frontiers in Genetics

HYPOTHESIS AND THEORY
published: 23 May 2017
doi: 10.3389/fgene.2017.00063

An Enhancer's Length and Composition Are Shaped by Its Regulatory Task

Lily Li and Zeba Wunderlich *

Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA, United States

Enhancers drive the gene expression patterns required for virtually every process in metazoans. We propose that enhancer length and transcription factor (TF) binding site composition—the number and identity of TF binding sites—reflect the complexity of the enhancer's regulatory task. In development, we define regulatory task complexity as the number of fates specified in a set of cells at once. We hypothesize that enhancers with more complex regulatory tasks will be longer, with more, but less specific, TF binding sites. Larger numbers of binding sites can be arranged in more ways, allowing enhancers to drive many distinct expression patterns, and therefore cell fates, using a finite number of TF inputs. We compare ∼100 enhancers patterning the more complex anterior-posterior (AP) axis and the simpler dorsal-ventral (DV) axis in *Drosophila* and find that the AP enhancers are longer with more, but less specific binding sites than the (DV) enhancers. Using a set of ∼3,500 enhancers, we find enhancer length and TF binding site number again increase with increasing regulatory task complexity. Therefore, to be broadly applicable, computational tools to study enhancers must account for differences in regulatory task.

Keywords: enhancer, transcription factor, embryogenesis, *Drosophila melanogaster*, gene regulation

OPEN ACCESS

Edited by:
Shihua Zhang,
Academy of Mathematics and Systems Science (CAS), China

Reviewed by:
Ka-Chun Wong,
City University of Hong Kong, Hong Kong
Zhi-Ping Liu,
Shandong University, China

***Correspondence:**
Zeba Wunderlich
zeba@uci.edu

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *17 March 2017*
Accepted: *08 May 2017*
Published: *23 May 2017*

Citation:
Li L and Wunderlich Z (2017) An Enhancer's Length and Composition Are Shaped by Its Regulatory Task. Front. Genet. 8:63. doi: 10.3389/fgene.2017.00063

INTRODUCTION

Nearly every aspect of an organism, from its development to its immune response, is dependent on precise spatiotemporal control of gene expression. This control is mediated by the binding of transcription factor (TF) activators and repressors to stretches of regulatory DNA called enhancers.

Given their role in diverse biological processes, it is not surprising that enhancers vary widely in architecture—length, number of TF binding sites, and the average binding specificity of the TFs that bind them. Enhancers can be ∼10–1,000 bps long, with a couple to tens of TF binding sites (Blackwood and Kadonaga, 1998; Yáñez-Cuna et al., 2013). Several theories have been put forth to explain why enhancers are built so differently. For example, differences in evolutionary pressures and TF cooperativity are invoked to explain why many developmental enhancers are robust to rearrangements of TF binding sites within them while some immune-responsive enhancers are intolerant to even point mutations (Thanos and Maniatis, 1995; Kim and Maniatis, 1997; Munshi et al., 2001; Arnosti and Kulkarni, 2005).

Although enhancers vary in architecture, some constraints apply to all enhancers, e.g., an enhancer's need to be distinguishable from the rest of the genome. Because eukaryotic TFs are highly degenerate, TF binding sites litter the genome, and an enhancer can only achieve

distinguishability if it consists of a cluster of TF binding sites within a short distance (Wasserman and Fickett, 1998; Berman et al., 2002; Frith et al., 2002; Halfon et al., 2002; Markstein et al., 2002; Rebeiz et al., 2002; Li et al., 2007; Wunderlich and Mirny, 2009; Hardison and Taylor, 2012). Enhancer length, number of TF binding sites, and average specificity of TFs binding an enhancer can be combined in different ways to achieve distinguishability. For example, an enhancer with higher average TF specificity requires fewer TF binding sites than one with lower average TF specificity to be distinguishable from the genomic background.

We propose that the complexity of an enhancer's "regulatory task"—the process that it controls—is one force that shapes enhancer architecture. In development, task complexity can be defined as the number of cell fates being specified in a set of roughly homogeneous cells at one time. When a cell can be driven to one of many cell fates, the task complexity is high; when a cell is making binary decisions between cell fates, the task complexity is low (**Figure 1A**). Since cell fate is largely specified by gene expression patterns, the more cell fates being specified, the more distinct expression patterns are needed. To accommodate this need using a limited set of TFs, these enhancers need to contain a larger number of TF binding sites, which allow for more rearrangements and, presumably, more expression patterns. Thus, we propose that enhancers with more binding sites can accommodate higher task complexity. Though intuitive, this proposal has never been verified systematically.

