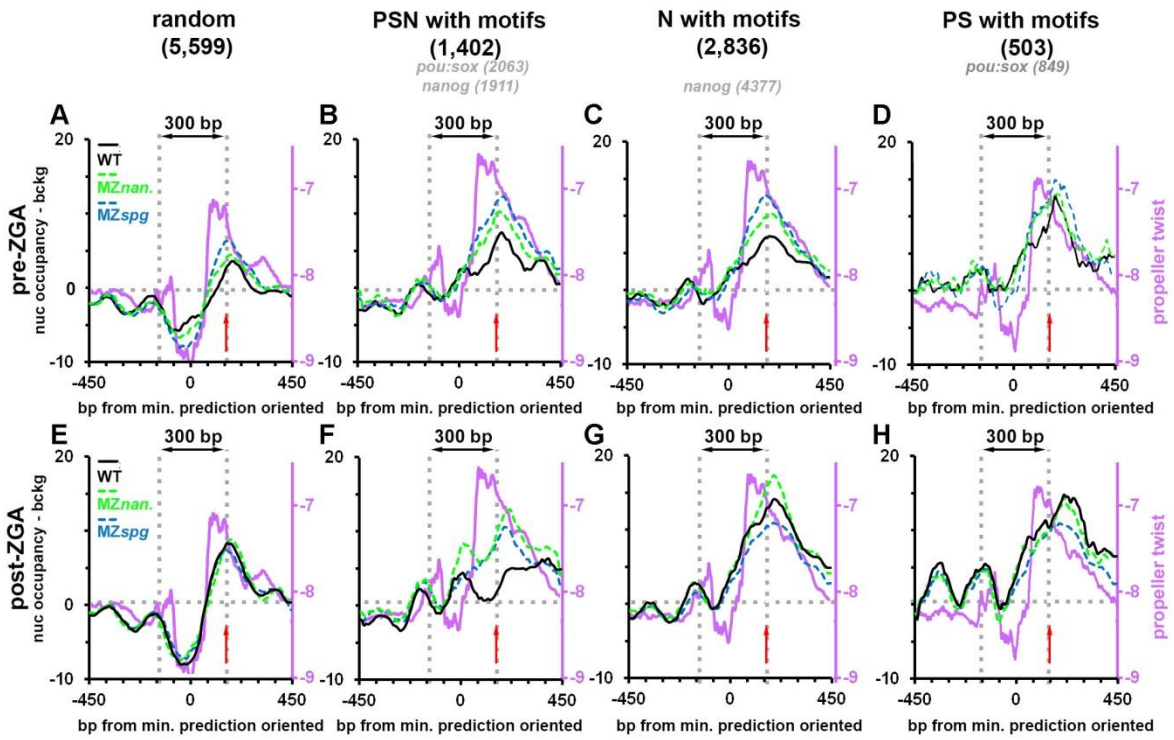

Fig. S19.

Pre-ZGA and post-ZGA changes of real vs. predicted minimal nucleosome density on PSN, N and PS HNARs compared to random control inter-nucleosomal region. (*A-H*) Genomic regions indicated on top aligned at [nucmin] and oriented by increasing predicted nucleosome occupancy (Table S1). Gray dotted vertical lines 300 bp apart and show the borders of periodic frame around [nucmin], horisontal lines show the background. Right axis-PT, left axis – nucleosome occupancy pre-ZGA (*A-D*) and post-ZGA (*E-H*). Violet lines show smoothened PT values over 80 bp, red arrows are average [nucmax] position (+150 bp from [nucmin]). Note that in PSN region the [nucmax] is cleared from nucleosomes in the wild-type (red arrows), which is not the case in N and PS.

