Enhancer sharing promotes neighborhoods of transcriptional regulation across eukaryotes

Porfirio Quintero-Cadena* and Paul W. Sternberg†,
*Division of Biology and Biological Engineering, California Institute of Technology; Howard Hughes Medical Institute, Pasadena, California, USA.

ABSTRACT Enhancers physically interact with transcriptional promoters, looping over distances that can span multiple regulatory elements. Given that enhancer-promoter (EP) interactions generally occur via common protein complexes, it is unclear whether EP pairing is predominantly deterministic or proximity guided. Here we present cross-organismic evidence suggesting that most EP pairs are compatible, largely determined by physical proximity rather than specific interactions. By re-analyzing transcriptome datasets, we find that the transcription of gene neighbors is correlated over distances that scale with genome size. We experimentally show that non-specific EP interactions can explain such correlation, and that EP distance acts as a scaling factor for the transcriptional influence of an enhancer. We propose that enhancer sharing is commonplace among eukaryotes, and that EP distance is an important layer of information in gene regulation.

INTRODUCTION Enhancers mediate the transcriptional regulation of gene expression, enabling isogenic cells to exhibit remarkable phenotypic diversity (Davidson and Peter 2015). In complex with transcription factors, they interact with promoters via chromatin looping (Marsman and Horsfield 2012), finely regulating transcription in time and space. A prevailing view is that most enhancers have a mechanism to selectively loop to a target promoter (van Arensbergen et al. 2014). Examples in this category usually require specific transcription factor binding at both enhancer and promoter sites (Davidson and Peter 2015), which could explain why some enhancers seem to influence different promoters in varying degrees (Gehrig et al. 2009). On the other hand, EP looping is generally mediated by common protein complexes (Kagey et al. 2010; Malik and Roeder 2010), conflicting with the specific molecular interactions required by such a model at a larger scale. Examples of non-specific EP pairing also seem to be common (Butler and Kadonaga 2001). Yet given that this model could result in transcriptional crosstalk, it appears inconsistent with our current paradigm of gene regulation. The predominant EP pairing scheme –specific or non-specific– and its determinants are thus unclear. Here we ask to what extent are potential EP pairs compatible through a meta-analysis of the genome-wide transcription of gene neighbors in five species. We propose that enhancer sharing occurs widely across eukaryotes, test key aspects of this hypothesis in C. elegans, and analyze its implications in other genomic phenomena.

MATERIALS AND METHODS

Computational biology

For each analyzed organism, Ensembl (Flicek et al. 2014) protein-coding genes were grouped by chromosome, sorted by position, and paired with the 100 nearest neighbors within the same chromosome. A list of duplicated gene pairs for H. sapiens and M. musculus was obtained from the Duplicated Genes Database (Ouedraogo et al. 2012) (http://dgd.genouest.org). A list of C. elegans genes predicted to be in operons was obtained from Allen et al. (2011), and gene pairs present in the same operon were removed from the analysis to prevent co-transcriptional bias. Processed RNA-seq data was obtained from multiple sources (Gerstein et al. 2010; Atrill et al. 2016; Ellahi et al. 2015; The ENCODE Project Consortium 2012) and converted to transcripts per million (TPM) (Li et al. 2010) when necessary. Formatted datasets are available upon request. Genes detected in less than 80% of experiments were discarded. To compute the Spearman correlation coefficient, TPM values were ranked in each RNA-seq experiment and the pairwise Pearson correlation coefficient was computed on the ranked values according.
to the following equation:

\[ \rho = \frac{\text{cov}(\text{gene}_1, \text{gene}_2)}{\sigma_{\text{gene}_1} \sigma_{\text{gene}_2}} \]

where \( \text{gene}_1 \) and \( \text{gene}_2 \) are the corresponding ranks of each paired gene in a given RNA-seq experiment, \( \text{cov} \) their covariance and \( \sigma \) their standard deviation. The list of gene pairs with intergenic distances and correlation coefficients was sorted by increasing intergenic distance, and subsequently smoothed using a sliding median with window size of 1000 gene pairs. The result was then fitted to an exponential decay function of the form:

\[ \rho(d) = \rho_0 e^{-\lambda d} + c \]

where \( \rho_0 \) is the median Spearman correlation coefficient of the closest neighboring genes, \( d \) the intergenic distance and \( c \) the baseline correlation. The mean distance at which a pair of genes remain correlated was then computed as:

\[ d_{\text{exp}} = 1/\lambda \]

To compute the background correlation, each gene was paired with 20 randomly selected genes from a different chromosome and the 95% median confidence interval was computed by bootstrapping with 10,000 samples. A list of genes annotated with RNA in situ hybridization data (Tomancak et al. 2007; Hammonds et al. 2013; Tomancak et al. 2002) was obtained from the Berkeley Drosophila Genome Project (http://insitu.fruitfly.org). Insulator ChIP-chip data was obtained from Negre et al. (2010) (GSE16245); the intersection of replicates was used. HiC data was obtained from Rao et al. (2014) (GSE63525, GM12878 primary replicate HiCUPS looplist). Functional protein classification was conducted using Panther (Mi et al. 2016). Genomic manipulations were conducted using Bedtools v2.24.0 (Quinlan and Hall 2010). Data analysis was conducted using Python 2.7.9 and the Scipy library (McKinney et al. 2016). Plots were generated using Matplotlib 1.5 (Hunter 2007).