To evaluate this hypothesis, we characterize two sets of enhancers in *Drosophila melanogaster* and analyze the correlation between regulatory task complexity and enhancer architecture. In a set of ∼100 early embryonic enhancers, those that pattern the more complex anterior-posterior (AP) axis are longer, have more binding sites, and have lower average TF specificity compared to those patterning the simpler dorsal-ventral (DV) axis. In a set of ∼3,500 enhancers active throughout embryogenesis, we find enhancers active early are longer and have more binding sites than those active late, reflecting the general trend that task complexity decreases with developmental time. We conclude that the complexity of an enhancer's regulatory task is one of many forces shaping its architecture.

RESULTS

To understand the properties required for an enhancer to be distinguishable from the genomic background, we calculate the probability of finding an enhancer with a particular length, number of TF binding sites, and average TF binding specificity (Wunderlich and Mirny, 2009) (Supplementary Material within). As a proxy for TF binding specificity, we use p, the probability of finding a "hit" or match to the TF binding motif in the genomic background (see Materials and Methods). Note that a larger p corresponds to a lower binding specificity. The probability of finding an enhancer of length w, with k TF binding sites is:

$$P(k) = \binom{w}{k} p^k (1-p)^{w-k}$$

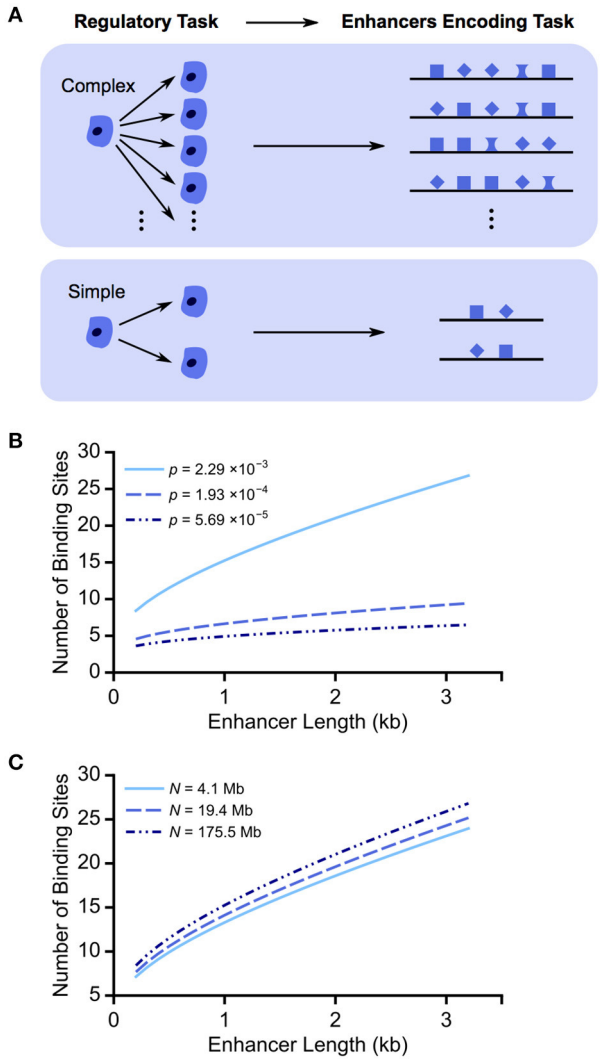

FIGURE 1 | Regulatory task complexity can shape enhancer length and binding site composition. (A) We propose that more complex regulatory tasks, e.g., cell patterning decisions, are associated with longer enhancers with more binding sites. More binding sites can be arranged within an enhancer in more ways, allowing for the specification of a wider variety of expression patterns and, therefore, more complex tasks. **(B)** We plot the minimum number of TF binding sites required for enhancers of varying lengths to achieve distinguishability from the genomic background. We show the results for three motif hit probabilities, corresponding to the median, first and third quartiles of *Drosophila* TF binding specificities. As motif hit probability p decreases from ∼2 in 1 kb (2×10^{-3}) to ∼6 in 100 kb (6×10^{-5}), an enhancer of the same length requires fewer binding sites to be distinguishable from the background. **(C)** To test the effect of genome accessibility, we plot the minimum number of TF binding sites required for enhancers of varying length in the context of different accessible genome sizes (N). Varying the accessible regions of the genome has a minor impact on the trend of numbers of TF binding sites increasing with enhancer length.

