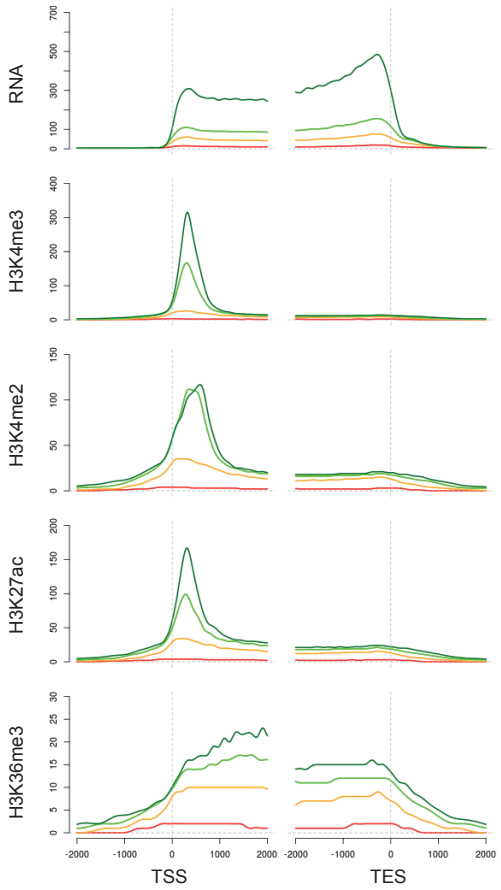
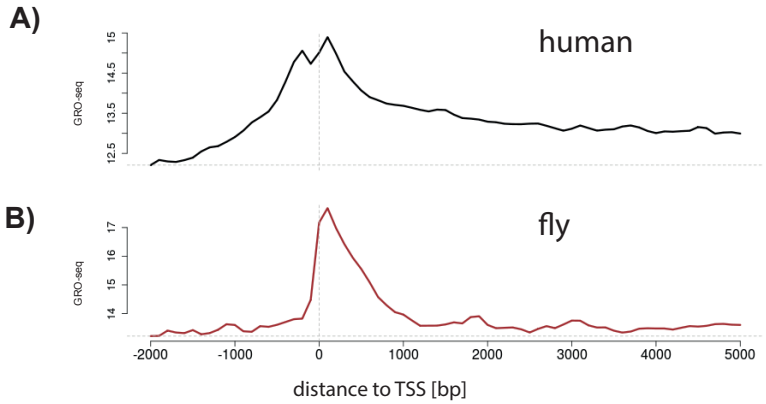
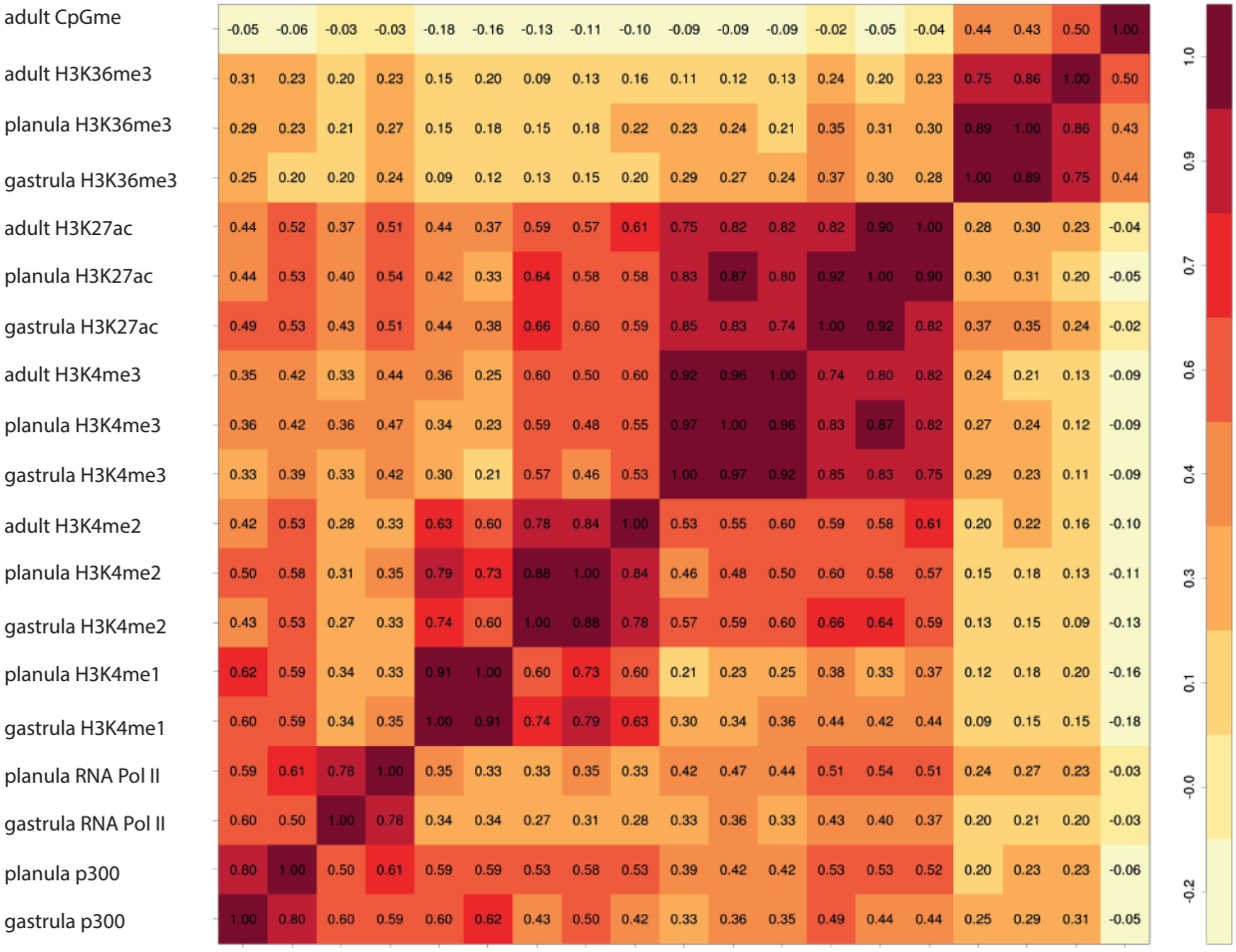
Schwaiger Supplemental Figure S1



