Supplemental Tables

Supplemental Table S1: Histone modifications used in this study and their preferred genomic localization relative to genes and gene regulatory elements in bilaterian model organisms (reviewed in (Rando and Chang 2009)).

Modification	Localization
Trimethylation of Lysine 4 of Histone H3 (H3K4me3)	Promoter
Dimethylation of Lysine 4 of Histone H3 (H3K4me2)	Promoter, enhancer
Monomethylation of Lysine 4 of Histone H3 (H3K4me1)	Enhancer
Trimethylation of Lysine 36 of Histone H3 (H3K36me3)	Transcribed regions
Acetylation of Lysine 27 of Histone H3 (H3K27ac)	Promoter, enhancer

Supplemental Table S2: Number of mapped high quality reads and pearson correlation coefficients of biological replicates of ChIP experiments. We obtained very high pairwise correlations between biological replicates of ChIP experiments, with the exception of the RNA Pol II dataset in planula, where one replicate showed slightly better enrichments then the other. However, this does not affect the conclusions of the paper in any way.

Stage	Antibody	Mapped (q30) reads	Correlation coefficient
Adult female	H3K27ac	29802613	0.99
Gastrula	H3K27ac	41322872	0.92
Planula	H3K27ac	20756340	0.95
Adult female	H3K36me3	31997931	0.8
Gastrula	H3K36me3	116860476	0.94
Planula	H3K36me3	33636747	0.95
Gastrula	H3K4me1	26497612	0.9
Planula	H3K4me1	14785024	0.94
Adult female	H3K4me2	42144245	0.99
Gastrula	H3K4me2	47159402	0.98
Planula	H3K4me2	43215016	0.97
Adult female	H3K4me3	34080968	0.99
Gastrula	H3K4me3	43170348	0.99
Planula	H3K4me3	41259256	0.99
Gastrula	p300	23850207	0.83
Planula	p300	105787830	0.8
Gastrula	RNA Pol II	118020758	0.78

Planula	RNA Pol II	42192101	0.57

Supplemental Methods

Western Blot

We extracted proteins from *Nematostella vectensis* planula larvae using RIPA Buffer with 1mM DTT and Complete Protease Inhibitor (Roche). Extracts were run on a NuPAGE Novex Tris-Acetate 3-8% Gradient Gel (Invitrogen) and blotted according to manufacturer's instructions. Membranes were blocked in 5% milk in TBST and incubated with p300 antibodies (1:1000) o/n at 4 degrees. After incubation with a secondary anti-rabbit POD antibody (GE Healthcare, 1:10000), the signal was detected using ECL+ (GE Healthcare). For peptide competition experiments, the antibody was incubated with 5ug of p300 peptide in PBS for 30 minutes on ice prior to incubation with the membrane containing the protein extracts. p300 peptides were run on a 15% Polyacrylamide gel prior to blotting and detection as described for protein extracts.

Gene model and transcriptome datasets

We used the following gene model datasets for all analysis related to enhancer positions. *Danio rerio*: NCBI RefSeq (accessed 2013-03-25, genome version: danRer7), *Drosophila melanogaster*: FlyBase release 5.46 (genome version: dm3), *Nematostella vectensis*: NveGenes2.0 (genome version: NemVec1) (http://www.cnidariangenomes.org/, Fredman et al., in prep.). RNA-seq reads were mapped to the *Nematostella* genome using Tophat (Trapnell et al. 2009) and expression values (FPKM) were calculated based on polyA-selected RNAseq datasets for NveGenes2.0 using Cuffdiff 2 (Trapnell et al. 2012).

Clustering

K-means clustering of H3K4me3 datasets of 200bp windows at 100bp step-size across genes was performed using Cluster 3.0 (Eisen et al. 1998). Genes were aligned at their transcription start sites (from 2 kb upstream to 5 kb into the gene body) and only those, which were not overlapping with other genes, longer than 5kb and had good mappability (> 5000 simulated reads across the 7kb region) were considered.