To achieve distinguishability, $P(k)$ must be less than $1/N$, where N is the genome size accessible for TF binding. Thus, the number of required binding sites increases with enhancer length and motif hit probability (**Figure 1B**). Considering the median, first and third quartiles of all *Drosophila* TF binding specificities, the

corresponding number of TF binding sites required in a 1 kb enhancer decreases from 16 to 7 to 5 as TF binding specificity increases (or motif hit probability decreases).

To take into account the compaction of the genome, we consider different values of N. We use DNase I hypersensitivity profiles to estimate the accessible regions (Thomas et al., 2011). Whether we use a conservative estimate of accessible regions during development (4.1 Mb), a more relaxed estimate (19.4 Mb), or the entire genome (175.5 Mb) (Thomas et al., 2011; Ellis et al., 2014), the same trends are seen (**Figure 1C**). For a 1 kb enhancer with binding sites for a TF with relatively low binding specificity, $p = 2 \times 10^{-3}$, the number of required binding sites increases from 13 to 15 as N increases from 4.1 to 175.5 Mb, and thus the number of required binding sites is only weakly dependent on accessible genome size.

To test whether task complexity shapes the characteristics of enhancer architecture, we need a set of enhancers that drive regulatory tasks of different complexities and knowledge of the transcription factors (TFs) that regulate them. The *Drosophila* embryonic AP and DV patterning systems neatly fit these criteria. The AP axis is more complex than the DV axis, with the AP axis consisting of 14 parasegments (Nasiadka et al., 2002) and the DV axis consisting of six germ layers and sublayers (Levine and Davidson, 2005), and therefore the patterning of the AP axis requires enhancers that drive more unique gene expression patterns. Years of work from many groups have identified ∼40 principal TFs (Fowlkes et al., 2008; MacArthur et al., 2009) whose binding to ∼100 characterized enhancers (Papatsenko et al., 2009) drives AP and DV patterning.

To identify the TF binding sites within these enhancers, we use a computational approach. Though ChIP can experimentally identify TF-bound regions, existing data sets in the *Drosophila* embryo are low resolution, with ∼100 base pair peaks (Li et al., 2008; MacArthur et al., 2009; Roy et al., 2010), which are longer than the ∼10 bp TF binding sites (Zhu et al., 2011). Therefore, we predict TF binding sites using experimentally measured binding motifs (Stormo, 2015). To select a threshold above which a sequence is deemed a "true" binding site, we develop a principled approach, scoring the aligned sequences used to create the motifs and setting a threshold such that 75% of these aligned sequences are predicted as "true" (see Materials and Methods).

We analyze 60 AP and 39 DV enhancers, identifying binding sites for 24 AP and 10 DV TFs. Consistent with our predictions, AP enhancers (median length = 1.3 kb) are longer than DV enhancers (median = 0.8 kb; **Figure 2A**; $p = 4.5 \times 10^{-4}$; Mann Whitney rank-sum test). AP enhancers also have a larger number of TF binding sites (median = 47) than DV enhancers (median = 9; **Figure 2B**; $p = 1.5 \times 10^{-13}$; Mann Whitney rank-sum test). To ensure that the difference is not due to the larger number of AP TFs, we also calculated the number of TF binding sites per enhancer, normalized by the number of TF motifs used to search the enhancer, and find the difference holds (AP median = 2.0, DV median = 0.9; **Figure 2C**; $p = 6.1 \times 10^{-6}$; Mann Whitney rank-sum test). AP enhancers also have a higher average motif hit probability (AP median = 4.9×10^{-3}, DV median = 3.0×10^{-3}; **Figure 2E**; $p = 7.5 \times 10^{-9}$; Mann Whitney rank-sum test), which is a result of differential binding rather than the TFs considered, as the specificity of AP and DV TFs have a similar distribution (**Figure 2D**, $p = 0.247$; Mann Whitney rank-sum test). This difference is likely driven by the fact that the key TFs that act as morphogens for these axes show markedly different binding specificities, with the key AP axis TFs having low binding specificities and the key DV axis TF having high binding specificity (Figure S1). In summary, we find that the enhancers that encode the lower complexity task of specifying the DV axis are composed of fewer binding sites, as predicted by our hypothesis. The DV enhancers require fewer binding sites because they are both shorter and use more specific TFs than the AP enhancers. Our hypothesis does not require that both enhancer length and motif specificity both differ, though in this case they do.

