



RESEARCH ARTICLE

Genome-wide measurement of spatial expression in patterning mutants of *Drosophila melanogaster* [version 1; referees: 2 approved]

Peter A. Combs¹, Michael B. Eisen²

¹Graduate Program in Biophysics, University of California, Berkley, USA

²Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, USA

v1 First published: 12 Jan 2017, 6:41 (doi: [10.12688/f1000research.9720.1](https://doi.org/10.12688/f1000research.9720.1))
 Latest published: 12 Jan 2017, 6:41 (doi: [10.12688/f1000research.9720.1](https://doi.org/10.12688/f1000research.9720.1))

Abstract

Patterning in the *Drosophila melanogaster* embryo is affected by multiple maternal factors, but the effect of these factors on spatial gene expression has not been systematically analyzed. Here we characterize the effect of the maternal factors Zelda, Hunchback and Bicoid by cryosectioning wildtype and mutant blastoderm stage embryos and sequencing mRNA from each slice. The resulting atlas of spatial gene expression highlights the intersecting roles of these factors in regulating spatial patterns, and serves as a resource for researchers studying spatial patterning in the early embryo. We identify a large number of genes with both expected and unexpected patterning changes, and through integrated analysis of transcription factor binding data identify common themes in genes with complex dependence on these transcription factors.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 12 Jan 2017	 report	 report

- 1 **Thomas Gregor**, Princeton University
USA
- 2 **Angela H. DePace**, Harvard University
USA

Discuss this article

Comments (0)

Corresponding authors: Peter A. Combs (pcombs@stanford.edu), Michael B. Eisen (mbeisen@gmail.com)

How to cite this article: Combs PA and Eisen MB. **Genome-wide measurement of spatial expression in patterning mutants of *Drosophila melanogaster* [version 1; referees: 2 approved]** *F1000Research* 2017, 6:41 (doi: [10.12688/f1000research.9720.1](https://doi.org/10.12688/f1000research.9720.1))

Copyright: © 2017 Combs PA and Eisen MB. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by a Howard Hughes Medical Institute investigator award to MBE. PAC was supported by the National Institutes of Health under a Genomics Training Grant (5T32HG000047-13) .
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: MBE is on the International Advisory Board for F1000.

First published: 12 Jan 2017, 6:41 (doi: [10.12688/f1000research.9720.1](https://doi.org/10.12688/f1000research.9720.1))

Introduction

In the early *Drosophila melanogaster* embryo, the spatially restricted activity of several maternally deposited factors establishes the main body axes of the animal by triggering cascades of patterned gene expression. Until recently, however, it was not practical to systematically characterize the effects of these factors on patterned expression, as the dominant method for studying spatial expression, *in situ* hybridization, does not scale well.

We previously introduced a method for the genome-wide measurement of spatial patterns of expression in the *Drosophila* embryo based on cryosectioning individual embryos along the anteroposterior axes and sequencing the mRNA from each slice. Here we extend this method to characterize embryos mutant for three maternal transcription factors (TFs): Bicoid, Zelda, and Hunchback.

All of these factors are crucial for proper patterning of the embryo. Bicoid is the primary, maternally provided anterior specifier that directly regulates many key gap and pair-rule genes¹⁻⁴. More recently, the role of Zelda (also known as vfl) has been identified as an important factor in establishing cis-regulatory chromatin domains of patterned genes⁵⁻¹². Finally, Hunchback is both maternally and zygotically expressed, and helps to specify the expression domains and levels of various gap and pair-rule genes in the thorax^{4,13,14}. Although broad-reaching examinations of the effects of mutating these genes on their targets has not previously been possible, mutating *bicoid* results in the anterior adopting an anterior-like fate^{15,16}, mutating *zelda* leads to shifts of expression patterns in time and space¹², and Hunchback binding has been implicated in multiple *eve* stripes, among many other expression patterns.^{4,17}; all of these mutations are lethal. Given the crucial roles of each of these factors in spatial patterning, we expected that perturbing their levels would lead to widespread direct and indirect effects on patterned genes.

Results

Genome-wide atlases of the blastoderm stage of multiple dosage mutants

We sliced embryos and sequenced the resulting mRNA from 4 mutant genotypes (Figure 1A): a *zld* germline clone, an RNAi knockdown for *bcd*, a knockdown for *hb*, and an overexpression line for *bcd* with approximately 2.4× wildtype expression. We chose two time points: cycle 13 (determined using nuclear density of either DAPI stained embryos or of the Histone-RFP present in the *zld* line) and mid-to-late cycle 14 (determined using 50–65% membrane invagination at stage 5) (Figure 1B). Genes expressed in cycle 13 are towards the end of the early round of genome activation and are enriched for Zld binding^{10,18,19}, but are early enough that the majority of patterning disruptions are likely to be direct effects of the mutants. By contrast, we chose the stage 5 time point in order to highlight the full extent of the patterning changes across the network.

In order to show the range of patterning differences observed, we generated heatmaps of all the gene expression present in the dataset (Figure 2). Of the 7104 genes with at least 15FPKM in at least one slice, approximately 3000 had uniform expression in all the wild-type embryos that was not greatly perturbed in any of the

mutants. The total number of expressed genes is very consistent with previous estimates of the number of maternally deposited and zygotically transcribed genes¹⁹.

The set of genes with anterior or posterior localization recapitulate the known literature¹⁶ and general expectations in the *bcd*-case: those expressed in the anterior typically lose expression (Figure 3A), and those in the posterior also frequently gain an expression domain in the anterior (Figure 3B). Surprisingly, most of these patterns are qualitatively unaffected in the other mutants. In the absence of *zld*, most of these genes are able to retain the proper anterior patterning (although they may have differences in expression levels). Similarly, these genes seem not to be strongly dependent on maternal Hb for patterning information, with most genes retaining a distinct anterior expression domain. As described in Liang *et al.*¹², there are some genes that are normally ubiquitously expressed in the wild-type that become localized to the poles in the *zld*-embryo (Figure 3C).

Most spatial patterns are robust to mutation

We next compared expression patterns from each of the mutant lines at late cycle 14 to similar expression patterns in wild-type. Because there are a different number of slices both between the wild type and mutant flies, and between replicates of the mutant flies, we decided to use Earth Mover's Distance (EMD) to compare patterns²⁰. This metric captures intuitive notions about what kinds of patterns are dissimilar, yielding higher distances for dissimilar distributions of RNA, and zero for identical distributions. Patterns were normalized to have the same maximum expression, in order to highlight changes in positioning of patterns, rather than changes in absolute level. In contrast to traditional RNA-seq differential expression metrics, this approach takes advantage of the spatial nature of the data, and with the fine slices, adjacent slices are able to function as "pseudo-replicates". Adjacent slices are, on average, much more similar than those from farther away in the same embryo (Figure S1).

The overall level of divergence in pattern across all genes is, in most cases, slightly larger than when comparing nearby time-points in wild-type or replicates of the same genotype and time point (Figure 4). Notably, the *zld*-mutant is more similar to wild-type than the mutants of the other, spatially distributed transcription factors. This suggests that Zld is a categorically different TF, consistent with its role as a pioneer factor rather than a direct activator. However, the low level of divergence is a reflection of the fact that the majority of genes are not dynamically expressed in either time or space.