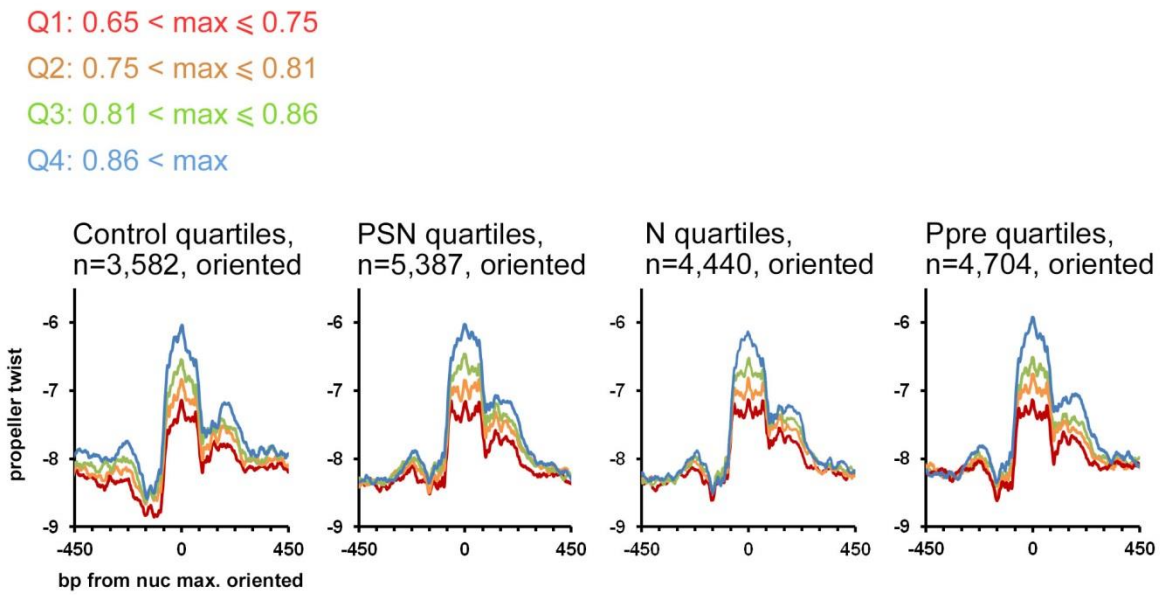

Fig. S20.

PT oriented profiles by quartile. 20,747 genomic regions, with max. nucleosome predicted value within 320 bp around the TF binding site/center more than 0.65, were divided to four quartiles. Max. values per quartile are shown above. Genomic regions were aligned at [nucmax], and were oriented so that minimum nucleosome prediction value within +/-160 bp is upstream (at the left side) from [nucmax] position. Mean PT values per quartile in the indicated groups were calculated in TFBShape server and plotted in Excel. [nucmax] values are listed in Table S1.