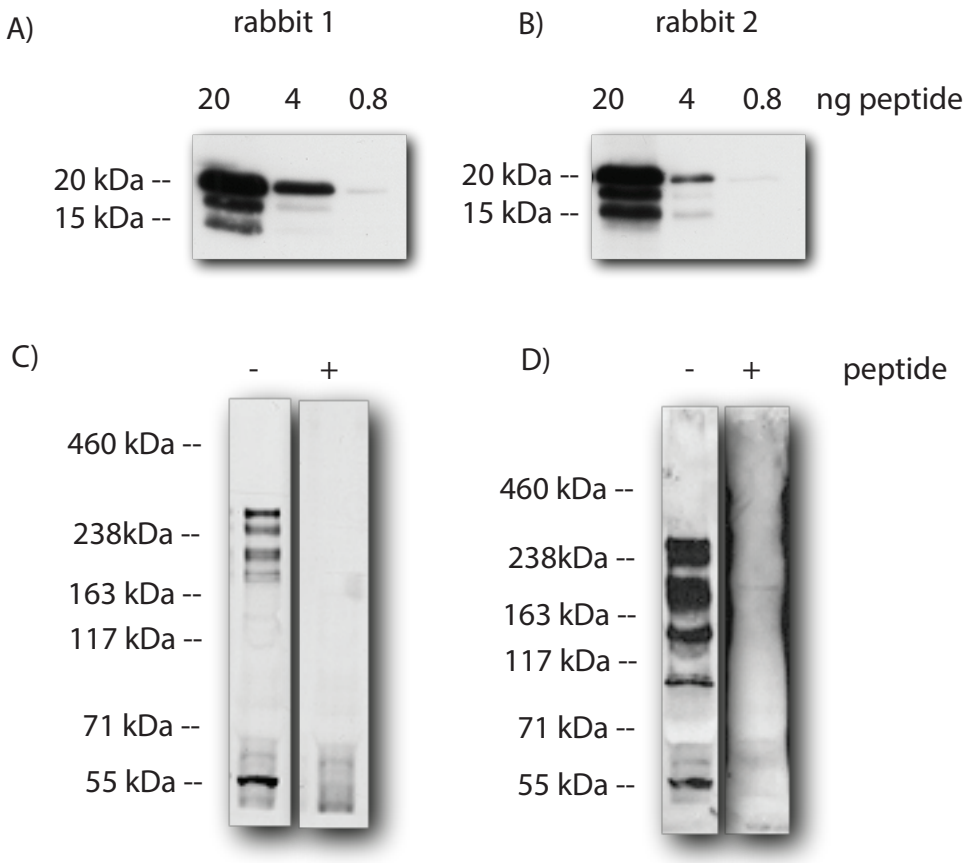
Supplemental Figure S1: The distribution of chromatin modifications across *Nematostella* genes. Genes were aligned relative to their annotated transcription start (left plots) and end (right plots) sites. The x-axis in each plot represents the position within the gene relative to transcription start sites and 3' ends. The y-axis in each plot represents the relative enrichment for RNA-seq reads and several histone modifications (H3K4me3,H3K4me2,H3K27ac,H3K36me3) in *Nematostella* polyps. Red line = non-expressed genes (FPKM < 1.5). Orange line = Lowly expressed genes. Green line = Medium expressed genes. Dark green line = Highly expressed genes. Expressed genes = FPKM > 2. To define lowly, medium and highly expressed genes, all expressed genes were sorted based on FPKM values and seperated into three groups containing roughly equal numbers of genes.



Supplemental Figure S2: The distribution of GRO-seq reads across human and fly genes. The distribution of genome-wide nuclear run-on sequencing (GRO-seq) reads (y-axis) is shown around human (A) (Core et al., 2008) and *Drosophila* (B) (Kharchenko et al., 2012) genes. Genes were aligned relative to their annotated transcription start sites. The x-axis in each plot represents the position within the gene relative to transcription start sites. The y-axis in each plot represents the number of GRO-seq reads. Note the enrichment of nascent transcripts upstream of TSSs in human, but not in *Drosophila*, correlating well with the distribution of H3K4me3 shown in Figure 1.

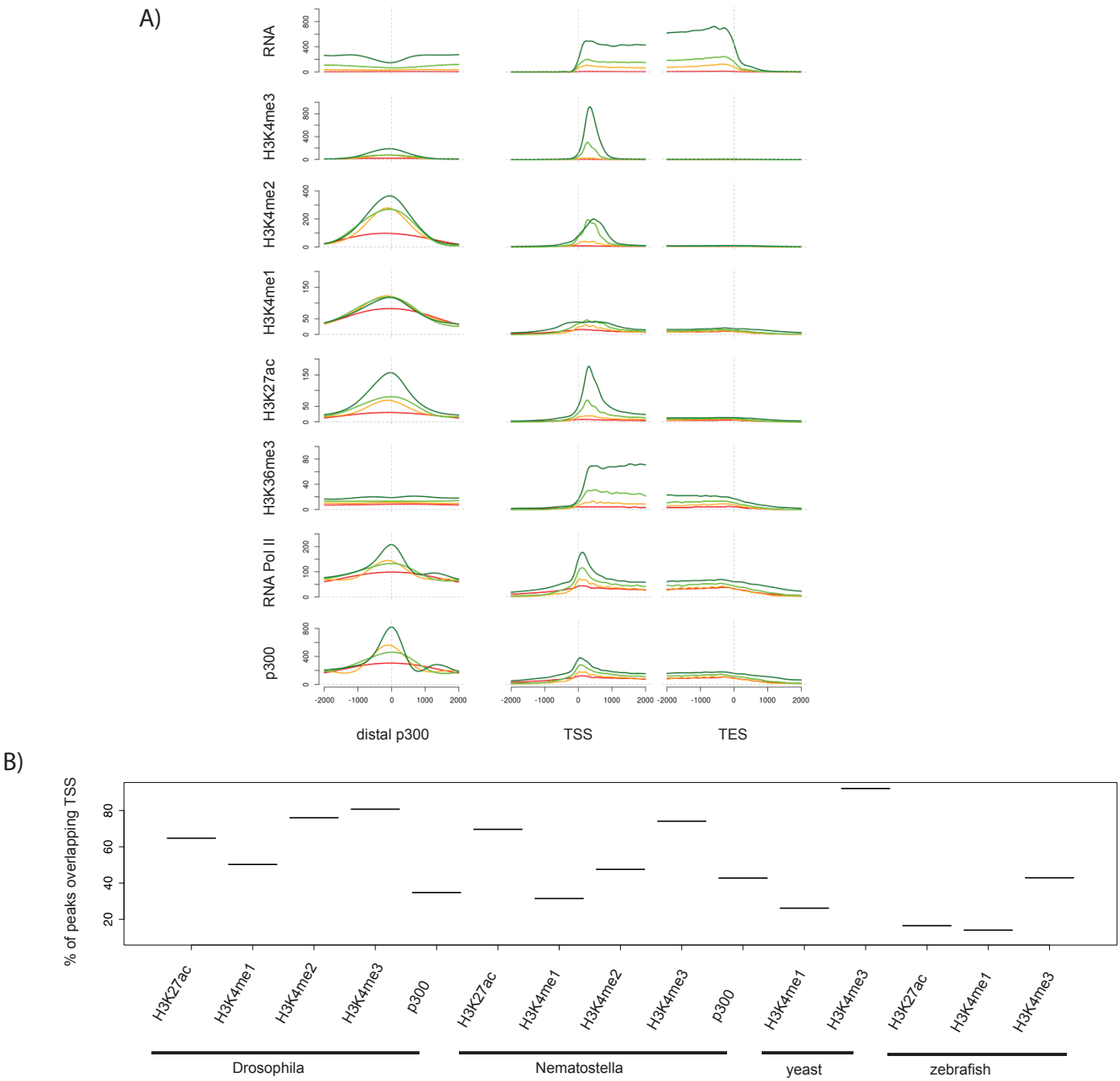


Supplemental Figure S3: DNA methylation correlates with H3K36me3 but not H3K4me3. Pearson correlation coefficients between chromatin modifications, p300, RNA Polymerase II and CpG methylation (Zemach et al., 2010). Correlation coefficients between all experiments (combined biological replicates) are shown. Underlying colors indicate the similarity between the different datasets. Histone modifications show high correlations between developmental stages. H3K36me3 is the only chromatin mark which correlates with DNA methylation.