In order to demonstrate that these mutants are more likely to affect already known Bcd regulatory systems, we examined genes that were close to 64 Bcd dependent enhancers previously identified²¹⁻²⁷. Although the bulk of these enhancers do not have validated associations with particular genes, we assumed that they would be relatively close to the genes that they drive. Of the 66 genes whose transcription start sites (TSSs) were the closest in either direction and within 10kb of the center of the tested CRM, only 32 were expressed at greater than 10FPKM in at least one slice of any of the wild-type embryos. Of these, only 10 had an

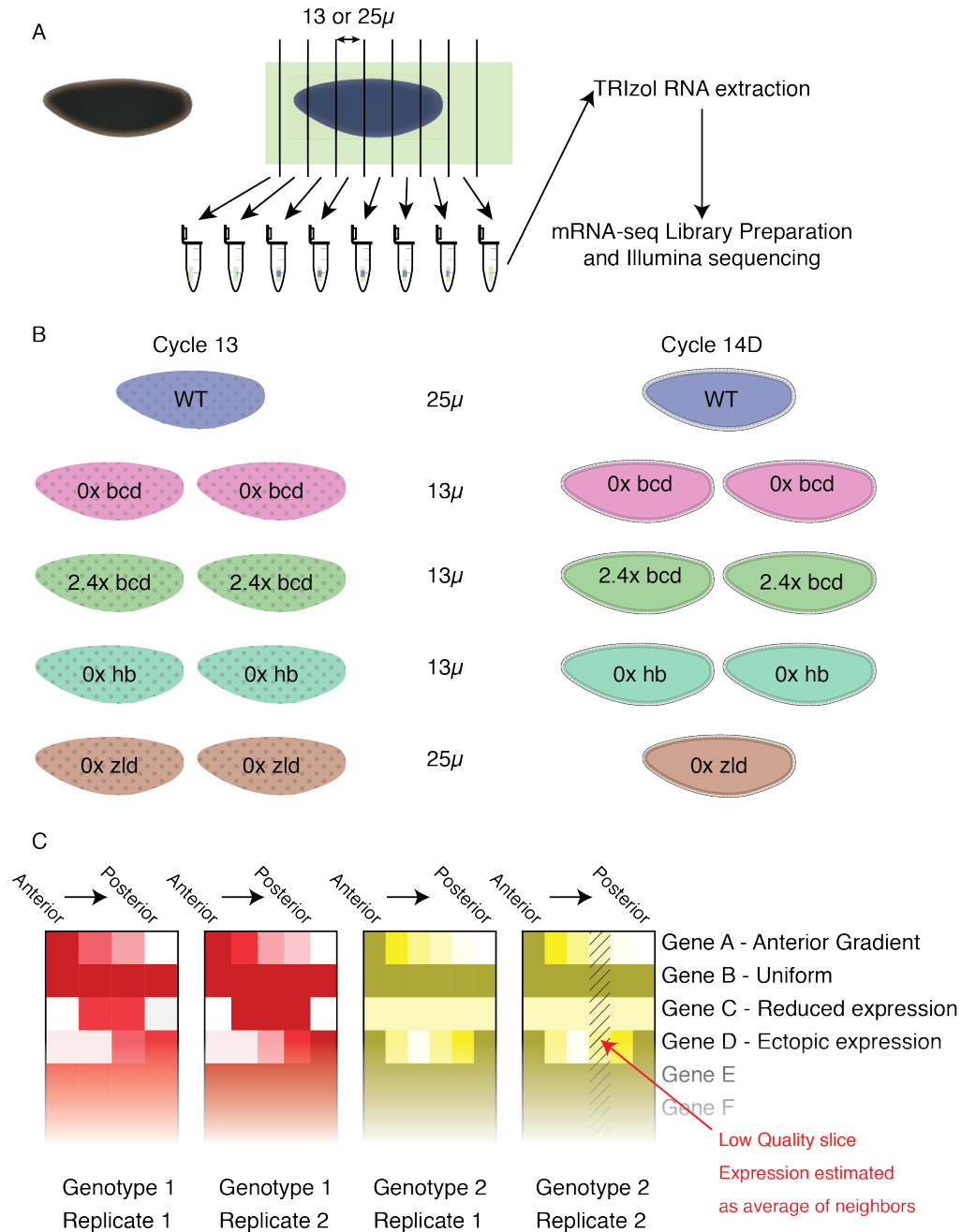


Figure 1. Schematic of experimental approach. **A)** We fixed embryos in methanol, then selected individual embryos at the correct stage, aligned them in sectioning cups, and sliced to the indicated thickness. We extracted RNA from individual slices, prepared barcoded libraries, then pooled them prior to sequencing. **B)** Overview of the mutant genotypes used. Two replicates per time point at two time points, based on nuclear density and morphology. **C)** Cartoon of heatmaps. Each genotype is assigned its own color (matching those in **B**), with darker colors representing higher expression and white representing no expression detected in that slice. Each boxed column represents a single individual, and within that column, slices are arranged posterior to the left and anterior to the right.

obvious anterior localization bias (31%), with the majority of the rest being approximately uniformly expressed across the embryo (Figure 5). The majority of genes with ubiquitous or central localization did not radically change in either the Bicoid overex-

pression or knockdown conditions. As expected, genes with anterior localization suffered a loss of patterning in the depletion mutant, and a posterior shift in the over-expression condition. We assume that genes that are not localized to the anterior are either driven by

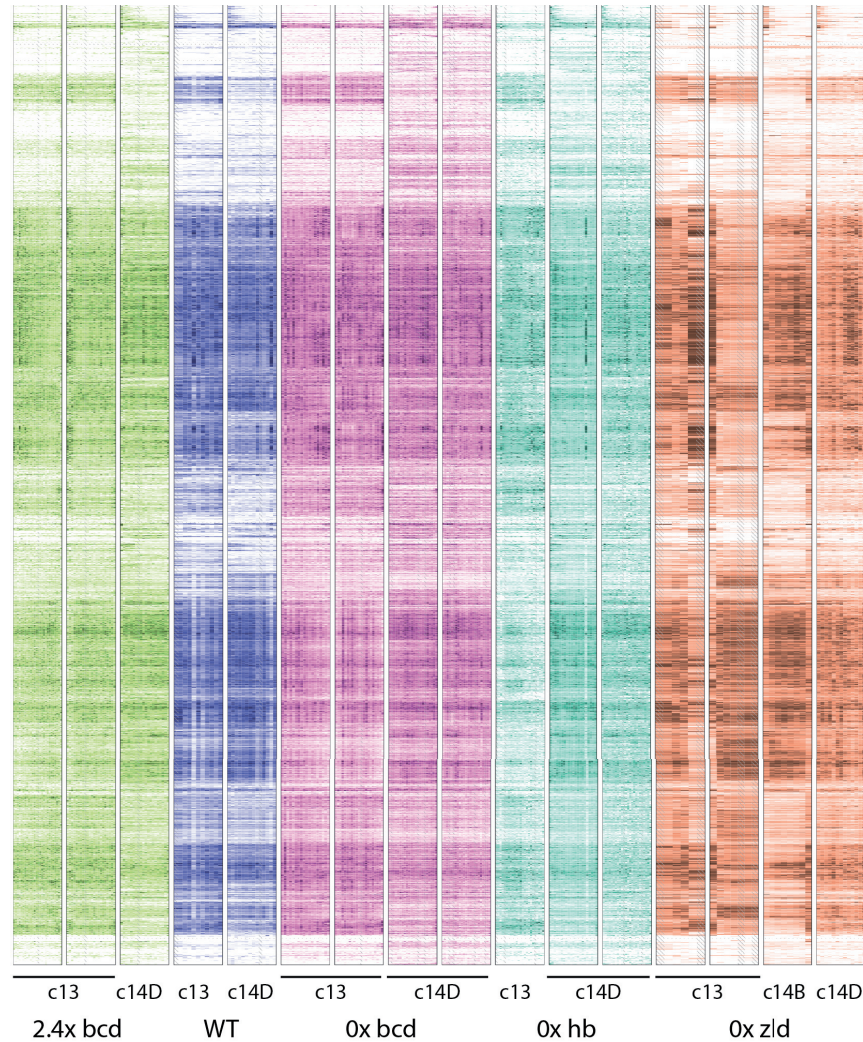


Figure 2. Heatmaps of gene expression patterns for all expressed genes. Each individual embryo is represented by one boxed column in the heatmap. Within a column, slices are arranged anterior to the left, and posterior to the right. Each embryo is colored according to genotype, with green for the *bcd* over-expression, blue for wild-type, magenta for *bcd* knockdown, cyan for *hb* knockdown, and orange for the *zld* mutant. Within a genotype, darker colors correspond to higher expression and white to zero expression, on a linear scale normalized for each gene separately to the highest expression for that gene in the embryo or to 10FPKM, whichever is greater. Slices that did not match quality control standards are replaced by averaging the adjacent slices, and are marked with hash marks. Rows are arranged by using Earth Mover's Distance to perform hierarchical clustering, so that genes with similar patterns across all of the embryos are usually close together.

multiple enhancers, such that loss of expression from one does not severely affect the overall expression, or that they are merely close to the enhancer, but unrelated.