To determine whether these tradeoffs in enhancer architecture apply to a larger, if less well-characterized dataset, we analyze the Vienna Tile enhancers (Kvon et al., 2014), which drive expression throughout *Drosophila* embryogenesis. To produce this dataset, the Stark lab measured the expression patterns driven by 7,705 enhancer candidates and found 4,480 enhancers that were active during development. Of these enhancers, we consider the 3,580 enhancer candidates that were successfully refined to the putative minimal enhancers using functional genomics (Kvon et al., 2014). To determine the relevant TFs, we match the stages of the active enhancers with the concurrently expressed TFs (Tomancak et al., 2002, 2007; Hammonds et al., 2013).

We assume that as development progresses, the task complexity decreases, approaching binary decisions between two cell fates. We found that enhancer length monotonically decreases over development (**Figure 3A**; Table S1). While stages 4–6 and 7–8, and stages 9–10 and 11–12, have very similar length distributions for active enhancers ($p = 1$, $p = 0.1$, respectively; Mann Whitney rank-sum test), all other intervals have significantly different distributions of enhancer length (**Figure 3D**, $p < 0.05$; Mann-Whitney rank sum test with Bonferroni correction applied). Number of TF binding sites and average motif hit probability, in contrast, do not show a clear trend (**Figures 3B,E**). However, there is a large increase in the number of TFs expressed in the final two time intervals (see Table S1), and when the number of binding sites is normalized by the number of binding motifs used to search the enhancer, the binding site trend mirrors the enhancer length trend (**Figures 3C,F**). We also verified that these trends were not unduly influenced by enhancers driving ubiquitous expression patterns (Figure S2, Table S2). Thus, decreasing complexity again is associated with decreasing enhancer length and with decreasing TF binding site number, when normalized appropriately. In this case, there is no clear trend with regards to motif specificity, which, as we note above, is not necessarily at odds with our hypothesis.

DISCUSSION

We hypothesize that an enhancer's regulatory task complexity shapes its architecture. In the case of *Drosophila* axis patterning, the AP axis has higher task complexity than the DV axis, and accordingly, enhancer length, number of TF binding sites, and average motif hit probability increase with task complexity. In the case of *Drosophila* embryogenesis, where we posit that task

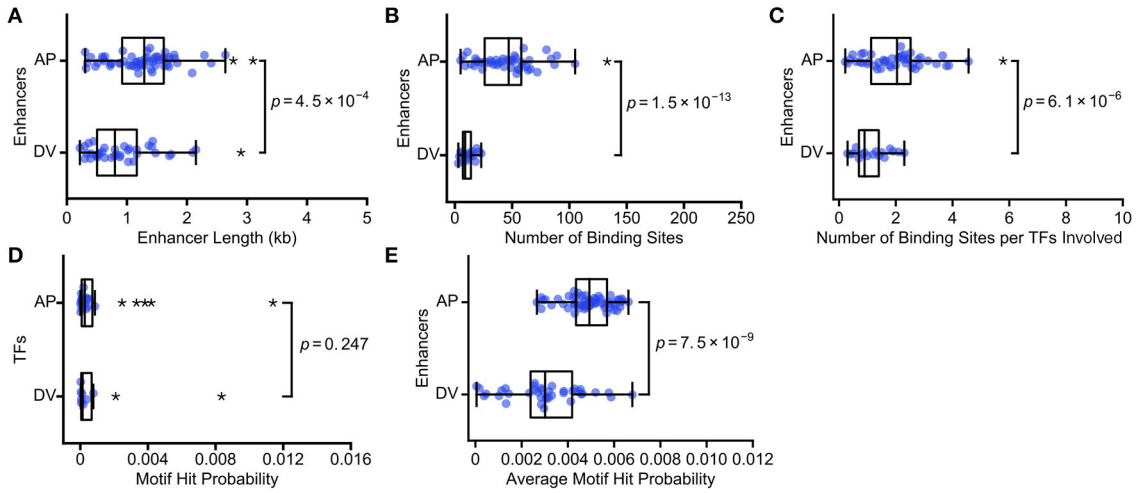

FIGURE 2 | The more complex AP axis is patterned by enhancers with more TF binding sites. We show the scatterplots and associated boxplots of **(A)** the length of AP and DV enhancers, **(B)** the number of TF binding sites predicted in AP and DV enhancers, **(C)** the number of TF binding sites normalized by the number of TFs involved, **(D)** the motif hit probability of TFs involved in AP and DV patterning, and **(E)** the average motif hit probability of AP and DV enhancers. These data are consistent with our hypothesis that enhancers carrying out more complex regulatory tasks will have more binding sites, in this case because AP enhancers are both longer and have lower average TF binding specificity. In all box plots, the boxes indicate the lower and upper quartiles, with the line within the box indicating the median. Whiskers extend to 1.5*IQR (interquartile range) plus or minus the upper and lower quartile, respectively, and the stars indicate outliers that fall outside the whiskers. *P*-values from Mann-Whitney rank tests are shown.