**Molecular biology**

*C. elegans* was cultured under standard laboratory conditions (Stieragle 2006). For enhancer additivity experiments, transgenic *C. elegans* lines carrying extra-chromosomal arrays were generated by injecting each plasmid at 50 ng/µL into unc-119 mutant animals. The minimal Δpes-10 promoter (Fire et al. 1990) and nuclear localized GFP (Lyssenko et al. 2007) were used in all constructs. Minimal regions of the myo-2 and unc-54 enhancers (Okkema et al. 1993) able to drive tissue specific expression were used. The BWM enhancer was obtained from the upstream region of F44B9.2; the BWM/intestine enhancer was obtained from the upstream region of rps-1. Animals were imaged at 40x using a GFP filter on a Zeiss Axioskop microscope.

For the enhancer promoter distance and ectopic enhancer experiments, we defined an EP distance of 0 to be the enhancer placed just upstream of the Δpes-10 promoter, which is ~350 bp away from the start codon of gfp. To ensure neutrality yet maintain a similar GC content as non-coding sequences in *C. elegans*, we used non-overlapping AT-rich DNA spacers obtained from the genome of *Escherichia coli*. Constructs were generated by integration into chromosome IV via CRISPR-Cas9 using plasmids provided as gifts by Dr. Zhiping Wang and Dr. Yishi Jin (unpublished results). Briefly, plasmids containing the following expression cassettes were co-injected: reporter and hygromycin resistance genes flanked by homologous arms for recombination-directed repair (10 ng/µL), single-guide RNA (10 ng/µL), Cas9 (10 ng/µL), and extra-chromosomal array reporter for expression of either rfp or gfp outside the tissue of interest (10 ng/µL). Transformants were selected for using hygromycin at 10 µg/µL, and those not carrying extra-chromosomal transgenes, lacking of gfp or rfp expression outside the tissue of interest, were subsequently isolated. Animals homozygotic for the insertion were identified by polymerase-chain reaction (PCR) and Sanger sequencing.

Quantitative PCR was carried out as previously described (Ly et al. 2015) using pmp-3 as a reference gene (Zhang et al. 2012). Briefly, third-stage larval (L3) worms, when expression from the test enhancers is maximal according to RNA-seq data, were synchronized at 20° via egg-laying. 15 animals were lysed in 1.5 µL of Lysis Buffer (5 mM Tris pH 8.0 (MP Biomedicals), 0.5% Triton X-100, 0.5% Tween 20, 0.25 mM EDTA (Sigma-Aldrich)) with proteinase-K (Roche) at 1.5 µg/µL, incubated at 65° for 10 minutes followed by 85° for one minute. Reverse transcription was carried out using the Maxima H Minus DNA synthesis kit (Thermo Fisher) by adding 0.3 µL H2O, 0.6 µL 5x enzyme buffer, 0.15 µL 10mM dNTP mix, 0.15 µL 0.5 µg/µL oligo dT primer, 0.15 µL enzyme mix and 0.15 µL DNase, and incubated for 2 minutes at 37°, followed by 30 minutes at 50° and finally 2 minutes at 85°. The cDNA solution was diluted to 15 µL and 1 µL was used for each qPCR reaction, so that on average each well contained the amount of RNA from a single worm. All qPCR reactions were performed with three technical replicates and at least three biological replicates using the Roche LightCycler® 480 SYBR Green I Master in the LightCycler® 480 System. Crossing point-PCR-cycle (Cp) averages were computed for each group of three technical replicates; these values were then subtracted from the respective average Cp value of the reference gene.

**Data and reagent availability**

Strains are available upon request. Relevant DNA sequences, including spacers, enhancers, primers, sgRNA and homology arms are available in Table S1. Correlation datasets are available in File S1. qPCR data is available in Table S2. Python source code, and links to all expression datasets used in this study, are available for download on the following github repository: https://github.com/WormLabCaltech/QuinteroSternberg2016.git.