Effects of TF depletion on patterned genes

We next sought to demonstrate that the technique of cryoslicing mutants is useful for identifying the effects of these early patterning genes. In comparison to Figure 3, where we looked for known patterning changes that we would expect from the literature, we also want to make sure that the largest and most common patterning changes that naturally arise from the data recapitulate the known literature. For each mutant genotype, we identified the 100 genes with the largest patterning change in that genotype, then averaged the pattern in all of the cycle 14 embryos (Figure 6).

Unsurprisingly, depletion of TFs known to be important for patterning are likely to make an otherwise non-uniform pattern more so (Figure 1). Of the 465 genes that have clearly non-uniform patterning in the wild-type at cycle 14D, 12–20% are affected in each depletion mutant, either losing expression entirely or becoming uniform. The over-expression line is at the low end of this range, also at 12%.

However, this is not always simply abrogating expression—a large number of genes seem to have higher expression everywhere. In the case of *bcd* depletion, approximately a third of these cases are genes that are restricted to the anterior in wild-type that become approximately uniform throughout the embryo (Figure 7). While some of these are due to genes with an early

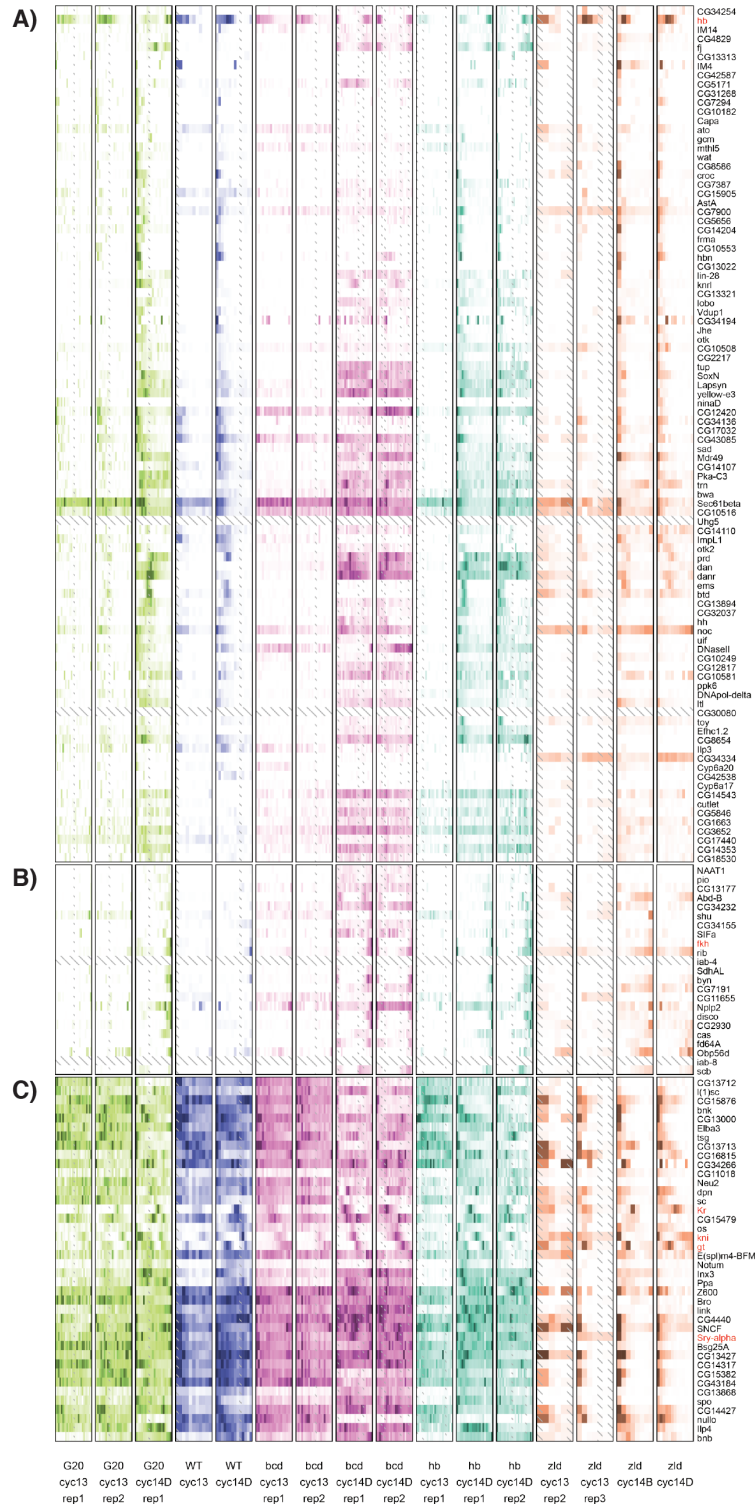


Figure 3. Heatmaps of gene expression patterns for anterior and posterior genes recapitulate expected patterning changes. We manually selected subsets of the larger heatmap in Figure 2 that showed clear differences between the genotypes. Each individual embryo is represented by one boxed column in the heatmap. Within a column, slices are arranged anterior to the left, and posterior to the right. Each embryo is colored according to genotype, with green for the *bcd* over-expression, blue for wild-type, magenta for *bcd* knockdown, cyan for *hb* knockdown, and orange for the *zld* mutant. Within a genotype, darker colors correspond to higher expression and white to zero expression, on a linear scale normalized for each gene to the highest expression for that gene in the embryo or to 10FPKM, whichever is greater. Genes identified in Liang *et al.*¹², Nien *et al.*⁵³, and Staller *et al.*¹⁶ as responsive to either Bcd or Zld have red labels. **A)** Genes with anterior expression in the wild type embryos. **B)** Genes with posterior expression in wild type embryos. **C)** Genes with broad expression in wild type and anterior expression in later stage *zld*-embryos.