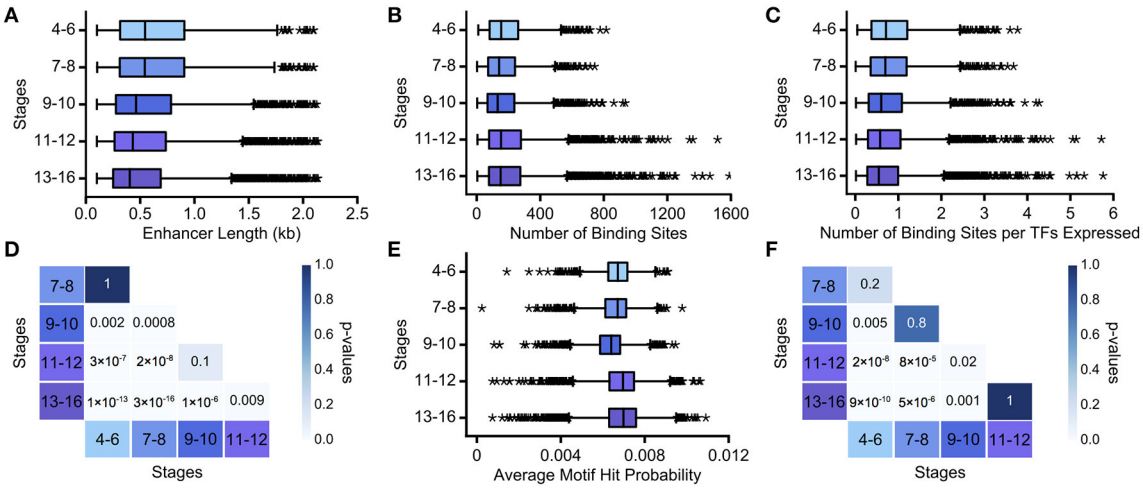

FIGURE 3 | Decreasing regulatory task complexity over embryogenesis is associated with decreasing enhancer length. We show boxplots of **(A)** the length of minimal Vienna Tile enhancers, **(B)** the number of TF binding sites predicted in minimal Vienna Tile enhancers, **(C)** the number of TF binding sites predicted in minimal Vienna Tile enhancers normalized by TFs concurrently expressed, and **(E)** the average motif hit probability of minimal Vienna Tile enhancers over developmental stages 4–16. The heatmaps display the Bonferroni-adjusted *p*-values from the Mann-Whitney rank test between **(D)** pairwise distributions of Vienna Tile enhancer length and between **(F)** pairwise distributions of the number of TF binding sites predicted in Vienna Tile enhancers per TFs concurrently expressed. In all box plots, the boxes indicate the lower and upper quartiles, with the line within the box indicating the median. Whiskers extend to 1.5*IQR plus or minus the upper and lower quartile, respectively, and the stars indicate outliers that fall outside the whiskers.

complexity decreases over time, enhancer length and binding site number decrease accordingly.

Though the well-characterized *Drosophila* axis patterning systems are ideal for studying how an enhancer's regulatory tasks shape its design (Markstein et al., 2002; Papatsenko et al., 2002; Lifanov et al., 2003; Papatsenko and Levine, 2005; Halfon et al., 2008), the systems still have limitations. For example, autoregulatory enhancers like ftz_up, ftz_zebra, and gt_minus1 (Hiromi et al., 1985; Hiromi and Gehring, 1987; Hoermann et al., 2016) have lower task complexity because they reinforce the expression patterns determined by other enhancers, and therefore, may not be consistent with

the observed trends. However, we find these autoregulatory enhancers have parameters that generally fall within the bulk of the distribution. In addition, the enhancer boundaries in this dataset were determined one-at-a-time. However, in the Vienna Tile enhancer set, in which boundaries are determined in a uniform manner, we still find that enhancer length decreases with developmental time and regulatory task complexity.

In contrast to the axis patterning data set, choosing principal TFs for the Vienna Tile enhancers is challenging because there is no consistent annotation of the expression patterns of TFs and enhancers. We match stage-specific expression of enhancers and TFs without considering tissue-specific expression, which impacts both the number of TF binding sites and the average motif hit probability and undoubtedly obscures the clarity of those trends.