**RESULTS AND DISCUSSION**

**Gene neighbors are transcriptionally correlated genome-wide**

We reasoned that widespread EP compatibility should result in transcriptional correlation among gene neighbors. Indeed, gene coexpression clusters have been extensively reported in eukaryotic genomes (e.g. Sémon and Duret September 2006; Roy et al. 2002; Lercher et al. 2002; Lercher and Hurst 2006; Williams and Hurst 2002; Singer et al. 2005; Lercher et al. 2003; Williams and Bowles 2004; Spellman and Rubin 2002; Purmann et al. 2007; Zhan et al. 2006; Boutanaev et al. 2002; Kalmykova et al. 2005; Caron et al. 2001; Rubin and Green 2013), in spite of order of magnitude variations in genome size (e.g. ~12 Mb in *S. cerevisiae* vs ~3 Gb in *H. sapiens*). An early informative example is the discovery of chromosomal domains of gene expression in *S. cerevisiae* (Cohen et al. 2000) which exhibit features that strongly support enhancer-sharing, mainly distance-dependence in transcriptional correlation that qualitatively resemble chromosome contact matrices (e.g. Rao et al. 2014), and instances in which a single enhancer seems to be responsible for the coexpression of adjacent gene pairs. The ubiquity of these features across eukaryotes would support the idea that EP interactions are largely determined by physical proximity rather than by specific interactions. Given the accumulation of transcriptome
sequencing data, we decided to investigate the transcriptional correlation of gene neighbors in representative eukaryotic species as a first step to explore the average EP pairing scheme.

We paired every protein-coding gene of five organisms (Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus and Homo sapiens) with its 100 nearest neighbors within the same chromosome. This yielded lists of around 600,000 (Figure S4a). This analysis revealed that close neighbors have a spatio-temporal relation pattern across organisms, and also decays exponentially as the distance between contacting sites increases (Ringrose et al. 1999; Rao et al. 2014), with the same pattern as observed here at least in some documented cases (e.g. H. sapiens, Figure S5) ii) the average distance between a large fraction of studied EP interactions scales with genome size in ranges often consistent with $d_{exp} < 1$ kb in S. cerevisiae, (Dobi and Winston 2007); $< 10$ kb in C. elegans, (Araya et al. 2014); and $120$ kb in H. sapiens, (Sanyal et al. 2012) iii) common protein complexes such as the mediator seem to be widely utilized bridges in EP looping (Kagey et al. 2010; Malik and Roeder 2010) iv) a high frequency of chromatin interactions are observed within topologically associated domains identified through high-resolution Chromosome Conformation Capture (Hi-C) (Rao et al. 2014) and v) studied enhancers often do not show promoter specificity (Butler and Kadonaga 2001). This line of reasoning suggests a model where, as opposed to only having a specific target gene (Figure 2a), the average enhancer has a range of action in which it can influence any active promoter within its reach (Figure 2b). A concrete example consistent with this idea is the upregulation of neighboring genes upon enhancer activation by fibroblast growth factor in mammalian cells (Ebisuya et al. 2008). Transcriptome analysis could thus provide indirect evidence of genome and condition-wide EP looping that is difficult to access through Hi-C (Rao et al. 2014) due to the low signal-to-noise ratio of short-range interactions.

Because of its compact genome, rapid development and availability of tissue specific enhancers (Corsi et al. 2015), we decided to use C. elegans to test the validity of a non-specific EP pairing model. We first postulated that unrelated enhancers should generally be compatible, showing qualitative additivity when placed upstream of a single promoter. We thus paired the well characterized myo-2 pharyngeal enhancer with a body wall muscle (BWM) and a BWM plus intestine enhancer, placed them upstream of a minimal promoter and a gfp reporter, and examined expression in transgenic animals. In both cases, we observed fluorescence in the corresponding tissues (Figure 2c, d, e). This observation is consistent with typical enhancer studies in artificial constructs (Dupuy et al. 2004) and agrees with some EP compatibility studies (Butler and Kadonaga 2001).

Given that both chromatin looping and expression correlation decay exponentially, we reasoned that transcription of a given gene should also decay exponentially with increasing EP distance if the observed correlation is to be explained by enhancer sharing. To test this hypothesis, we first built a series of genetic constructs with increasing neutral EP distances (0, 1, 1.5 and 2 kb) for two different enhancers, myo-2 and unc-54 (~400 and 300 bp, respectively). We then integrated each construct in single copy into the genome of C. elegans and used quantitative PCR to i) measure the influence of EP distance on the reporter gene in native chromatin and ii) analyze the impact of the perturbation on the two genes that natively flank the site of transgene insertion (dpy-13 and col-34, Figure 2h), which we reasoned should be affected in two counteracting ways. First, the ectopic enhancers should promote their expression. Second, the increased EP distance caused by the addition of spacers should reduce their expression by scaling down the influence of both ectopic and native enhancers (the latter of unknown identity and location) to the opposite side of the spacer.

We found that transcriptional levels of the reporter gene indeed fall rapidly with increasing EP distance with both enhancers (Figure 2f, g); this occurred at a rate that seems congruent or faster than the genome-wide correlation decay, likely because of the dramatic separation of every regulatory element at once, as opposed to gradual separation from individual enhancers in a native environment; this dramatic effect suggests complex interactions.