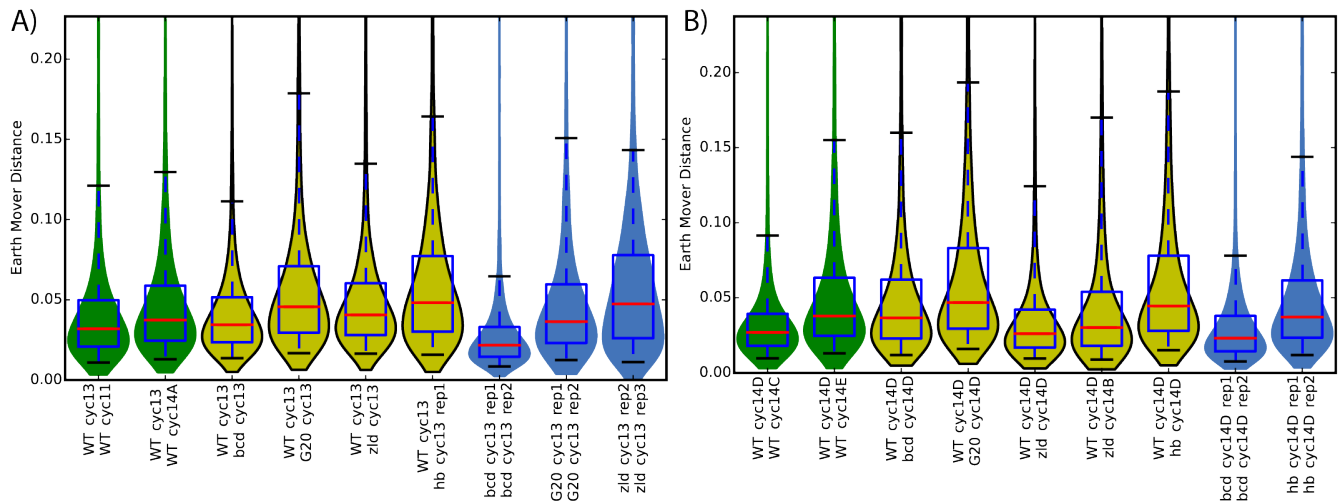


Figure 4. Distributions of patterning differences show that mutants have wide-spread subtle patterning effects and more genes with large patterning differences than replicates. Adjacent time points from the wild-type dataset in Combs and Eisen⁴⁴ are colored green, and replicates of the same genotype and time point are colored blue. Median distances are marked in red. **A)** Cycle 13 and **B)** mid cycle 14.

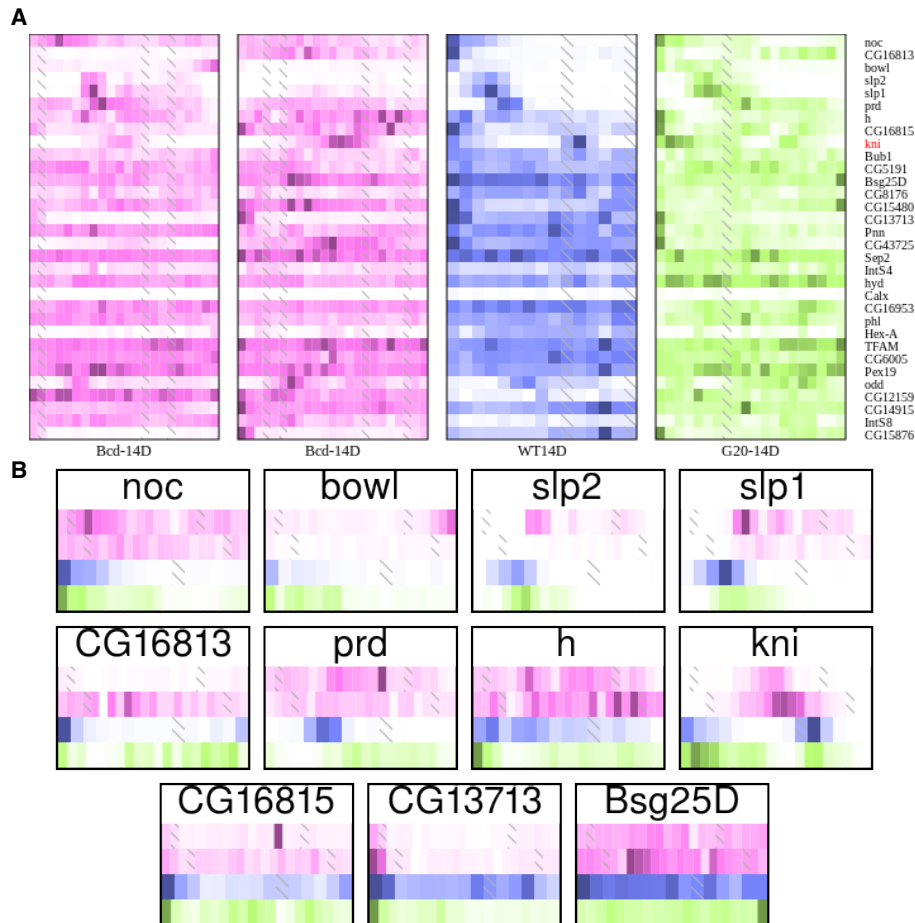


Figure 5. Patterning changes of genes near Bcd-dependent enhancers in *bcd* knockdown and overexpression are clearly visible in anterior-localized genes. **A)** Each individual is represented in its own heatmap. The magenta heatmaps are from the *bcd*-embryos, blue from wild-type, and green from 2.4x *bicoid*. **B)** Each gene with anterior localized expression in WT, with data from each individual as its own row, to highlight position changes across the mutant genotypes.

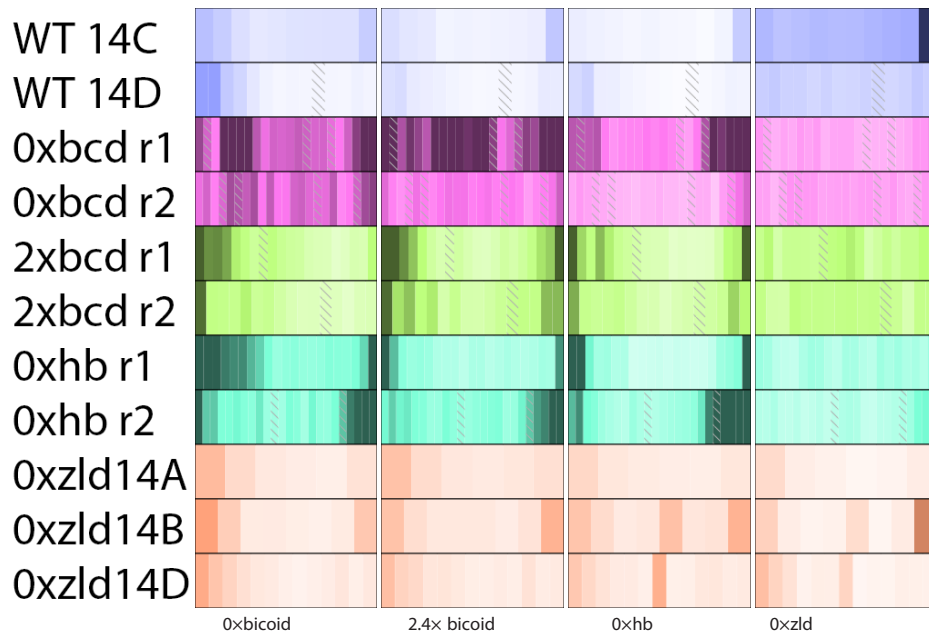


Figure 6. Averaging patterning changes in each genotype recapitulates known TF localization and function. Each individual is represented in its own heatmap. The magenta heatmaps are from the *bcd*- embryos, blue from wild-type, and green from 2.4x bicoid. A–D) The average pattern in each of the embryos of the 100 genes with the greatest change in *bcd* knockdown (A), *bicoid* overexpression (B), *hunchback* knockdown (C), and *zelda* knockout (D).