We expect that there are many other forces shaping enhancer architecture, like protein-protein interactions between TFs or between TFs and cofactors, and therefore do not expect that regulatory task complexity alone can explain enhancer architecture. Additionally, a particular TF may be employed in an enhancer because it is expressed in the right place at the right time, and not because of its TF binding specificity, though an analysis of the binding specificities of the TFs encoded in the *Drosophila* genome shows that there is a wide distribution of TF specificities that is relatively independent of developmental stage (Figures S3–S6). However, we can make educated guesses about the ways that enhancer architecture may vary depending on regulatory task and use this information to improve our ability to predict and design putative enhancers. As increasingly large sets of enhancers are identified in a variety of biological settings, we will undoubtedly uncover other forces impacting why an enhancer is built in a particular way.

MATERIALS AND METHODS

Datasets Used in this Study

The 60 AP and 39 DV patterning enhancers were collected by Papatsenko et al. (2009) and provided here as Supplementary Data Sheet 1, and in this dataset, we considered the binding of 33 early patterning TFs (Fowlkes et al., 2008; MacArthur et al., 2009). The Vienna Tile enhancer project tested the activity of 7,705 potential enhancers (Kvon et al., 2014). In this analysis, we considered the 3,580 enhancer candidates that were active during embryogenesis and whose boundaries had been refined using DHS regions or CBP/P300-bound and H3K4me1-marked regions. Here, we analyzed the binding of TFs that were concurrently expressed with active enhancers based on the Berkeley *Drosophila* Genome Project *in situ* annotations (Tomancak et al., 2002, 2007; Hammonds et al., 2013). These TF lists are available in Supplementary Data Sheets 2, 3. 51.7% of the AP enhancers and 30.8% of the DV enhancers at least partially overlapped with the Vienna Tile enhancers. The five completely overlapping enhancers showed expression in the same stage (i.e., stages 4–6), except for the DV enhancers pnr and rho (Table S3). Outside of those two, the greatest amount of overlap that did not result in expression at the same stage was 60.5%.

Transcription Factor Binding Site Prediction

Transcription factor (TF) binding sites were computationally predicted using Patser (Hertz and Stormo, 1999) with the position weight matrices (PWMs) from FlyFactor Survey (Zhu et al., 2011). Pseudocounts were added to each element in the PWM in proportion to the intergenic frequency of the corresponding base to a total of 0.01. For those TFs with more than one PWM, the PWM derived from the largest set of aligned sequences was used, except in the case of giant and daughterless.

As no TF binding sites were identified in the set of AP enhancers when using the giant PWM selected using the previous criteria, a switch was made to another available giant PWM in FlyFactor Survey with which binding sites could be predicted. In the case of daughterless, which often binds as a heterodimer and has been identified as one of the key TFs in DV patterning, the PWM that was created using only daughterless and not any heterodimeric partners was used.

For the early patterning dataset, 33 of the 37 TFs (Fowlkes et al., 2008; MacArthur et al., 2009) determined to be principal regulatory factors for AP and DV patterning had available PWMs and were used. PWMs were not available for croc, Stat92E, tsh, or Dad. For the Vienna Tile dataset, a total of 292 TFs that had available PWMs and were expressed during embryogenesis were used.

To determine ln (p-value) cutoffs in a systematic manner, the aligned sequences from which the PWMs are derived are scored by Patser, and a 75th percentile ln (p-value) was chosen as a cutoff such that 75% of the aligned sequences are considered "true" binding sites. Cutoffs at multiple percentiles were considered, but the overall trends for relative numbers of putative TF binding sites identified remained constant regardless of the chosen cutoff (Figure S7).

Some PWMs were generated from DNase I footprints curated in the FlyReg database (Halfon et al., 2008); aligned versions of these footprints were not directly provided by FlyFactor Survey. The raw sequences were retrieved from FlyReg v2.0 and were aligned when possible. Note that Patser can only score sequences that are the same length or longer than the PWM, so some sequences used to create the PWM have been omitted when determining the percentile cutoffs.