Enhancer sharing explains the transcriptional correlation of gene neighbors

The pervasive nature of proximal gene coexpression supported the idea of widespread EP compatibility. This connection is, in turn, supported by several other observations in literature: i) enhancers regulate transcription by making contact with promoters via chromatin looping (Marsman and Horsfield 2012), whose incidence also decays exponentially as the distance between contacting sites
between multiple EP loops that are disrupted with the insertion of DNA sequences devoid of regulatory elements. Transcription was still well detected even when the enhancers were placed 2 kb away, supporting the hypothesis that EP distance is a scaling factor on the enhancer’s influence. Expression of dpy-13 and col-34 was reduced with the introduction of the 2 kb spacer when compared to transgenic lines without it (Figure 2i, j). On the other hand, spacer-free lines were comparable to wild-type, suggesting the incorporation of ectopic enhancers compensated for the EP distance increase caused by the addition of the genetic construct itself. These observations seem to fit the corollaries of our model, even amid the complexity of a native regulatory environment. However, the distance over which we see an effect on col-34 falls outside our \( d_{\text{exp}} \) estimate for \( C.\ elegans \) (8 kb). Its expression is impacted by the presence of the 2 kb spacer outside of the interval between the \( myo-2/unc-54 \) enhancer, suggesting that enhancers >12 kb away can still influence its expression. As evidenced with the discrepancy in \( D.\ melanogaster \) when using in situ or RNA-seq data, this observation suggests that \( d_{\text{exp}} \) is only a rough estimate of the average enhancer rage of action; this is useful to gain insight into genome-wide mechanisms but not for precise individual predictions.

Chromatin modifications have been shown to have a significant impact on enhancer function (Calo and Wysocka 2013) and thus likely influence EP pairing. Thus, chromatin features and enhancer sharing might be mutually inclusive rather than stand alone explanations for the observed correlation domains. From this perspective, transcriptionally correlated genes would have similar chromatin states, facilitated by their physical proximity, that make them accessible to enhancer action.

The existence of multiple, independent but similar enhancers is an alternative possible explanation. However, since we are looking at genome-wide averages, this would mean that most gene neighbors have a functionally redundant set of independent enhancers that function through distinct molecular interactions. Although possible, this is a rather intricate explanation.

In agreement with the enhancer sharing hypothesis, it can be argued that the scaling of correlation domains is a result of the ability to connect EP loops over longer distances in larger genomes. Yet in spite of having a compact genome, \( D.\ melanogaster \) is able to form many long-range EP interactions (> 50 kb) (Ghavi-Helm et al. 2014), which is considerably different to the range of its estimated \( d_{\text{exp}} \) (6-22 kb). Additionally, these long-range interactions appear to be particularly specific, with enhancers selectively activating their target promoters (Ghavi-Helm et al. 2014; Kwon et al. 2009). It is thus possible that in compact genomes, long-range EP interactions would need to be specific, whereas nearby interactions would tend to fall in the non-specific pairing scheme, ultimately resulting in the observed correlation domain size.

Enhancer-promoter distance insulates neighboring genes

We next wished to determine the extent to which enhancer sharing could explain other genomic phenomena. Previous reports have suggested that divergent, parallel and convergent gene pairs tend to have distinct correlation profiles (e.g. Chen and Stein 2006). To explore this hypothesis, we compared the distribution of intergenic distances of gene pairs oriented in parallel, divergent and convergent orientations (Figure 3a, Figure S6). As expected, divergent gene pairs tend to be closest, followed by parallel and finally convergent genes. We then confirmed that each group appears to have different distributions of correlations \( D.\ melanogaster \) (Figure 3b, Figure S6). To consider the influence of EP distance, we sampled gene pairs from each orientation controlling for intergenic size. This resulted in distributions of correlations that exactly overlap (Figure 3c, Figure S6), an observation that is supported by previous reports in specific cases (Ghanbarian and Hurst 2015; Cohen et al. 2000). We thus conclude that the apparent influence of gene orientation in the transcriptional relationship of neighboring gene pairs is consistent with the enhancer sharing hypothesis. In this scenario, the effect of gene orientation can be simply explained by the different EP distance distributions for each configuration.

From the regulatory perspective, EP distance provides independence to most gene pairs, as the vast majority have an intergenic distance that puts them in the baseline correlation regime (Figure 1c). To study the enhancer-blocking influence of insulators (Bushey et al. 2009) genome-wide, we analyzed each group of genes flanked by insulator binding sites, which were previously determined using Chromatin ImmunoPrecipitation coupled with microarrays (ChiP-chip) for six known insulators in \( D.\ melanogaster \): BEAF-32, CP190, CTCF, GAF, Mod(mdg4) and Su(Hw) (Negre et al. 2010). We observed that gene neighbors closer than 10 kb bound by each of the insulators tend to be less correlated in gene expression than gene pairs not bound by them (Figure 3e), supporting their role in genome-wide insulation and agreeing with the observation that insulators tend to separate differentially expressed genes (Negre et al. 2010; Xie et al. 2007). Nevertheless, the same groups of gene pairs also tend to have much larger intergenic distances than genes that are not flanked by insulator binding sites (Figure 3d). After controlling for the distribution of intergenic distances, we found very similar correlation distributions between insulator and not insulator flanked gene pairs (Figure 3f). This finding agrees with previous reports suggesting that insulators do not block enhancers everywhere they bind, but rather act only on very specific genomic contexts (Schwartz et al. 2012; Liu et al. 2015; Ong and Corces 2014); it also reconciles the lack of known insulator orthologs in \( C.\ elegans \) (Heger et al. 2009) in the context of local enhancer-blocking. In combination, these studies strongly suggest that EP distance is the general source of transcriptional independence for close gene neighbors.