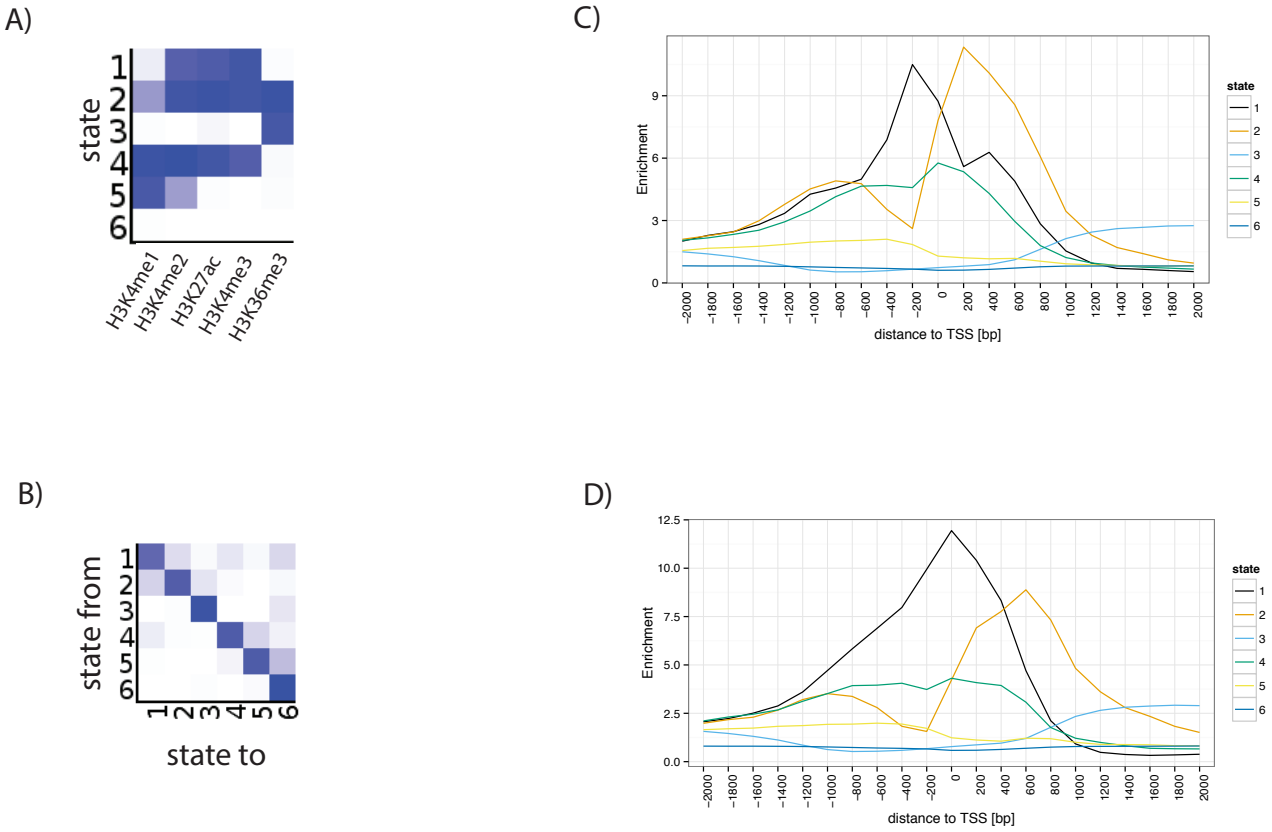


Supplemental Figure S4: The p300 antibodies recognize endogenous Nematostella p300 protein. A-B) A fragment of the p300 protein (predicted molecular weight 15kDa) containing a 6xHis tag is recognized by the affinity purified p300 antibody raised against this peptide in rabbit #1 (A) and #2 (B). C-D) The affinity purified p300 antibody from rabbit #1 (C) and #2 (D) recognizes different isoforms of the p300 protein (predicted molecular weight 250kDa) in Nematostella vectensis whole embryo extracts. The p300 protein is not recognized when the antibody is incubated with the p300 peptide first. P300 antibodies from both rabbits were used for p300 ChIP-seq, where each biological replicate was performed with antibody from one rabbit.