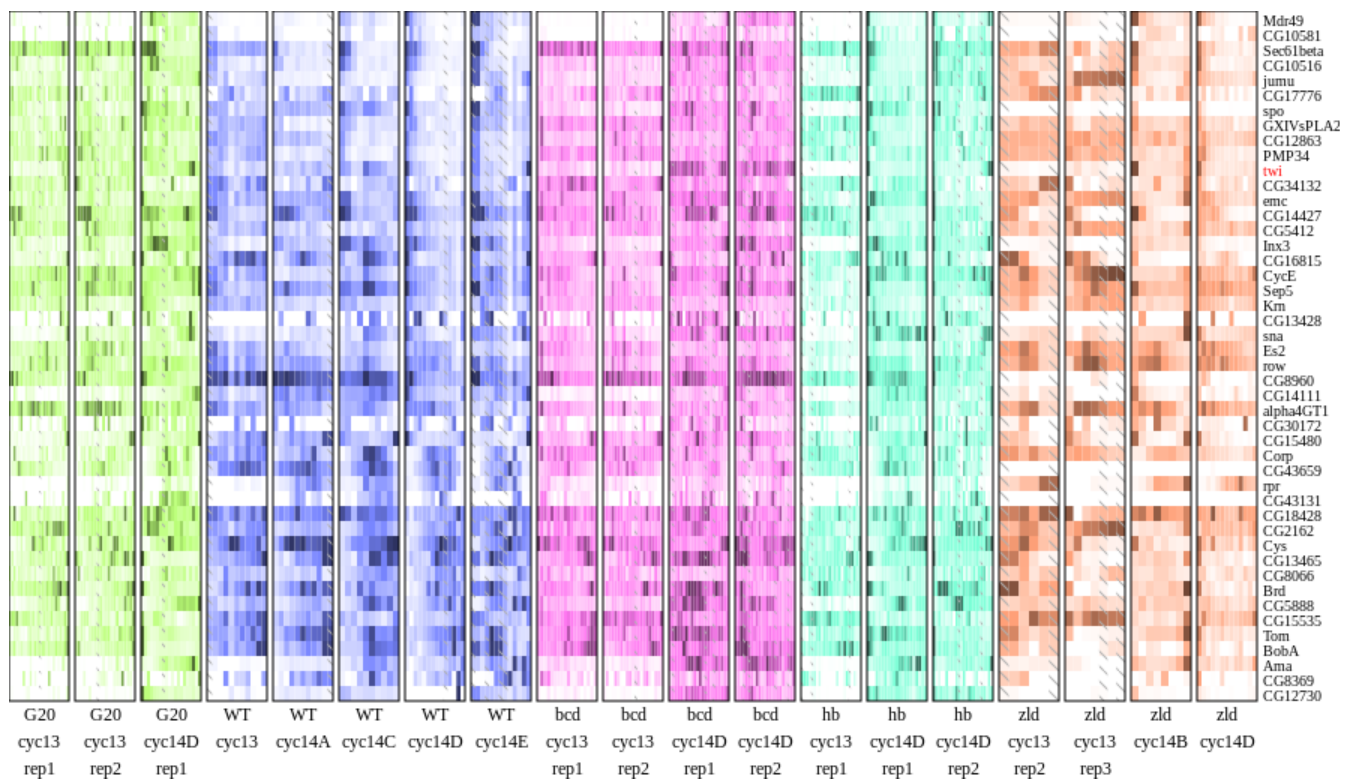


Figure 7. Patterned genes in wild-type that become uniformly expressed are widespread in *bcd*-. Each embryo is normalized independently.

uniform pattern that fails to properly resolve into spatially restricted domains, approximately half are true ectopic expression (Figure S2).

As a first step to identifying likely regulatory motifs, we used binding data for 9 non-pair-rule AP TFs²⁸ and for Zld¹⁰ to search for factors with differential rates of binding among the sets of genes with patterning changes (Figure 2). This analysis highlights that Zelda operates in a qualitatively different manner from the other transcription factors—in its absence other TFs are likely to continue expression, though in abnormal patterns. Additionally, Zld is crucial for maintaining patterned expression, as the most common change is from patterned genes integrating one or more AP factors to minimal overall expression.

Furthermore, *bicoid* stands out as a major factor involved in AP patterning. In all of the mutant conditions except *bcd* and *zld* depletion, having a Bcd binding site is associated with an increase in patterned expression. In all of the conditions, a Bcd binding site is associated with a ubiquitous becoming patterned, and this pattern is often anterior expression.

In addition to patterning changes, some genes with ubiquitous localization actually showed the same response in absolute level as a result of both *bcd* depletion and over-expression. Of these genes, 1002 showed at least 1.5 fold higher expression on both conditions, and 414 showed a 1.5 fold decrease in expression. Such a scenario suggests that these genes are, at wild-type levels, tuned to a particular level of Bicoid expression.

It is difficult to reconcile increases of expression in the posterior with any local model of transcription factor action. Bcd protein is only present at approximately 5nM at 50% embryo length, and negligible levels more posterior²⁹. It is conceivable that Bcd activates a repressor gene somewhere in the anterior, which then diffuses more rapidly than Bcd to cover at least some of the posterior of the embryo. Nevertheless, there have previously been hints that Bicoid can function as far to the posterior as *hairy* stripe 7³⁰.

Genes are likely to change in similar ways in different mutant conditions

We next asked whether patterning changes in one genotype could be used to predict whether the pattern changes in another. Therefore, we plotted the EMD between wild-type and the *bicoid* RNAi line on the X axis, and wild-type to the *zelda* GLC on the Y axis (Figure 8). Unsurprisingly, the majority of genes did not change, but of those that did, only a small fraction of them changed in one condition but not the other (the blue and green regions near the axes). We grouped genes according to whether they were in the top 20% of the EMD distribution for each genotype independently, then performed a Pearson's χ^2 test of independence of change in *bcd*- versus *zld*-. The result was highly significant ($p < 1 \times 10^{-100}$), with the largest overrepresentation coming from the case where both changed. Repeating this across all combinations of wild-type and two other mutant genotypes yielded the same results: in every case, there were between 2.2 to 2.7 times as many genes that changed in both categories as would be expected (Figure S3).

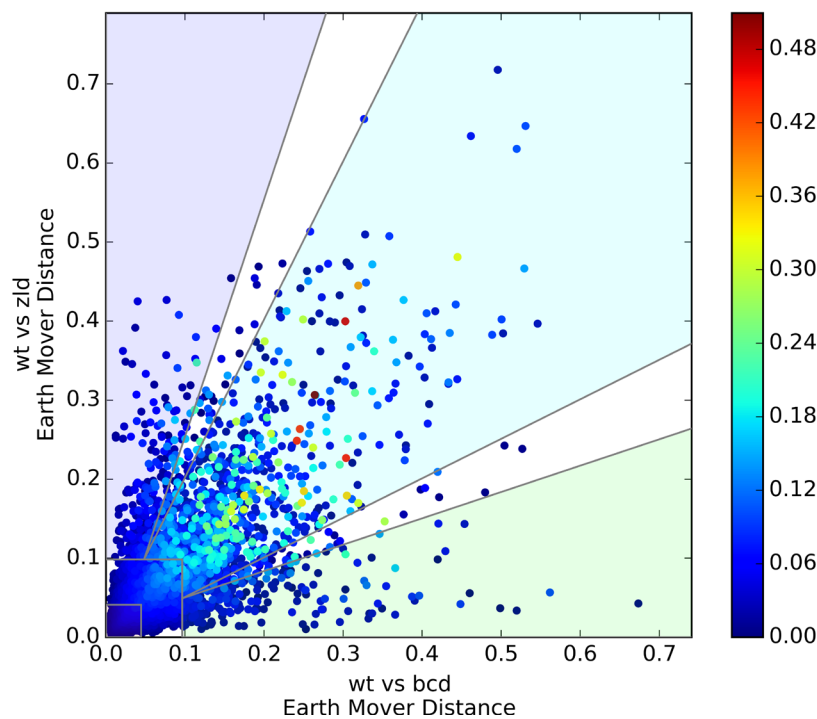


Figure 8. Genes that change in *bcd*- are likely to change in the same way in *zld*-, and vice-versa. Change versus the wild-type is plotted on the x and y axes. Each point is colored according to its ΔD score, calculated in Equation 1, in order to highlight genes that change differently between the two conditions.