Previous EP compatibility studies suggest that EP specificity is widespread (Gehrig et al. 2009), while others that it is restricted to a smaller subset of enhancers (Butler and Kadonaga 2001). Although our results support the latter, views arising from these studies might not be mutually exclusive, as it is likely that enhancers have specificity to promoter classes (Danino et al. 2015), whose limited number could result in general EP compatibility.

The implications from considering our observations are broadly applicable to gene regulation. Position effects, in which transgene expression levels are influenced by the insertion site (Gierman et al. 2007), are naturally expected from enhancer sharing. Chromosomal translocations and mutations involving regulatory elements likely impact genetic contexts rather than individual genes. Furthermore, enhancer sharing and distance-based scaling of enhancer influence potentially provides an additional layer of information in gene regulation, as the transcriptional output of a given gene would be the result of scaled contributions from multiple shared enhancers. Such a feature could by itself be under selective pressure, leading to a roughly constant size of the correlation domain in number of genes regardless of absolute physical distance, as observed in this study. Our analysis provides a clarifying perspective of gene regulation consistent with both mechanistic and genome-wide studies.
ACKNOWLEDGMENTS

Our work was supported by the Howard Hughes Medical Institute, of which P.W.S is an investigator. We thank Zhiping Wang and Yishi Jin for plasmids for Crispr-Cas9 single copy insertion, Carmie Robinson for discussions, experimental suggestions and comments on the manuscript, Han Wang for discussions, technical advise and comments on the manuscript, Hillel Schwartz, Mitchell Guttman, Mihoko Kato, David Angeles-Albores, Jonathan Liu, Barbara Wold, Isabelle Peter and Angelike Stathopoulos for discus-sions and comments on the manuscript, the Encode and Mod-Encode consortiurns, FlyBase, WormBase and Ensembl databases, the Wold Lab and the Guigo Lab for data accessibility.

CONTRIBUTIONS

P.Q.C. performed the experiments and analyzed the data. P.Q.C. and P.W.S. designed the experiments and wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.
Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of C. elegans trans-splicing. Genome Res. 21: 255–264.

Araya, C. L., T. Kawli, A. Kundaje, L. Jiang, B. Wu, D. Vafeados, R. Terrell, P. Weissdepp, L. Gevirtzman, D. Mace, W. Niu, A. P. Boyle, D. Xie, L. Ma, J. I. Murray, V. Reinke, R. H. Waterston, and M. Snyder, 2014 Regulatory analysis of the C. elegans genome with spatiotemporal resolution. Nature 512: 400–405.

Attrill, H., K. Falls, J. L. Goodman, G. H. Millburn, G. Antonazzo, A. J. Rey, S. J. Marygold, and the FlyBase consortium, 2016 Flybase: establishing a gene group resource for drosophila melanogaster. Nucleic Acids Res. 44: D786–D792.

Boutanaev, A. M., A. I. Kalmykova, Y. Y. Shevelyov, and D. I. Nurminsky, 2002 Large clusters of co-expressed genes in the drosophila genome. Nature 420: 666–669.

Bushey, A. M., E. R. Dorman, and V. G. Corces, 2009 Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol. Cell 32: 1–9.

Butler, J. E. and J. T. Kadonaga, 2001 Enhancer–promoter specificity mediated by dpe or tata core promoter motifs. Genes & Development 15: 2515–2519.

Calo, E. and J. W. Wysocka, 2013 Modification of enhancer chromatin: What, how, and why? Molecular Cell 49: 825 – 837.

Caron, H., B. v. Schaik, M. v. d. Mee, F. Baas, G. Riggins, P. v. Sluis, M.-C. Hermus, R. v. Asperen, K. Boon, P. A. Voûte, S. Heisterkamp, A. v. Kampen, and R. Versteeg, 2001 The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. Science 291: 1289–1292.

Chen, N. and L. D. Stein, 2006 Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans. Genome Res. 16: 606–617.

Cohen, B. A., R. D. Mitra, J. D. Hughes, and G. M. Church, 2000 A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet 26: 183–186.

Corsi, A. K., B. Wightman, and M. Chalfie, 2015 A transparent window into biology: A primer on Caenorhabditis elegans. Genetics 200: 387–407.