Schwaiger Supplemental Figure S5

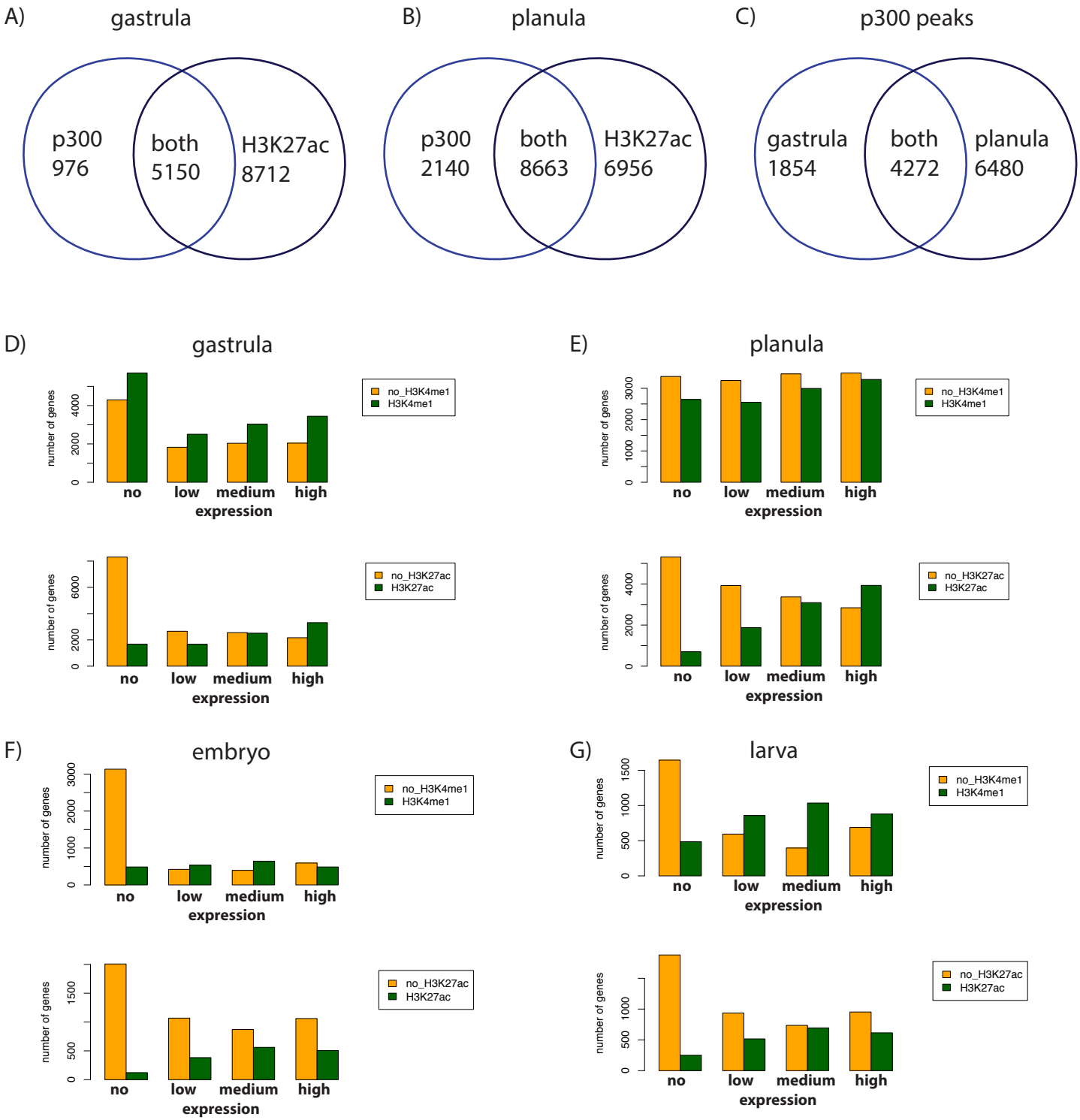


Supplemental Figure S5: Enhancer related chromatin modifications are associated with distal p300 peaks, while promoter related chromatin modifciations are more often associated with TSSs. A) Distribution of chromatin marks, RNA Polymerase II and p300 across distal p300 peaks and genes. Gastrula p300 peaks which do not overlap with TSSs were aligned relative to their peak summit (left plots) and genes relative to their annotated transcription start (middle plots) and end (right plots). The x-axis in each plot represents the position within the gene relative to peak summits, transcription start sites and 3' ends. The y-axis in each plot represents the relative enrichment for epigenomic variables such as several histone modifications in the gastrula stage. Red line = non-expressed genes (FPKM < 1.5). Orange line = Lowly expressed genes (=log2(FPKM) > 1 and < 2.3). Green line = Medium expressed genes (=log2(FPKM) > 2.3 and < 3.8). Dark green line = Highly expressed genes (=log2(FPKM) > 3.8). Expressed genes = FPKM > 2. The expressed genes were divided into three bins of equal number of genes according to their FPKM values to define the low, middle, and highly expressed genes. To define lowly, medium and highly expressed genes, all expressed genes were sorted based on gastrula stage FPKM values and seperated into three groups containing roughly equal numbers of genes. B) We determined the overlap of histone modification peaks in *Nematostella, Drosophila,* yeast, and zebrafish with TSSs. The percentage of peaks overlapping TSSs is shown on the y-axis (median of the results from several different developmental stages analyzed (Nematostella: 2-3 stages, Drosophila: 3 stages, zebrafish:3 stages, yeast: 1 sample.)). While the number of peaks overlapping a TSS varies greatly between all species, the relationship between the different histone modifications is highly similar.



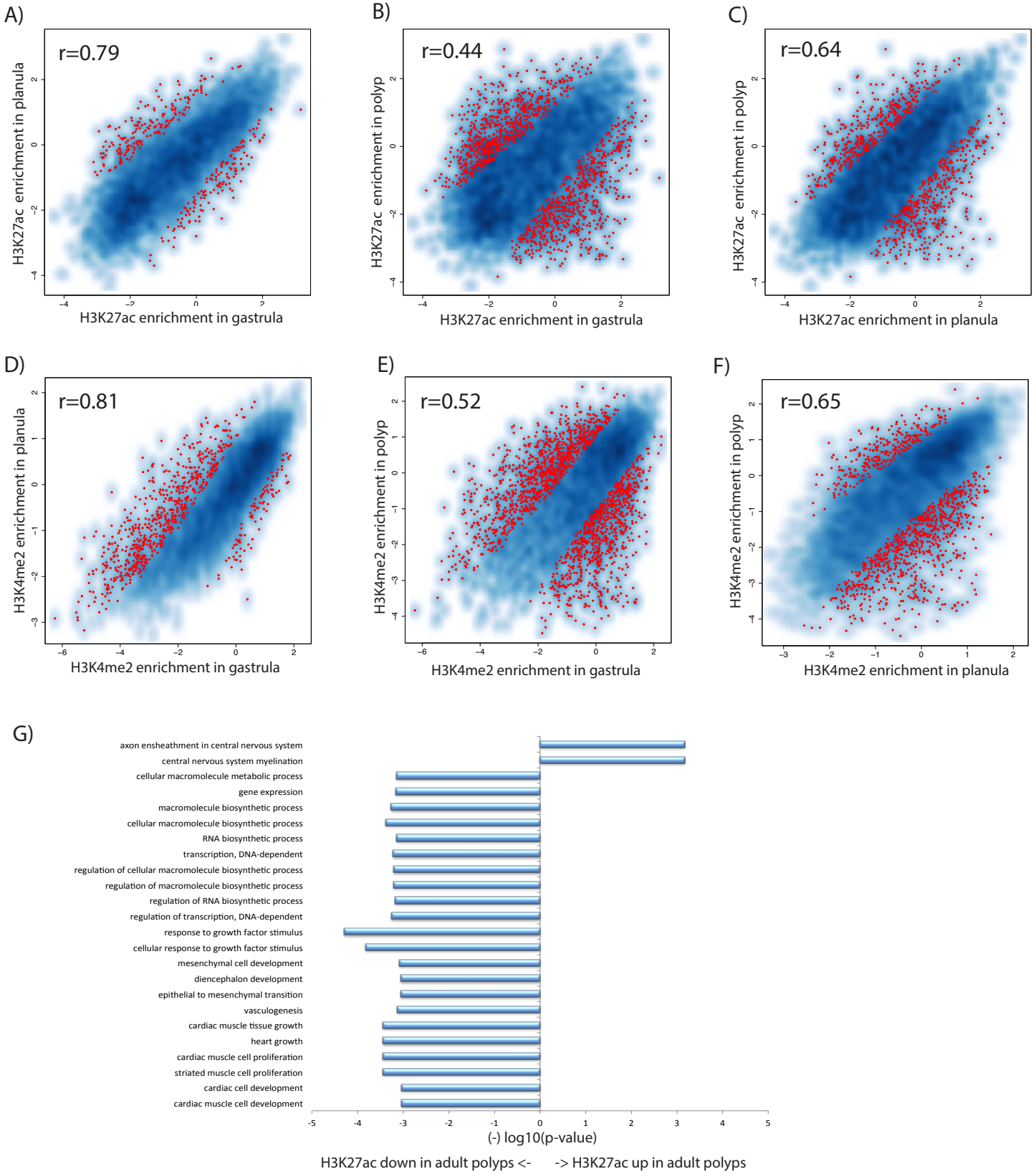
Supplemental Figure S6: Definition of chromatin states using ChromHMM. A) A six state hidden Markov model (Ernst and Kellis 2012) was trained on all chromatin modifications in gastrula and planula together, resulting in the depicted emission parameters (frequency with which a given mark is found at genomic positions corresponding to the chromatin state, where darker blue color indicates higher frequency on a scale of 0-1) of each modification (x-axis) in the six states (y-axis). B) Transition parameters (frequency with which a given state changes into another state at the neighboring location on the chromosome, where darker blue color indicates higher frequency on a scale of 0-1) between states of the hidden Markov model. C-D) Enrichment (y-axis) of each chromatin state in planulae (C) or gastrulae (D) in 200bp windows around aligned TSSs of all genes (x-axis).