Evaluating Transcription Factor Specificity

Information content is a measure of TF specificity. To measure information in a motif, we calculated the Kullback-Leibler distance (Schneider et al., 1986; Stormo and Fields, 1998) between the motif and the composition of the intergenic regions of the genome

$$I = \sum_{i=1}^{L} \sum_{b \in \{A,C,G,T\}} p_i(b) \, log_2 \frac{p_i(b)}{q(b)}$$

where L is the length of the motif, $p_i(b)$ is the frequency of base b at position i in the motif, and $q(b)$ is the frequency of base b in the intergenic regions of the genome. Note that $p = 2^{-I}$ is roughly the probability of a motif hit in the genome for a TF of information

content I (Berg and von Hippel, 1987). For a set of TF binding sites in an enhancer, an average motif hit probability

$$p_{av} = \frac{\sum_j n_j 2^{-I_j}}{n_{\text{total}}}$$

is calculated, where n_j = number of binding sites for TF j, I_j = information content of TF j, and n_{total} = the total number of TF binding sites in a particular enhancer. If a cluster is composed of sites of m different TFs with identical motif hit probability p, the probability of finding a cluster of k binding sites within w bps is

$$P(k) = \binom{w}{k} (mp)^k (1 - mp)^{w-k}$$

Therefore, to characterize the average specificity of TFs employed in a specific enhancer, we choose to compute the average motif hit probability p, as opposed to the average information content I.

Quantitation and Statistical Analysis

Mann-Whitney rank tests were performed to compare all distributions, and the p-values were reported. The Mann-Whitney test was chosen because it does not require the assumption that the distributions to be compared are normally distributed. When multiple comparisons were made, the Bonferroni correction was applied.

Data and Software Availability

Python code for enhancer architecture analysis is available at GitHub: https://github.com/WunderlichLab/Info_Content.

The axis patterning enhancers originally collected by Papatsenko et al. are available as supplemental data file Supplementary Data 1 (Papatsenko et al., 2009).

AUTHOR CONTRIBUTIONS

LL and ZW conceived of the study. LL carried out the analysis, and LL and ZW wrote the paper.

FUNDING

This work is supported by NIH grants R00HD073191 (to ZW) and T32EB009418 (to LL).

ACKNOWLEDGMENTS

We thank Ali Mortazavi and Rahul Warrior for useful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2017.00063/full#supplementary-material

REFERENCES

Arnosti, D. N., and Kulkarni, M. M. (2005). Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.* 94, 890–898. doi: 10.1002/jcb.20352

Berg, O. G., and von Hippel, P. H. (1987). Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* 193, 723–50. doi: 10.1016/0022-2836(87)90354-8

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., et al. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila genome*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 757–762. doi: 10.1073/pnas.231608898

Blackwood, E. M., and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* 281, 60–63. doi: 10.1126/science.281.5373.60

Ellis, L. L., Huang, W., Quinn, A. M., Ahuja, A., Alfrejd, B., Gomez, F. E., et al. (2014). Intrapopulation genome size variation in *D. Melanogaster* reflects life history variation and plasticity. *PLoS Genet.* 10:4522. doi: 10.1371/journal.pgen.1004522

Fowlkes, C. C., Hendriks, C. L. L., Keränen, S. V. E., Weber, G. H., Rübel, O., Huang, M.-Y. Y., et al. (2008). A quantitative spatiotemporal atlas of gene expression in the drosophila blastoderm. *Cell* 133, 364–374. doi: 10.1016/j.cell.2008.01.053

Frith, M. C., Spouge, J. L., Hansen, U., and Weng, Z. P. (2002). Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences. *Nucleic Acids Res.* 30, 3214–3224. doi: 10.1093/nar/gkf438

Halfon, M. S., Gallo, S. M., and Bergman, C. M. (2008). REDfly 2.0: an integrated database of cis-regulatory modules and transcription factor binding sites in Drosophila. *Nucleic Acids Res.* 36, D594–D598. doi: 10.1093/nar/gkm876

Halfon, M. S., Grad, Y., Church, G. M., and Michelson, A. M. (2002). Computation-based discovery of related transcriptional regulatory modules and motifs using a experimentally validated combinatorial model. *Genome Res.* 12, 1019–1028. doi: 10.1101/gr.228902

Hammonds, A. A. S., Bristow, C. C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., et al. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biol.* 14:R140. doi: 10.1186/gb-2013-14-12-r140

Hardison, R. C., and Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* 13, 469–483. doi: 10.1038/nrg3242

Hertz, G. Z., and Stormo, G. D. (1999). Identifying, DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* 15, 563–577. doi: 10.1093/bioinformatics/15.7.563

Hiromi, Y., and Gehring, W. J. (1987). Regulation and function of the *Drosophila segmentation* gene fushi tarazu. *Cell* 50, 963–974. doi: 10.1016/0092-8674(87)90523-X

Hiromi, Y., Kuroiwa, A., and Gehring, W. J. (1985). Control elements of the *Drosophila segmentation* gene fushi tarazu. *Cell* 43, 603–613. doi: 10.1016/0092-8674(85)90232-6