Danino, Y. M., D. Even, D. Ideses, and T. Juven-Gershon, 2015 The core promoter: At the heart of gene expression. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1849: 1116 – 1131.

Davidson, E. H. and I. S. Peter, 2015 Chapter 1 - the genome in development. In Genomic Control Process, edited by E. H. Davidson and I. S. Peter, pp. 1–40. Academic Press, Oxford.

Dobi, K. C. and F. Winston, 2007 Analysis of transcriptional activation at a distance in saccharomyces cerevisiae. Molecular and Cellular Biology 27: 5575–5586.

Dupuy, D., Q.-R. Li, B. Deplancke, M. Boxem, T. Hao, P. Lamesch, R. Sequerra, S. Bosak, L. Doucette-Stamm, I. A. Hope, D. E. Hill, A. J. Walhout, and M. Vidal, 2004 A first version of the caenorhabditis elegans promoterome. Genome Res. 14: 2169–2175.

Ebisuya, M., T. Yamamoto, M. Nakajima, and E. Nishida, 2008 Ripples from neighbouring transcription. Nat Cell Biol 10: 1106–1113.

Ehahori, A., D. M. Thurtle, and J. Rine, 2015 The chromatin and transcriptional landscape of native saccharomyces cerevisiae telomeres and subtelomeric domains. Genetics 2: 505–521.

Fire, A., S. W. Harrison, and D. Dixon, 1990 A modular set of lac fusion vectors for studying gene expression in caenorhabditis elegans. Gene 93: 189 – 198.

Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, E. Kulesha, F. J. Martin, T. Maurel, W. M. McLaren, D. N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, M. Ruffier, D. Shepheard, K. Taylor, A. Thommann, S. J. Trevanion, A. Vullo, S. P. Wilder, M. Wilson, A. Zadissa, B. L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T. J. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D. R. Zerbino, and S. M. Searle, 2014 Ensembl 2014. Nucleic Acids Res. 42: D749–D755.

Gehrig, J., M. Reischl, E. Kalmar, M. Ferg, Y. Hadzhiev, A. Zauker, C. Song, S. Schindler, U. Liebel, and F. Muller, 2009 Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. Nat Meth 6: 911–916.

Gerstein, M. B., Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff, T. Liu, K. Y. Yip, R. Robinollo, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateignier, M. Perry, M. Morris, R. K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhiessorraek, A. Aragwal, R. P. Alexander, G. Barber, C. M. Brdlik, J. Brennan, J. N. J. Brouillet, A. Carr, M.-S. Cheung, H. Clawson, S. Contrino, L. O. Dannenberg, A. F. Dernburg, A. Desai, L. Dick, A. C. Döse, J. Du, T. Egelhofer, S. Erkan, G. Euskirchen, B. Ewing, E. A. Feingold, R. Gassmann, P. J. Good, P. Green, F. Gullier, M. Gutwein, M. S. Guyer, L. Habegger, T. Han, J. G. Henikoff, S. R. Henz, A. Hinrichs, H. Holster, T. Hyman, A. L. Iniguez, J. Janette, M. Jensen, M. Kato, W. J. Kent, E. Kephart, V. Khivansara, E. Khurana, J. K. Kim, P. Kolasinska-Zwierz, E. C. Lai, I. Lalorere, A. Leahey, S. Lewis, P. Lloyd, L. Lovchovsky, R. F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S. D. Mackowiak, M. Mangone, S. McKay, D. Mecenas, G. Merrihew, D. M. Miller, A. Muroyama, J. I. Murray, S.-L. Ooi, H. Pham, T. Phippen, E. A. Preston, N. Rajewsky, G. Rätsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sadhisharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F. J. Slack, C. Slightam, R. Smith, W. C. Spencer, E. O. Stinson, S. Taing, T. Takasaki, D. Vafeados, K. Vorinina, G. Wang, N. L. Washington, C. M. Whittle, B. Wu, K.-K. Yan, G. Zeller, Z. Zha, M. Zhong, X. Zhou, J. Ahringer, S. Strome, K. C. Gunsalus, G. Micklem, X. S. Liu, V. Reinke, S. K. Kim, L. W. Hillier, S. Henikoff, F. Piano, M. Snyder, L. Stein, J. D. Lieb, and R. H. Waterston, 2010 Integrative analysis of the caenorhabditis elegans genome by the modencode project. Science 330: 1775–1787.

Ghanbarian, A. T. and L. D. Hurst, 2015 Neighboring genes show correlated evolution in gene expression. Mol. Biol. Evol. 32: 1748–1766.

Ghavi-Helm, Y., F. A. Klein, T. Pakozdi, L. Siglar, D. Noordermeer, W. Huber, and E. E. Furlong, 2014 Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512: 96–100.

Gierman, H. J., M. H. Indemans, J. Koster, S. Goetze, J. Seppen, D. Geerts, R. van Driel, and R. Versteeg, 2007 Domain-wide regulation of gene expression in the human genome. Genome Res. 17: 1286–1295.