Schwaiger Supplemental Figure 7

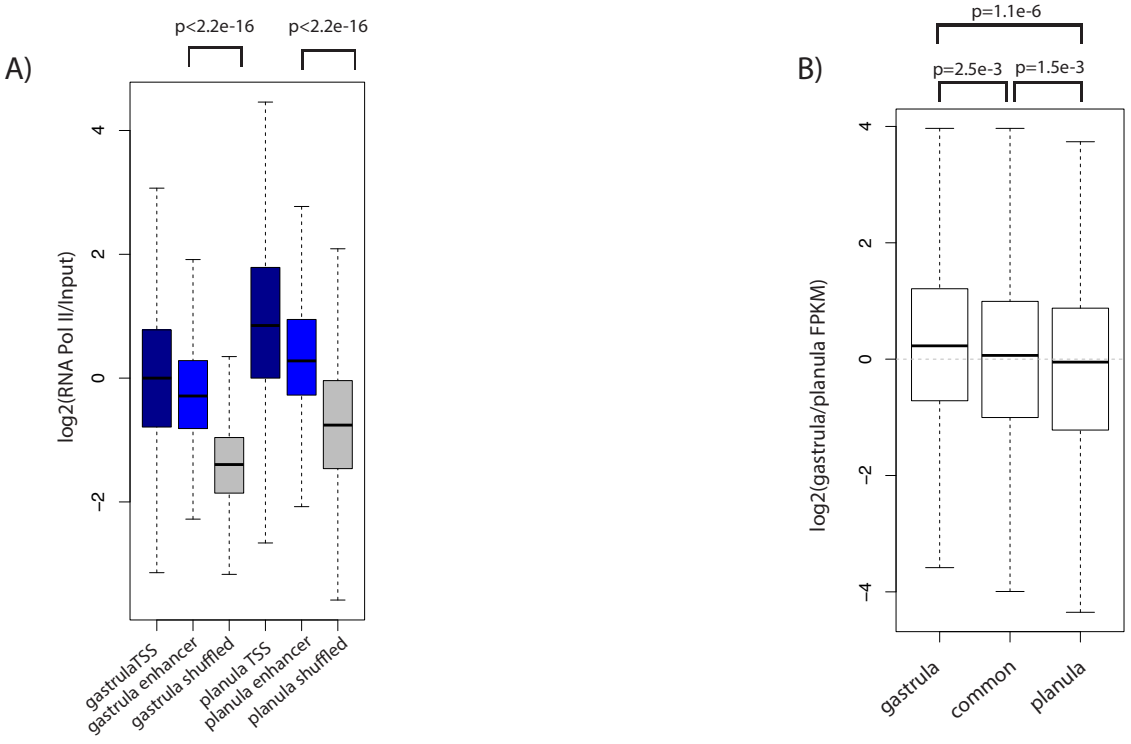


Supplemental Figure S7: Dynamics of chromatin modifciations and p300 peaks. A-B) We determined the overlap of p300 and H3K27ac peaks in gastrula (A) and planula(B) stage embryos. C) Overlap of p300 peaks in gastrula and planula. D-G) We devided all Nematostella (D-E) or Drosophila (F-G) genes into non-expressed genes (FPKM <1.5) and expressed genes (FPKM > 2). To define lowly, medium and highly expressed genes, all expressed genes were sorted based on FPKM values and seperated into three groups containing roughly equal numbers of genes. Expression values from gastrula (D) planula (E), 0-4h embryos (F) and L3 larva (G) were used. The y-axis shows the number of genes in each expression category that are (green) or are not (orange) associated with (= +/- 3kb around TSS) any H3K4me1 (top) or H3K27ac (bottom) peaks. H3K27ac is enriched around expressed genes. Drosophila modEncode data have been downloaded from http://www.modencode.org.