Of these genes that do change in both conditions, the majority changed in effectively identical ways. We computed a modified EMD that down-weights genes that are very similar to wild-type in at least one mutant genotype:

$$\Delta D = \text{EMD}(M1, M2) - \frac{|\text{EMD}(M1, WT) - \text{EMD}(M2, WT)|}{2} \quad (1)$$

where $\text{EMD}(x, y)$ is the Earth Mover's Distance between identically staged embryos of genotype x and genotype y . Even among only the set of genes that change in both conditions, ΔD is small (mean of 3.5%, 95th percentile of 11.9%)—equivalent to a shift of the entire pattern by about 1 or 2 slices in either direction. However, there are 13 genes that change differently between wild-type, *bcd-*, and *zld-* ($\Delta D > 20\%$). These genes have noticeably different patterns in all three genotypes (Supplemental Figure S4).

Differential response to mutation is strongly associated with transcription factor binding

We sought to understand what is different about genes with a high ΔD , compared to those that change in response to wild-type,

but have a low ΔD (that is, those that change in the same way in response to distinct mutant conditions). We found that genes with a high ΔD score were strongly enriched for a number of TF binding sites (Table 3 and Table S1).

Table 1. TF depletion is more likely to make a non-uniform pattern uniform than vice versa.

	Low Expr to Patterned	Patterned to Uniform	Patterned to Low Expr	Uniform to Patterned
2.4×bcd	343	32	25	64
0×bcd	96	69	40	12
0×hb	348	36	20	43
0 ×zld	71	34	61	28

We measured EMD for each gene at cycle 14D in each genotype compared to a uniform distribution. We considered genes uniform if they had an $\text{EMD} < 0.04$, and non-uniform if they have an $\text{EMD} > 0.08$. We then considered genes with at least 15 FPKM in at least one slice in both wild-type and the mutant line.

Table 2. Patterning changes are strongly associated with increased TF binding.

	Low Expr to Patterned	Patterned to Uniform	Patterned to Low Expr	Uniform to Patterned
2.4×bcd	bcd,gt,kni,hkb,tll	—	—	bcd
0×bcd	gt,kni,tll	bcd,cad,gt	—	bcd,kni
0×hb	bcd,gt,kni,tll	—	—	bcd
0 ×zld	—	—	bcd,cad,gt,kni,tll	bcd,cad,gt,kni,tll

Using the genes with identified patterning changes in Figure 1, we performed a χ^2 test with a Bonferroni-corrected p-value of 0.05.

Table 3. TF binding is enriched near differentially changing genes between WT, *bcd-*, and *zld-*.

	odds ratio	base freq	p-value
bcd	4.69	17.40%	1.2e-09
kni	4.19	3.01%	0.000176
gt	3.79	19.00%	2.81e-08
cad	3.17	31.10%	6.73e-08
kr	3.1	56.58%	1.6e-07
tll	2.97	7.64%	0.000597
hb	2.21	39.53%	0.000143
hkb	2.1	23.86%	0.00115

χ^2 test results for TF binding within 10kb of the TSS for the wild-type/*bcd*-/*zld*- three-way comparison. We examined the top 50 genes by ΔD , compared to the 200 genes closest to the median ΔD of genes that change in response to both mutations. Base frequency indicates the fraction of genes expressed at this time point with at least one ChIP peak for that TF. In this comparison, only Dichaete and Zelda binding were not significant at a Bonferroni-corrected p-value of 0.05.

Next, we binned genes by ΔD score, then examined trends in combinatorial transcription factor binding. As ΔD score increases, genes are more likely to be bound by multiple TFs (Figure 9). Due to the high background rate of binding, assaying the presence of at least 3 factors is not readily able to distinguish between genes with high and low ΔD 's, as nearly 70% of all genes expressed have at least 3 TFs bound. Assaying for the presence of more factors is better able to identify which genes are likely to change, and the top 50 genes all have at least 8 factors bound.

We sought to understand the extent to which genes with the same pattern of upstream regulators had the same responses to perturbation. We grouped genes according to the complement of ChIP-validated TF binding sites near that gene, then examined the patterning changes. Although with 10 different TFs there are potentially over one thousand distinct combinations of binding patterns, in practice the dense, combinatorial patterns found around patterning enhancers reduces this set to a much more manageable 157 different combinations, of which only 52 had at least 30 genes.

Within these sets of genes with similar TF binding profiles, we then asked whether the distribution of patterning changes was any different from the distribution of patterning changes for all genes. For each gene, we summed the EMD scores for the $2.4 \times bcd$, bcd -, and zld -, then performed a KS-test between the summed EMD scores of genes with a given binding pattern and the summed scores for all expressed genes. We found only 2 binding patterns with a Bonferroni-corrected p-value less than .05. Both of these sets were highly bound, and they were also very similar to each other in their binding, differing only in the presence of a Knirps (Kni) binding site (Figure 10).

Despite the similar binding patterns near these genes, there is a wide range of responses. The wild-type expression patterns run nearly the complete gamut, including uniform expression, anterior stripes, posterior stripes, and central expression domains. Additionally, the presence of a Kni site seems to yield an increased number of genes with an anterior expression domain.

Discussion

We have generated a dataset that is unparalleled in its coverage assaying patterning changes in mutant conditions. When these patterning mutants have been described previously, either major morphological readouts like cuticle staining or *in situ* hybridization has been used to illustrate the effects on downstream target genes^{12,16,31}. However, *in situ* hybridization suffers from a strong selection bias in the genes that are chosen. By assaying spatial differences in the patterning of every gene in the genome, we demonstrate the full effect that these TFs have on developmental gene expression networks.

Despite the importance of the factors we chose for establishing spatially and temporally correct patterning, only a relatively small number of genes have significant expression pattern changes. Many of the targets that do show clearly abnormal expression patterns are, themselves, key transcription factors. This suggests that, even though key, maternally provided patterning factors bind to thousands of places throughout the genome²⁸, many of those binding sites are not functional in any meaningful sense. Certainly some of this binding is due to artifacts in the ChIP data, and even reproducible, non-artifactual binding should not be confused with function^{32,33}. However, the fact that genes near binding sites for multiple factors tend to have more complicated responses to mutation suggests that there is some truth to the idea that gene

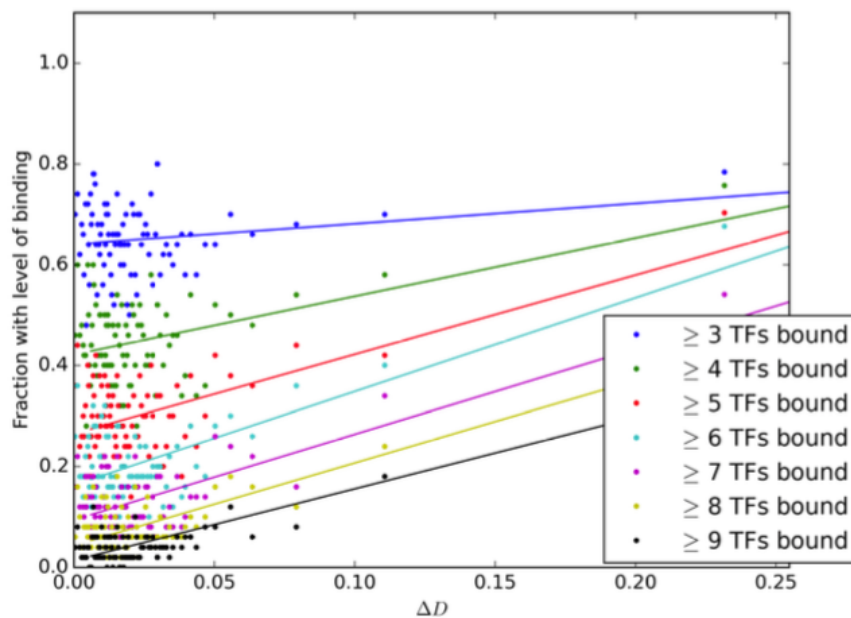


Figure 9. Higher ΔD scores are correlated with increased combinatorial binding. We grouped genes into non-overlapping windows of 50 genes by ΔD score, and calculated the fraction of those genes with at least 3, 4, ... etc. of the 10 early AP TFs bound (including Zld). We also plotted a simple linear regression on the binned points.