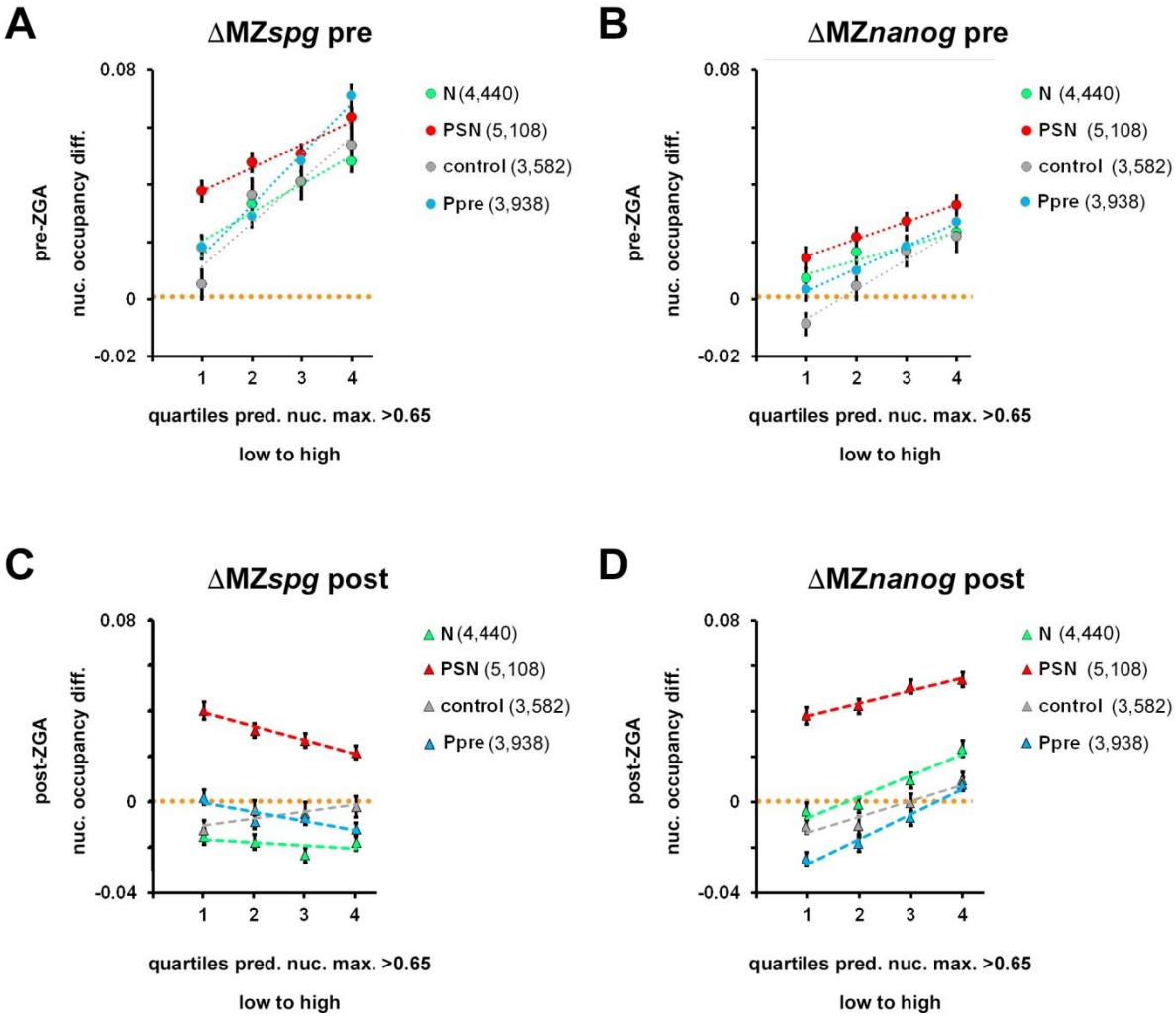

Fig. S21.

ΔMZ*spg* and ΔMZ*nanog* dependence on nucleosome footprint strength in PSN, N, Ppre and control groups. Normalized difference between the nucleosome occupancy in the mutant and wild-type (Δmut) for 320 bp region around [nucmax] position was calculated as (rpkm(mut)-rpkm(wt))/(rpkm(mut)+rpkm(wt)). (*A*) pre-ZGA in MZ*spg*, Δmut increases with predicted nucleosome occupancy in all groups. (*B*) Pre-ZGA in MZ*nanog*, Δmut increases with predicted nucleosome occupancy in all groups. (*C*) Post-ZGA in MZ*spg*, Δmut does not depend on predicted nucleosome occupancy in N and control, and decreases with predicted nucleosome occupancy in PSN and Ppre. groups. (*D*) Post-ZGA in MZ*nanog*, Δmut increases with predicted nucleosome occupancy in all groups. Note that however, when compared to pre-ZGA, the Δmut MZ*nanog* threshold is changed: in N and all groups except PSN, in two lower quartiles Δmut MZ*nanog* is zero or negative.

Table S3.

Summary of mapping.

stage/genotype	total mapped reads	scale factor (reads per bp to RPKM)	background, RPKM
Dome_WT	323,368,745	3.092444819	43.80891336
Dome_MZ*spg*	332,823,523	3.004595321	43.82624617
Dome_MZ*nanog*	275,837,911	3.625317479	43.3100867
512c_WT	179,050,944	5.585002668	44.90686207
512c_MZ*spg*	162,739,598	6.144785979	45.19895504
512c_MZ*nanog*	151,233,737	6.612281227	44.68716792

Table S4.
1-way ANOVA of post to pre ZGA differences nucleosome occupancy at different TF regions in the WT (Fig. S 5D), and of mutant to the wild type differences pre- and post-ZGA for MZ*nanog* and MZ*spg* mutants (Fig. S5B,C,E,F).

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Response: [delta]WT(post-pre)				Fig.S5 D	
TFs_binding	7	33.67	4.810	147.77	<.0001
Error	30506	992.91	0.033		
Response: [delta]Mut_spg-pre)				Fig.S5 B	
TFs_binding	7	14.48	2.069	63.53	<.0001
Error	30500	993.40	0.033		
Response: [delta]Mut_Nanog-pre)				Fig.S5 C	
TFs_binding	7	5.43	0.776	28.48	<.0001
Error	30495	830.96	0.027		
Response: [delta]Mut_spg-post)				Fig.S5 E	
TFs_binding	7	14.69	2.099	98.14	<.0001
Error	30507	652.44	0.021		
Response: [delta]Mut_Nanog-post)				Fig.S5 F	
TFs_binding	7	34.35	4.907	219.33	<.0001
Error	30508	682.49	0.022		

Table S5.
1-way ANOVA of mutant-WT nucleosome occupancy differences pre-ZGA (Fig. 1G, K) and post-ZGA (Fig. 2D, G) at TF-binding regions with different number of non-overlapping TF binding sites. Control values are not included into this analysis. Colors correspond to those on Fig. 1 and Fig.2.