Schwaiger Supplemental Figure S8

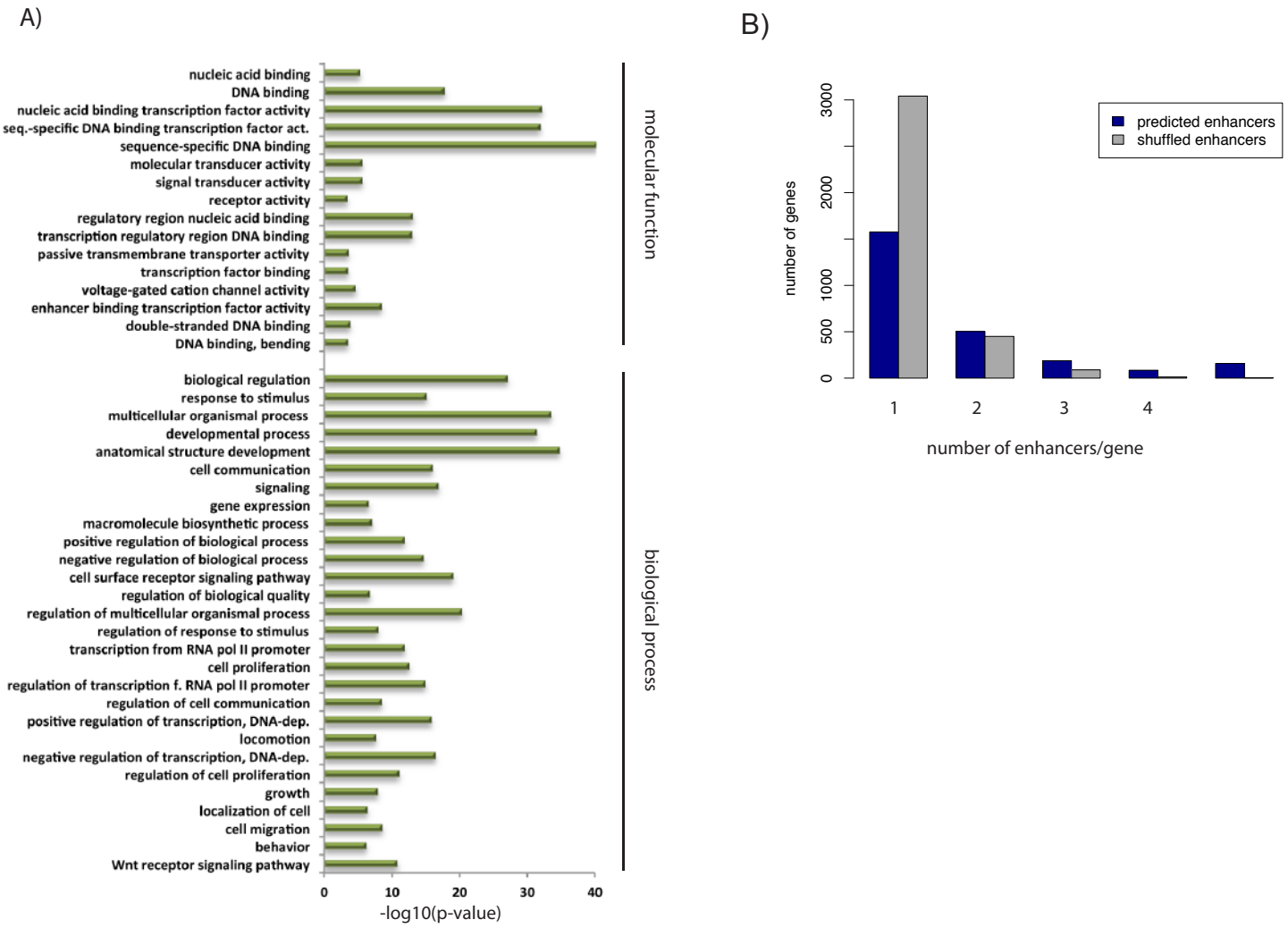


Supplemental Figure S8: Dynamics of enhancer modifiations between gastrula, planula, and adult female polyps. A-C) Pairwise correlations between H3K27ac (log2(bound/Input) in all predicted enhancer regions in gastrula and planula (A), gastrula and adult female polyp (B) or planula and adult female polyp (C). D-E) Pairwise correlation between H3K4me2 (log2(bound/Input) in all predicted enhancer regions in gastrula and planula (D), gastrula and adult female polyp (E) or planula and adult female polyp (F). The intensity of the blue color indicates the density of datapoints. Red points correspond to enhancers that are significantly (adj. p-value < 0.0001, > 2-fold change) different between the two stages. P-values were calculated using EdgeR (Robinson et al. 2010). Enhancer modifications correlate very well between the two embryonic stages, but show higher dynamics when compred to adult polyps. G) GO terms enriched among genes associated with enhancers that show higher H3K27ac in adults compared to planulae (positive log10 p-values) or higher H3K27ac in planulae compared to adults (negative log10 p-values).



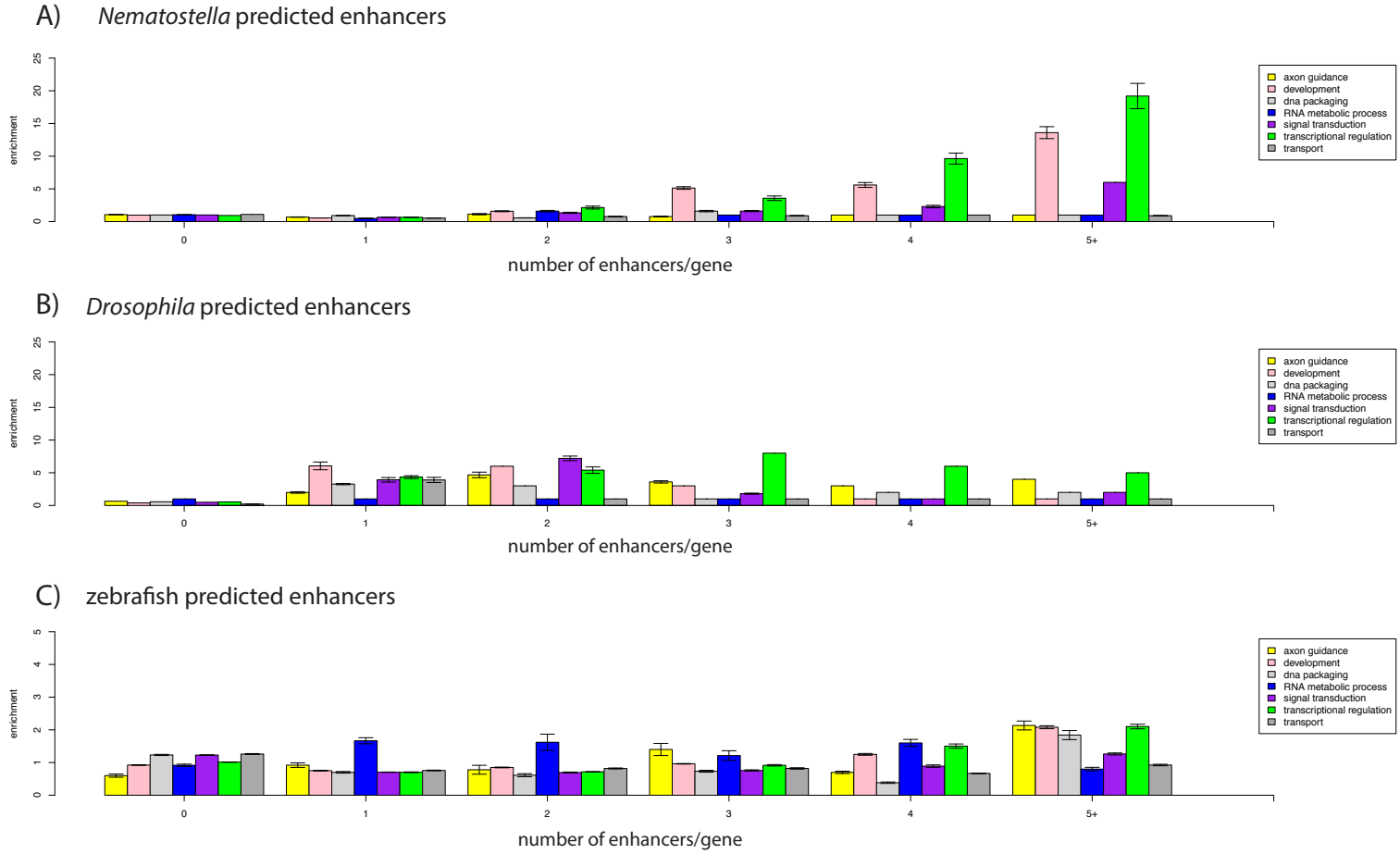
Supplemental Figure S9: Correlation of predicted enhancers with transcriptional activity. A) Enrichment of RNA Polymerase II (RNA Pol II, y-axis, calculated as number of reads divided by number of input reads) in TSSs of active genes (FPKM > 10), predicted enhancers in gastrulae and planulae (blue boxes) and regions of the same number and length as the predicted enhancers at random places in the genome, excluding TSSs ("shuffled enhancers", grey boxes). B) Boxplot showing the distribution of gene expression differences between gastrula and planula stage embryos (y-axis) for genes with gastrula specific enhancers (left), planula specific enhancers (right) and all remaining genes containing enhancers (middle). Genes with gastrula specific enhancers are expressed at higher levels in gastrula compared to the remaining enhancer proximal genes (p-value = 1.8e-4), genes with planula specific enhancers show higher expression levels in planula compared to all other enhancer neighboring genes (p-value = 1.1e-3), and also the gastrula and planula specific distributions differ significantly (p-value = 1.02e-06). Expression levels of genes in gastrula and planula have been quantile normalized, accordingly, a value of 0 means there was no change in gene expression of a given gene relative to all other genes. All p-values were calculated using the Wilcoxon rank sum test.