Enhancer prediction

Peak finding on each p300 and RNA Polymerase II datasets was performed using the peakzilla software (https://github.com/steinmann/peakzilla), using Input reads as control. Peaks with > 2-fold enrichment over control and a score > 3 (p300) or > 5 (Pol2) were considered significant for each replicate, if they overlapped more than 2-

fold enriched peaks in the other replicate. Finally, the significant peaks of the two replicates were combined.

To define different chromatin states across the genome, we used ChromHMM (Ernst and Kellis 2012) to separate a combination of all five chromatin marks into several states. To this end we concatenated the datasets generated in gastrulae and planula larvae to train a Hidden Markov Model composed of 6 states. Input data were used as a background control, and the program was run with standard parameters. While all parameters of the model are learned directly from the data, the number of states has to be defined a priori. We tried different numbers of states (between four and twelve) and found that when using six states there was little redundancy between states yet still enough states to cover all possible genomic locations (promoter, enhancer, gene body) that we expected to resolve given the selection of chromatin modifications we used. Enhancers were predicted as significant p300 peaks, which did not overlap TSSs (no intersection with 300bp upstream or 300bp downstream of the TSS), overlapped a state 4 or 5 region (containing typical enhancer chromatin patterns) based on the ChromHMM analysis.

Zebrafish enhancers were taken from gastrula stage enhancer predictions based on histone modifications (Bogdanovic et al. 2012). *Drosophila melanogaster* enhancers were predicted based on p300/CBP datasets for *Drosophila* embryos (Nègre et al. 2012). To this end we used peakzilla to identify p300/CBP peaks in the *Drosophila* genome, and subtracted peaks +/- 300bp from the TSS, as described for *Nematostella*.

Definition of GO functional categories, transcription factors and house keeping genes

For comparison of the enhancer distribution between genes with different functional characteristics, we constructed groups based on related GO terms as follows: transcriptional_regulation (GO:0006355, GO:0006351, GO:0009889, GO:0010468,GO:0010556, GO:0019219, GO:0045893, GO:0010628, GO:0031323, GO:0031326, GO:0032774, GO:0045944, GO:0051171, GO:0051252, GO:0080090, GO:2000112, GO:2001141, GO:0032583, GO:0045449, GO:0061019), development (GO:0007275, GO:0009790, GO:0009792, GO:0030154, GO:0043009, GO:0048513, GO:0048731, GO:0048856, GO:0032502, GO:0007389, GO:0045165), RNA_metabolic_process (GO:0016070), axon_guidance (GO:0007411), transport (GO:0006810), signal_transduction (GO:0007165), dna_packaging (GO:0006323, GO:0034728, GO:0031497, GO:0043933, GO:0043933, GO:0051276, GO:0006333, GO:0006334, GO:0006325, GO:0022607, GO:0006996, GO:0044085, GO:0034622, GO:0065004, GO:0065003). Transcription factor genes in *Nematostella* were identified

based on protein domains characteristic for transcription factors and supported by gene ontology terms. House keeping genes were defined as genes with functions "Signal transduction, Metabolic process, or Transport", and a temporal specificity of gene expression < 0.3 (based on RNA-seq of 7 developmental stages, Fredman et al., in prep.).

Alignment-free sequence similarity analysis

To test the similarity of predicted enhancers in different species, we calculated a correlation coefficient based on 6mer content (without allowing for mismatches) between sequences using N2 (Göke et al. 2012).

Supplemental References

Eisen MB, Eisen MB, Eisen MB, Spellman PT, Spellman PT, Spellman PT, Brown PO, Brown PO, Brown PO, Botstein D, et al. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**: 14863–14868. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9843981&retmode=ref&cmd=prlinks.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nature Publishing Group* **9**: 215–216.

Göke J, Schulz MH, Lasserre J, Vingron M. 2012. Estimation of pairwise sequence similarity of mammalian enhancers with word neighbourhood counts. *Bioinformatics* **28**: 656–663.

Rando OJ, Chang HY. 2009. Genome-Wide Views of Chromatin Structure. *Annu Rev Biochem* **78**: 245–271.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2012. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* **31**: 46–53.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.