Hammonds, A. S., C. A. Bristow, W. W. Fisher, R. Weiszmann, S. Wu, V. Hartenstein, M. Kellis, B. Yu, E. Frise, and S. E. Celniker, 2013 Spatial expression of transcription factors in drosophila embryonic organ development. Genome Biol. 14: R140.

Heger, P., B. Marin, and E. Schierenberg, 2009 Loss of the insulator protein CTCF during nematode evolution. BMC Mol. Biol. 5: 1–14.
Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, and R. A. Young, 2010 Mediator and cohesin connect gene expression and chromatin architecture. Nature 467: 430–435.

Kalmykova, A. I., D. I. Nurminsky, D. V. Ryzhov, and Y. Y. Shevelevyov, 2005 Regulated chromatin domain comprising cluster of co-expressed genes in drosophila melanogaster. Nucleic Acids Research 33: 1435–1444.

Kwon, D., D. Mucci, K. K. Langlais, J. L. Americo, S. K. DeVido, Y. Cheng, and J. A. Kassis, 2009 Enhancer-promoter communication at the drosophila engrailed locus. Development (Cambridge, England) 136: 3067–3075.

Lercher, M. J., T. Blumenthal, and L. D. Hurst, 2003 Coexpression of neighboring genes in caenorhabditis elegans is mostly due to operons and duplicate genes. Genome Research 13: 238–243.

Lercher, M. J. and L. D. Hurst, 2006 Co-expressed yeast genes cluster over a long range but are not regularly spaced. Journal of Molecular Biology 359: 825 – 831.

Lercher, M. J., A. O. Urrutia, and L. D. Hurst, 2002 Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31: 180–183.

Li, B., V. Ruotti, R. M. Stewart, J. A. Thomson, and C. N. Dewey, 2010 Rna-seq gene expression estimation with read mapping uncertainty. Bioinformatics 26: 493–500.

Liu, M., M. T. Maurano, H. Wang, H. Qi, C.-z. Song, P. A. Navas, D. W. Emery, J. A. Stamatoyannopoulos, and G. Stamatoyannopoulos, 2015 Genomic discovery of potent chromatin insulators for human gene therapy. Nat. biotechnol. 33: 198–203.

Ly, K., S. J. Reid, and R. G. Snell, 2015 Rapid RNA analysis of individual Caeorhabditis elegans. MethodsX 2: 59–63.

Lyssenko, N. N., W. Hanna-Rose, and R. A. Schlegel, 2007 Cognate putative nuclear localization signal effects strong nuclear localization of a gfp reporter and facilitates gene expression studies in caenorhabditis elegans. Biotechniques 43: 596–600.

Malik, S. and R. G. Roeder, 2010 The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. 11: 761–772.

Marsman, J. and J. A. Horsfield, 2012 Long distance relationships: Enhancer–promoter communication and dynamic gene transcription. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819: 1217 – 1227.

McKinney, W., 2010 Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, edited by S. van der Walt and J. Millman, pp. 51–56.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, 2016 Panther version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 44: D336–D342.

Michalak, P., 2008 Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics 91: 243 – 248.

Negre, N., C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison, S. Henikoff, M. Kellis, and K. P. White, 2010 A Comprehensive Map of Insulator Elements for the Drosophila Genome. Plos Genet. 6: e1000814.

Ong, C.-T. and V. G. Corces, 2014 Ctfc: an architectural protein bridging genome topology and function. Nat Rev Genet 15: 234–246.

Ouedraogo, M., C. Bettembourg, A. Bretaudeau, O. Sallou, C. Diot, O. Demeure, and F. Lecerf, 2012 The duplicated genes database: Identification and functional annotation of co-localised duplicated genes across genomes. PLoS ONE 7: 1–8.

Purmann, A., J. Toedling, M. Schueler, P. Carminci, H. Lehrah, Y. Hayashizaki, W. Huber, and S. Sperling, 2007 Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. Genomics 89: 580 – 587.

Quinlan, A. R. and I. M. Hall, 2010 Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Rao, S. S. P., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, 2014 A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–1680.

Ringrose, L., S. Chabanis, P. O. Angrand, C. Woodroofe, and A. F. Stewart, 1999 Quantitative comparison of dna looping in vitro and in vivo: chromatin increases effective dna flexibility at short distances. EMBO J. 18: 6630–6641.

Roy, P. J., J. M. Stuart, J. Lund, and S. K. Kim, 2002 Chromosomal clustering of muscle-expressed genes in caenorhabditis elegans. Nature 418: 975–979.

Rubin, A. F. and P. Green, 2013 Expression-based segmentation of the drosophila genome. BMC Genomics 14: 1–8.

Sanyal, A., B. R. Lajoie, G. Jain, and J. Dekker, 2012 The long-range interaction landscape of gene promoters. Nature 489: 109–113.