Schwaiger Supplemental Figure 10



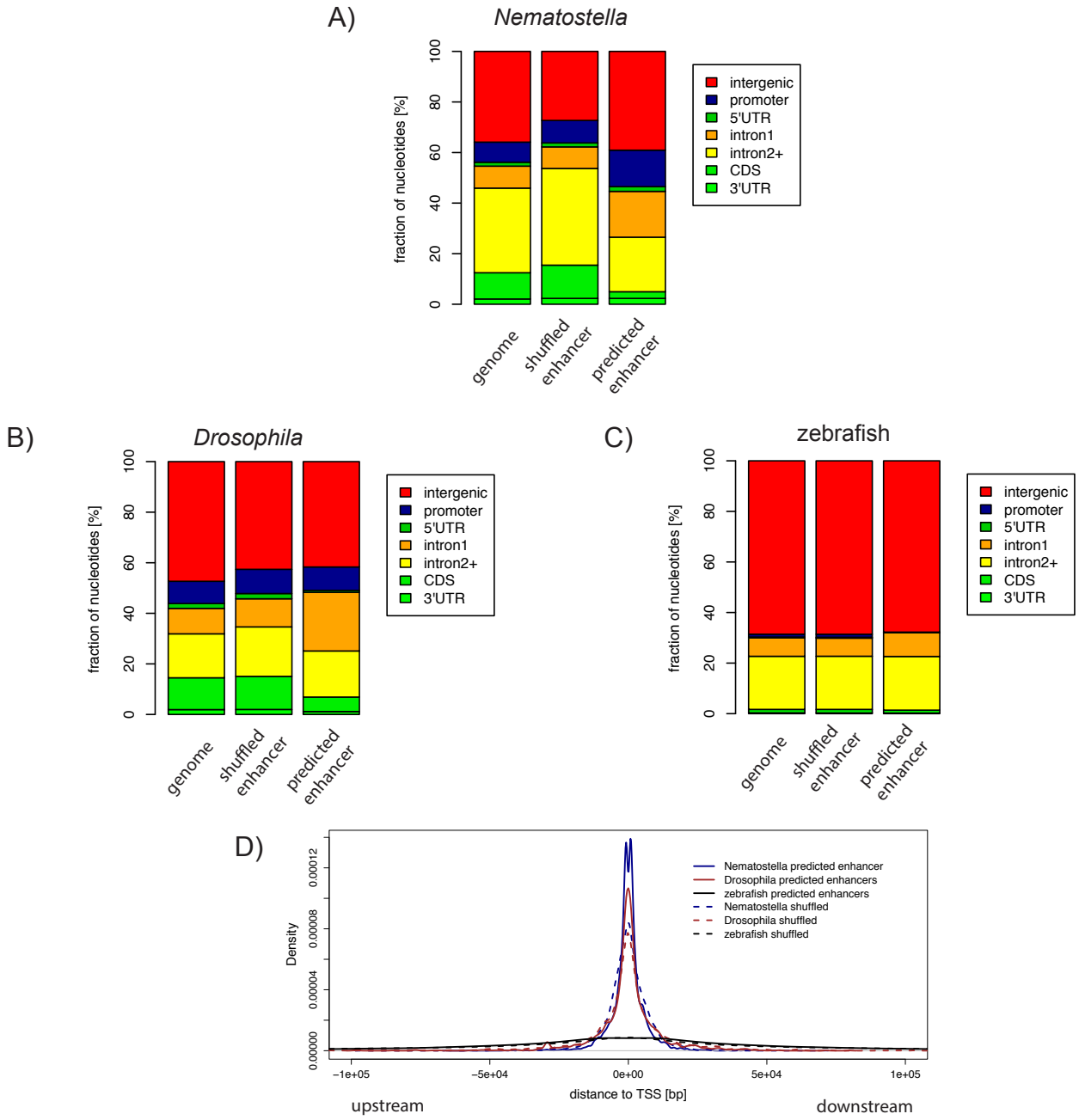
Supplemental Figure S10: GO terms enriched among Nematotella genes associated with enhancers. A) Gene ontology (GO) terms enriched among genes in proximity to predicted enhancers. GO terms are separated into two categories: Top: molecular function, bottom: biological process. The p-value of the GO term enrichment is indicated on the x-axis. B) The number of genes associated with 1,2,3,4, and 5 or more predicted enhancers in *Nematostella* is plotted for predicted enhancers (blue) and regions of the same number and length as the predicted enhancers at random places in the genome ("shuffled enhancers", grey). Predicted enhancers are clustered around genes more often than expected by chance based on shuffled regions of the same number and length as the predicted enhancers (p-value < 2.2e-16).