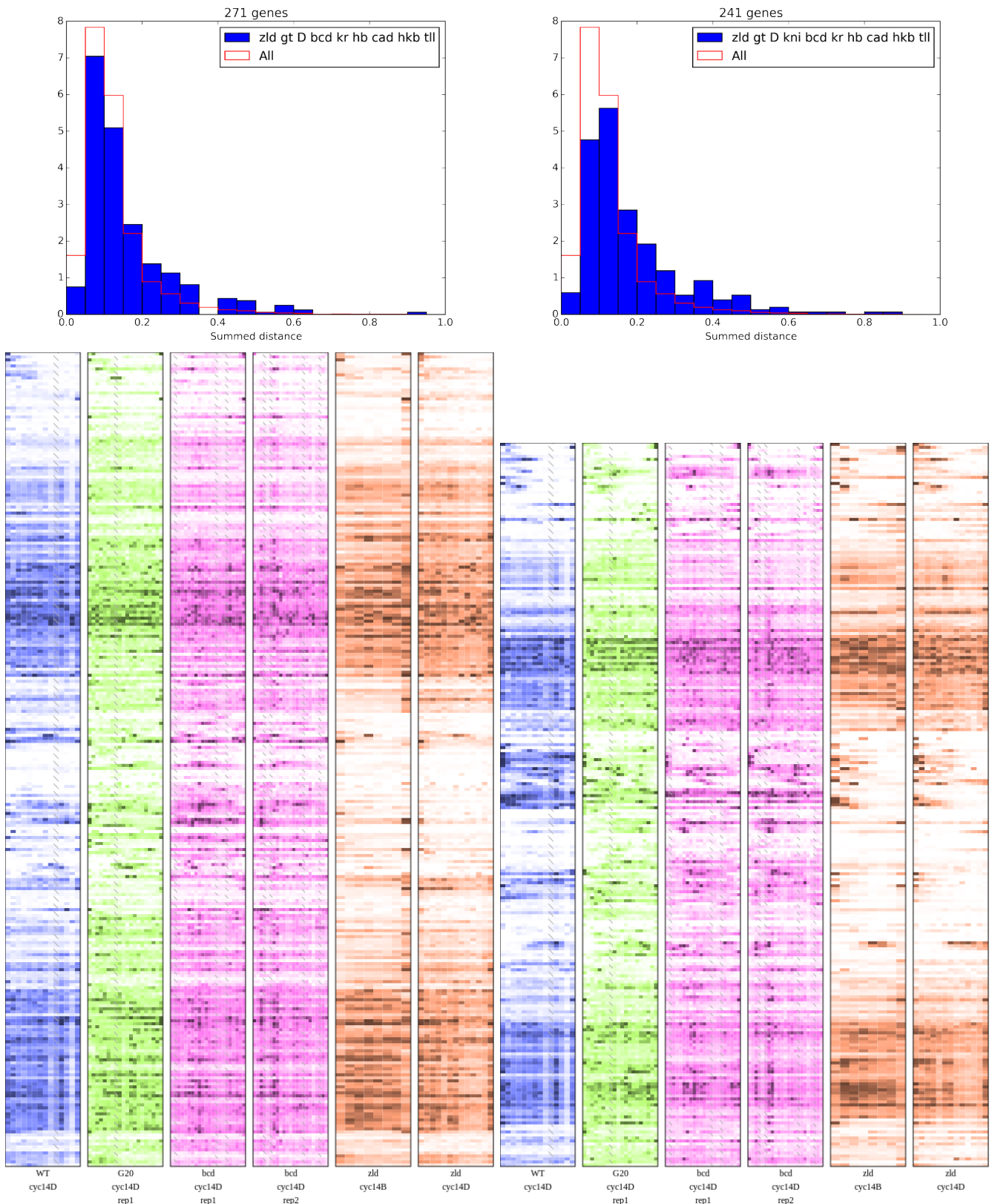


Figure 10. Identical binding patterns have a wide range of patterned responses.

regulation in complex animals tends to be combinatorial, even if the ChIP data are imperfect.

We were surprised how much proper *bicoid* expression seems to be required for proper patterning at all points along the embryo, not just in the anterior. The fact that *bcd* is normally understood to be an activator, while the plurality of genes with higher, ubiquitous expression in the mutant are normally localized to the anterior in wildtype suggests that this is normally mediated through one or more repressors that depend on *bcd*. As one of three TFs overrepresented at genes with this phenotype, *gt* is likely to be involved in this global derepression, but since it is itself neither ubiquitous throughout the embryo nor universally bound at the genes that change, it is likely not the only player.

The mutants we examined seemed to produce very similar changes in their downstream targets, despite the wild-type TFs having widely varying spatial distributions. Our initial expectation was that there would be many more ways to fail to properly pattern expression, and that different mutations would have different average effects from each other. Indeed, relying on different mutations having different responses has been the key to genetic analysis of fine scale patterns such as the *eve* stripes^{17,34–36}. Although averaging across the most different genes in a mutant genotype does yield different patterns (Figure 6), for any given gene excursions from the correct spatial expression pattern seems to be largely canalized (Figure 8). This seemingly-canalized expression change may be a consequence of the types of genes we can easily measure patterning changes among—we cannot resolve individual pair-rule stripes, for instance—so genes with coarser patterns may be more likely to have a single “failure” phenotype, as compared to those with finer patterns, which have more layers of regulation to perturb.

We do recognize a number of distinct limitations of this dataset towards predicting gene expression change as a function of mutation. The spatial resolution is still much coarser than *in situ* hybridization based experiments. This is especially concerning near regions where there are fine stripes of expression, which cannot be resolved between adjacent slices, or at regions where there is a transition between expression domains, where it is possible that the slicing axis is not perfectly aligned with the domain border. Finally, it is worth remembering that especially in the later stages examined, the gap gene positions will also be perturbed, so any observed changes in pattern positioning is likely to be a combination of direct effects and downstream effects of the original mutation.

A number of recent studies have used various technical or experimental techniques to improve the resolution of RNA-seq maps of gene expression in developing embryos. Iterated sectioning of different embryos in all three dimensions can be deconvolved to yield estimates of the original pattern³⁷. Similarly, sequencing mRNA from dissociated nuclei allows for the maximum possible spatial resolution, assuming the original location of those nuclei can be estimated^{38,39}. While these approaches are worthwhile for establishing a baseline map of expression patterns in wild-type embryos, the expense of sequencing still makes single-dimensional studies worthwhile. Furthermore, the single-cell approaches in Satija *et al.*³⁸ and Achim *et al.*³⁹ require some prior knowledge of spatial

gene expression, which may be significantly perturbed in patterning mutants. Other approaches for multiplexed *in situ* profiling of mRNA abundance have been described, but are not yet cheap or reliable enough to be readily useful for screening mutants^{40,41}.