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
	Response: [delta]Mut_Nanog-pre)			Fig. 1G	
N(nanog motifs) at N regions	4	0.10	0.024	1.08	0.37
Error	5448	120.46	0.022		
N(nanog motifs) at PSN regions	4	0.08	0.021	0.94	0.44
Error	6327	140.12	0.022		
	Response: [delta]Mut_Nanog-post)			Fig. 2D	
N(nanog motifs) at N regions	4	0.02	0.004	0.31	0.87
Error	5448	76.96	0.014		
N(nanog motifs) at PSN regions	4	1.86	0.466	21.65	9.10E-18
Error	6327	136.04	0.022		
	Response: [delta]Mut_spg-pre)			Fig. 1K	
N(pou:sox motifs) at PS regions	4	0.04	0.010	0.32	0.87
Error	941	29.73	0.032		
N(pou:sox motifs) at PSN regions	4	0.36	0.089	4.05	0.0028
Error	6327	139.61	0.022		
	Response: [delta]Mut_spg-post)			Fig. 2G	
N(pou:sox motifs) at PS regions	4	0.25	0.063	3.58	0.0067
Error	941	16.66	0.018		
N(pou:sox motifs) at PSN regions	4	7.81	1.953	103.46	1.74E-85
Error	6327	119.44	0.019		

Table S6.
T-tests of the difference between nucleosome occupancy (Δmut) for the TF bound regions lacking motifs and that in randomly chosen control regions, pre-ZGA (Fig. 1G, K) and post-ZGA (Fig. 2D, G)**.** Colors correspond to those on the figures.

	Mutant=MZ*nanog*				Mutant=MZ*spg*			
	Δmut_pre		Δmut_post		Δmut_pre		Δmut_post	
	t	p	t	p	t	p	t	p
Control vs. PSN *pou:sox* motifs=0	8.82	6.68E-19 (Fig. 1G, red)	16.09	1.71E-57 (Fig.2D, red)	14.11	9.59E-45 (Fig. 1K, red)	2.05	0.04 (Fig. 2G, red)
Control vs. PS *pou:sox* motifs=0					2.92	0.0035 (Fig. 1K, purple)	-0.31	0.76 (Fig. 2G, purple)
Control vs. N *nanog* motifs=0	5.93	3.10E-09 (Fig. 1G green)	4.94	7.81E-07 (Fig.2D, green)				

Table S7 and S8.
1-way ANOVA of the differences between [nucmax] and [nucmin] among TF binding groups. Related to Fig. S8.

	Response:	nucmax			
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
TFs_binding_group	7	61.49	8.78	521.15	0
Error	24294	409.47	0.02		
	Response:	nucmin			
Source					
TFs_binding_group	7	60.50	8.64	523.81	0
Error	24294	400.82	0.02		

Table S9.
1-way ANOVA of mutant-WT nucleosome occupancy differences pre- and post-ZGA at regions that fall into 1st through 4th quartiles of predicted nucleosome occupancy (Fig. 5D,H). Colors correspond to those on the figures.

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Response: Δmut MZ*nanog* pre-ZGA					
[nucmax] quartiles	3	1.238	0.413	16.69	7.92E-11
Error	20739	512.538	0.025		
Response: Δmut MZ*spg* pre-ZGA					
[nucmax] quartiles	3	3.547	1.182	42.61	1.93E-27
Error	20743	575.578	0.028		
Response: Δmut MZ*nanog* post-ZGA					
[nucmax] quartiles	3	2.403	0.801	38.60	7.28E-25
Error	20743	430.485	0.021		
Response: Δmut MZ*spg* post-ZGA					
[nucmax] quartiles	3	0.156	0.052	2.77	0.040
Error	20743	389.867	0.019		

Table S1. (separate file)
Genomic regions used in this study, nucmax and nucmin values.

Table S2. (separate file)
Positional Weight Matrices for the motifs found in this study.