Schwaiger Supplemental Figure S11

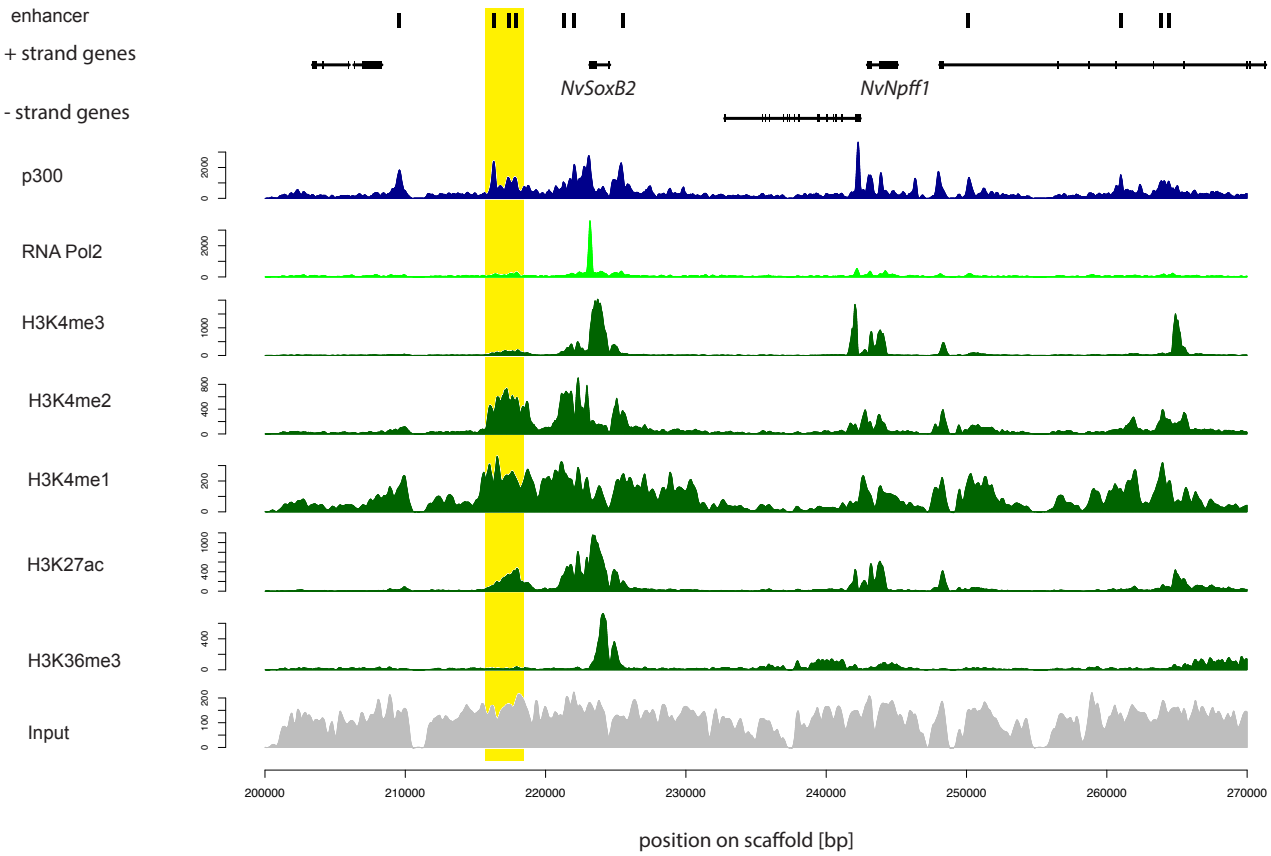


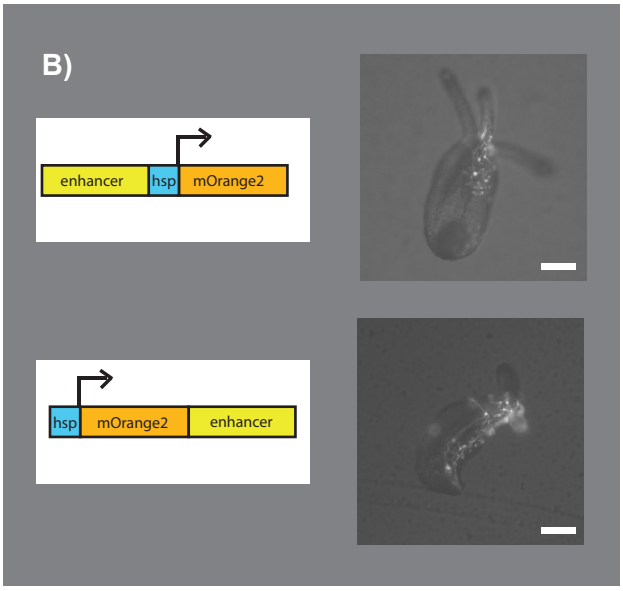
Supplemental Figure S11: Developmental regulatory genes are associated with multiple enhancers in *Nematostella*. The number of genes associated with 1,2,3,4, and 5 or more predicted enhancers in *Nematostella* (A), *Drosophila* (B) and zebrafish (C) is plotted for genes associated with different GO categories. The counts of genes with a given number of predicted enhancers have been normalized to the counts of genes associated with a given number of shuffled predicted enhancers (enrichment, y-axis). Error bars show the standard deviations of five different shuffling rounds. The following GO terms were used: axon guidance (75 genes in Nematostella, 607 genes in Drosophila, 22 genes in zebrafish), development (516 genes in Nematostella, 293 genes in Drosophila, 447 genes in zebrafish), DNA packaging (231 genes in Nematostella, 343 genes in Drosophila, 127 genes in zebrafish), RNA metabolic process packaging (31 genes in Nematostella, 67 genes in Drosophila, 17 genes in zebrafish), signal transduction packaging (430 genes in Nematostella, 637 genes in Drosophila, 921 genes in zebrafish), transcriptional regulation (783 genes in Nematostella, 1343 genes in Drosophila, 732 genes in zebrafish), transport (326 genes in Nematostella, 506 genes in Drosophila, 761 genes in zebrafish). In Nematostella, GO terms for transcriptional regulation, development, and signal transduction are strongly enriched among genes with multiple enhancers. In Drosophila, genes with multiple enhancers are associated with GO terms for transcriptional regulation and axon guidance. In zebrafish we could observe a weak enrichment of several GO categories, including transcriptional regulation. This only rather small enrichment is very likely due to the less stringent enhancer prediction method and the fact that it is more difficult to associate an enhancer with the correct target gene in the larger zebrafish genome.

Schwaiger Supplementary Figure S12



Supplemental Figure S12: Comparison of the genomic of predicted enhancers between *Nematostella*, *Drosophila*, and zebrafish. A-C) Distribution of predicted enhancer regions across genomic annotations. Green=exonic regions, yellow=non first introns, orange=first introns, blue=promoter regions (TSS-1kb), red=remaining intergenic regions. The y-axis indicates the fraction of nucleotides covered by each annotation in the genome (left bar), shuffled predicted enhancer regions (middle bar), and predicted enhancer regions (right bar) in *Nematostella* (A), *Drosophila* (B, (Nègre et al. 2012)) and zebrafish (C, (Bogdanovic et al. 2012)) embryos. Shuffling of predicted enhancers was performed by restricting the shuffled regions to lie on the same chromosome as the original regions. In *Nematostella*, shuffled regions were not allowed to fall into regions where no Input reads could be mapped to, thereby excluding genome sequence possibly representing bacterial contamination. D) Density plot showing the distances of predicted enhancers to the closest transcription start site (TSS). The distance to the TSS (x-axis) is negative if a region is located upstream of the TSS, and positive if it is downstream of the TSS relative to the orientation of transcription from that TSS. Blue=*Nematostella*, Brown=*Drosophila*, Black=zebrafish predicted enhancers. Dashed lines indicate the distribution of predicted enhancers shuffled to a random position on the same chromosome.





Supplemental Figure S13: A distal enhancer element driving reporter gene expression in neurons. A) Genomic region showing predicted enhancers (black bars), gene models, distribution of p300, RNA Pol II and several histone modifications in planulae. x-axis=position on the scaffold. y-axis=number of reads. The enhancer region that was tested in vivo is marked with a yellow background. The *NvSoxB2* and *NvNpff1* genes are indicated. B) mOrange2 signal of primary polyps injected with a construct where the predicted enhancer region was placed upstream (top) or downstream (bottom) of the *Nvhsp70a* minimal promoter driving mOrange2. The white scale bars represent 100µm. The predicted enhancers more than 5kb upstream of the *NvSoxB2* gene drive reporter gene expression in all developmental stages. In primary polyps, the fluorescent signal is clearly specific to neurons, as fluorescent lines representing the axons are visible between fluorescent cell bodies. This suggests that *Nematostella* enhancers can activate gene expression across large genomic distances.