Hoermann, A., Cicin-Sain, D., and Jaeger, J. (2016). A quantitative validated model reveals two phases of transcriptional regulation for the gap gene giant in *Drosophila*. *Dev. Biol.* 411, 325–338. doi: 10.1016/j.ydbio.2016.01.005

Kim, T. K., and Maniatis, T. (1997). The mechanism of transcriptional synergy of an *in vitro* assembled interferon-beta enhanceosome. *Mol. Cell* 1, 119–129. doi: 10.1016/S1097-2765(00)80013-1

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., et al. (2014). Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* 512, 91–95. doi: 10.1038/nature13395

Levine, M., and Davidson, E. H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4936–4942. doi: 10.1073/pnas.0408031102

Li, L., Zhu, Q., He, X., Sinha, S., and Halfon, M. S. (2007). Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol.* 8:R101. doi: 10.1186/gb-2007-8-6-r101

Li, X., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila blastoderm*. *PLoS Biol.* 6:e27. doi: 10.1371/journal.pbio.0060027

Lifanov, A. P., Makeev, V. J., Nazina, A. G., and Papatsenko, D. A. (2003). Homotypic regulatory clusters in *Drosophila*. *Genome Res.* 13, 579–588. doi: 10.1101/gr.668403

MacArthur, S., Li, X.-Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., et al. (2009). Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 10:R80. doi: 10.1186/gb-2009-10-7-r80

Markstein, M., Markstein, P., Markstein, V., and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila embryo*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 763–768. doi: 10.1073/pnas.012591199

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G., and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. *Science* 293, 1133–1136. doi: 10.1126/science.293.5532.1133

Nasiadka, A., Dietrich, B. H., and Krause, H. M. (2002). Anterior-posterior patterning in the *Drosophila embryo*. *Adv. Dev. Biol. Biochem.* 12, 155–204. doi: 10.1016/S1569-1799(02)12027-2

Papatsenko, D. A., Makeev, V. J., Lifanov, A. P., Régnier, M., Nazina, A. G., and Desplan, C. (2002). Extraction of functional binding sites from unique regulatory regions: the *Drosophila* early developmental enhancers. *Genome Res.* 12, 470–481. doi: 10.1101/gr.212502

Papatsenko, D., Goltsev, Y., and Levine, M. (2009). Organization of developmental enhancers in the *Drosophila* embryo. *Nucleic Acids Res.* 37, 5665–5677. doi: 10.1093/nar/gkp619

Papatsenko, D., and Levine, M. (2005). Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA.* 102, 4966–4971. doi: 10.1073/pnas.0409414102

Rebeiz, M., Reeves, N. L., and Posakony, J. W. (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. *Proc. Natl. Acad. Sci. U S A.* 99, 9888–9893. doi: 10.1073/pnas.152320899

Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., et al. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330, 1787–1797. doi: 10.1126/science.1198374

Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* 188, 415–431. doi: 10.1016/0022-2836(86)90165-8

Stormo, G. D., and Fields, D. S. (1998). Specificity, free energy and information content in protein-DNA interactions. *Trends Biochem. Sci.* 23, 109–113. doi: 10.1016/S0968-0004(98)01187-6

Stormo, G. D. (2015). DNA Motif databases and their uses. *Curr. Protoc. Bioinformatics* 51, 2.15.1–2.15.6. doi: 10.1002/0471250953.bi0215s51

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN-beta gene expression requires the assembly of an enhanceosome. *Cell* 83, 1091–1100. doi: 10.1016/0092-8674(95)90136-1

Thomas, S., Li, X.-Y., Sabo, P. J., Sandstrom, R., Thurman, R. E., Canfield, T. K., et al. (2011). Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* 12:R43. doi: 10.1186/gb-2011-12-5-r43

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 3:RESEARCH0088. doi: 10.1186/gb-2002-3-12-research0088

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., et al. (2007). Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* 8:R145. doi: 10.1186/gb-2007-8-7-r145

Wasserman, W. W., and Fickett, J. W. (1998). Identification of regulatory regions which confer muscle-specific gene expression. *J. Mol. Biol.* 278, 167–181. doi: 10.1006/jmbi.1998.1700

Wunderlich, Z., and Mirny, L. A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* 25, 434–440. doi: 10.1016/j.tig.2009.08.003

Yáñez-Cuna, J. O., Kvon, E. Z., and Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. *Trends Genet.* 29, 11–22. doi: 10.1016/j.tig.2012.09.007

Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 39, 111–117. doi: 10.1093/nar/gkq858

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Li and Wunderlich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.