Schwarz, Y. B., D. Linder-basso, P. V. Kharchenko, M. Y. Tolorukov, M. Kim, H.-b. Li, A. A. Gorchakov, A. Minoda, G. Shanower, A. A. Alekseyenko, N. C. Riddle, Y. L. Jung, T. Gu, A. Plachetka, S. C. R. Elgin, M. I. Kuroda, P. J. Park, M. Savitsky, and G. H. Karpen, 2012 Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res. 11: 2188–2198.

Séminon, M. and L. Duret, September 2006 Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Molecular Biology and Evolution 23: 1715–1723.

Singer, G. A. C., A. T. Lloyd, L. B. Huminecki, and K. H. Wolfe, 2005 Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. Molecular Biology and Evolution 22: 767–775.

Spellman, P. T. and G. M. Rubin, 2002 Evidence for large domains of similarly expressed genes in the drosophila genome. Journal of Biology 1: 1–8.

Stiernagle, T., 2006 Maintenance of c. elegans. In WormBook, edited by The c. elegans Research Community, WormBook, The ENCODE Project Consortium, 2012 An integrated encyclopedia of dna elements in the human genome. Nature 489: 57–74.

Tomancak, P., A. Beaton, R. Weissmann, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, and G. M. Rubin, 2002 Systematic determination of patterns of gene expression during drosophila embryogenesis. Genome Biol. 3: research0088.1–88.14.

van Arensbergen, J., B. van Steensel, and H. J. Bussemaker, 2014 In search of the determinants of enhancer–promoter interaction specificity. Trends in cell biology 24: 695–702.
Williams, E. J. and D. J. Bowles, 2004 Coexpression of neighboring genes in the genome of arabidopsis thaliana. Genome Research 14: 1060–1067.
Williams, J. B. E. and D. L. Hurst, 2002 Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. Journal of Molecular Evolution 54: 511–518.
Xie, X., T. S. Mikkelsen, A. Grirke, K. Lindblad-toh, M. Kellis, and E. S. Lander, 2007 Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. USA 104: 7145–7150.
Zhan, S., J. Horrocks, and L. N. Lukens, 2006 Islands of co-expressed neighbouring genes in arabidopsis thaliana suggest higher-order chromosome domains. The Plant Journal 45: 347–357.
Zhang, Y., D. Chen, M. A. Smith, B. Zhang, and X. Pan, 2012 Selection of reliable reference genes in Capenorhabditis elegans for analysis of nanotoxicity. PLoS ONE 7: 1–7.
Figure 1 Neighboring genes are transcriptionally correlated genome-wide across eukaryotes. a) Sliding median of correlations between paired neighbors (blue line) and interquartile range (pale blue) with increasing intergenic distance. Median ± 95% confidence interval of randomly paired genes is shown as a horizontal gray line. Fit to an exponential decay function (red line) was used to compute the mean distance at which gene neighbors remain correlated ($d_{exp}$, vertical red dashed line). The genome size (G) is displayed for each organism. Distribution of intergenic distances between each gene and its nearest neighbor (b) and all paired genes (c). The organism analyzed in each case is indicated for each group of three subplots.
Figure 2 Enhancer sharing explains the transcriptional correlation of gene neighbors. Two possible models for EP relationship: a) Enhancers have specific target genes and b) enhancers have a range of action in which they influence genes by physical proximity. Tissue specific enhancers (c) are generally compatible. Pharynx and body wall muscle (d) and pharynx, body wall muscle and intestine (e) enhancers driving nuclear gfp expression. mRNA levels of gfp with increasing EP distance for lines with myo-2 (filled circles, f) and unc-54 (hollow circles, g) enhancers. h) Genomic context of the integration site. The inserted construct is shown over a dashed black line and includes a hygromycin resistance gene (hyg-R) regulated by a ribosomal enhancer (rps-0) and promoter in addition to the reporter (gfp) regulated by either the myo-2 or unc-54 enhancers; the native genes dpy-13 and col-34 flank the insertion site. Relative mRNA levels of dpy-13 (i) and col-34 (j) in wild-type and lines with and without the 2 kb spacer (*two tailed P-val<0.05, Mann-Whitney U test). The difference in crossing point-PCR-cycle (∆Cp) with the reference gene pmp-3 and the corresponding median for each group of biological replicates is shown for every qPCR experiment.

Figure 3 EP distance causes gene orientation-dependent correlation and provides regulatory independence to gene neighbors. Distribution of intergenic distances below 10 kb of gene pairs in D. melanogaster by configuration (~5 to 18 thousand gene pairs for each group, a) and flanking insulator binding sites identified through ChIP-chip (Negre et al. 2010) (~5 to 15 thousand pairs for each group, d). The corresponding distribution of correlations is shown for the same gene pairs (b, e) and pairs with controlled distributions of intergenic distances between 30 and 40 kb (~7 to 14 thousand pairs for gene orientation groups, ~10 to 18 thousand for insulator groups, c, f).