Additionally, the time and expense required for a single individual necessarily means that we have profiled only a small number of individuals. We were therefore careful to choose only highly penetrant mutations for analysis, and to choose individuals at as similar staging as we could. However, even for genes with a consistent, precise time-dependent response between individuals, the differences in staging are likely to be a significant contributor to variation. Furthermore, we only examined two relatively distant time points in this study (approximately 45 minutes apart), making comparisons across time fraught at best.

Nevertheless, this experiment suggests a number of genes for more detailed follow up studies. As our predictive power for relatively well-studied model systems, such as the *eve* stripes improves, it will be especially important to take these insights to other expression patterns in the embryo. The risk of over-fitting increases with the depth of study of any particular model system, even if any given study is relatively well controlled. Therefore, by demonstrating that particular insights hard-won in these model systems are broadly applicable, we can gain some confidence in the results, and we approach having a rigorous, broadly applicable predictive model of gene regulation.

Ultimately, we believe more datasets addressing chromatin state in response to different conditions will be necessary for accurate prediction of spatial responses to mutation. In a ChIP-seq dataset on embryos with different, uniform levels of *bicoid* expression, hundreds of peaks seem to vary with differing affinities to Bcd protein (Colleen Hannon and Eric Wieschaus, personal communication, March 2015). The zygotically expressed genes near these differential peaks also have different spatial localization in wild-type, and different average responses to the mutants presented here. In addition to spatially resolved expression measurements, spatially resolved binding and chromatin accessibility data will likely be necessary. While ChIP-seq experiments currently require several orders of magnitude more input material than can be reasonably collected from spatially resolved samples, recent method developments in measuring chromatin accessibility have shown that it is possible to collect data from as few as 500 mammalian nuclei⁴². A similar amount of DNA is present in a single *Drosophila* embryo, which suggests that spatially resolved chromatin accessibility data may be achievable.

Materials and methods

Fly lines, imaging, and slicing

Zelda germline clone flies (w *zld*-FRT/FM7a; His2Av RFP) were a gift of Melissa Harrison, and were mated and raised as described previously¹⁰. Embryos were collected from mothers 3–10 days old.

The construction of the *bcd* and *hb* RNAi flies has been described previously⁴³ and were obtained from the DePace Lab at Harvard Medical School. Briefly, we generated F1s from the cross of maternal tubulin Gal4 mothers (line 2318) with

UAS-shRNA-*bcd* or UAS-shRNA-*hb* fathers (lines GL00407 and GL01321 respectively), then collected embryos from the sibling-mated F1s. In order to take advantage of the slowed oogenesis and resulting greater RNAi efficiency, we aged the F1 mothers for approximately 30 days at 25°C.

The *bcd* overexpression lines were a generous gift of Thomas Gregor at Princeton University. We used line 20, which has 2.4× wild-type levels of eGFP-Bcd fusion. Flies were kept in uncrowded conditions, and embryos were collected at 25°C from 3–7 day old mothers.

We washed, dechorionated, and fixed the embryos according to our standard protocol (see 44), incubated in 3 μM DAPI for 5 minutes, washed twice with PBS, and then imaged on a Nikon 80i microscope with a Hamamatsu ORCA-Flash4.0 CCD camera. We did not DAPI stain the *zld*-embryos because they had a histone RFP marker. After selecting embryos with the appropriate stage according to density of nuclei in histone-RFP or DAPI staining and membrane invagination for the cycle 14 embryos, we washed embryos with methanol saturated with bromophenol blue (Fisher), aligned them in standard cryotome cups (Polysciences Inc), covered them with VWR Clear Frozen Section Compound (VWR, West Chester, PA), and froze them at -80C.

We sliced the embryos as in Combs and Eisen⁴⁴. Single slices were placed directly in non-stick RNase-free tubes (Life Technologies), and kept on dry ice until storage at -80C.

RNA extraction, library preparation, and sequencing

We performed RNA extraction in TRIzol as previously⁴⁴. All RNA quality was confirmed using a BioAnalyzer 2100 RNA Pico chip (Agilent).

We generated libraries of the *zld*-embryos using the TruSeq mRNA unstranded kit (Illumina). As described previously, we added in 70 ng of yeast total RNA as a carrier and performed reactions in half-sized volumes to improve concentration⁴⁴.

We generated libraries from the RNAi and overexpression embryos using the SMARTseq2 protocol; we skipped the cell lysis steps because RNA had already been extracted^{45,46}. As described previously, tagmentation steps were performed at 1/5th volume to reduce costs⁴⁶.

Data analysis and deposition

All data was compared to FlyBase genome version r6.03 (ftp://ftp.flybase.net/releases/FB2014_06/). Mapping was performed using RNA-STAR v2.3.0.1⁴⁷, and expression estimates were generated using Cufflinks v2.2.1 on only the *D. melanogaster* reads⁴⁸. Reads from Combs and Eisen⁴⁴ were re-mapped to the new genome version. When carrier RNA was used (data from Combs and Eisen⁴⁴ and the *zld*-embryos), we discarded as ambiguous reads with 3 or fewer mismatches to prefer one species or the other. Due to the extensive divergence between the yeast carrier RNA and fly target RNA, the vast majority of mapped reads (>99.99%) were unambiguous as to the species of origin. After mapping, we removed

samples that had fewer than 500,000 *D. melanogaster* reads and samples with less than a 70% mapping rate when no carrier RNA was used; no other filtering or corrections were performed.

Specific analysis code was custom-written in Python 2.7.6. Custom analysis and heatmap generation code is available from <https://github.com/petercombs/EisenLab-Code>. All analyses presented here and all data figures were made using commit 2c144be (doi: [10.5281/zenodo.160787](https://doi.org/10.5281/zenodo.160787)). EMDs were calculated using the python-emd package by Andreas Jansson (no version number available, version used archived under doi: [10.5281/zenodo.160797](https://doi.org/10.5281/zenodo.160797)). Violin plots, histograms, and scatter plots were made using Matplotlib v1.4.2 and Numpy v1.9.2,⁴⁹⁻⁵¹. Linear regressions (`stats.linregress`), Kolmogorov-Smirnov tests (`stats.ks_2samp`), and χ^2 tests (`stats.chi2_contingency`) were performed using Scipy v 0.14.0.

Newly generated sequencing reads have been deposited at the Gene Expression Omnibus under accession GSE71137. Additional files and a searchable database are available at <http://eisenlab.org/supplements/combs2016/>.

Data availability

Newly generated sequencing reads have been deposited at the Gene Expression Omnibus under accession GSE71137 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71137>).

Custom analysis code: <https://github.com/petercombs/EisenLab-Code>

Archived custom analysis code at the time of publication: <http://dx.doi.org/10.5281/zenodo.160787>

Author contributions

Conceived and designed the experiments: PAC MBE. Performed the experiments: PAC. Analyzed the data: PAC. Algorithms used in analysis: PAC. Contributed reagents/materials/analysis tools: PAC MBE. Wrote the paper: PAC.

Competing interests

MBE is on the International Advisory Board for F1000.

Grant information

This work was supported by a Howard Hughes Medical Institute investigator award to MBE. PAC was supported by the National Institutes of Health under a Genomics Training Grant (5T32HG000047-13).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to thank Steven Brenner, Han Lim, and Lior Pachter for helpful comments on the manuscript as part of PAC's dissertation.

Supplementary material

Supplemental Figure 1: Adjacent slices are more similar than distant ones. Violin plots of the Spearman Rank correlations between adjacent slices and pairs of slices separated by more than one third of the embryo length.

[Click here to access the data](#)

Supplemental Figure 2: Figure 7 normalized to expression in wild-type cycle 14D highlights absolute expression level changes. Slices with higher expression are clipped to the maximum expression in wild-type.

[Click here to access the data](#)

Supplemental Figure 3: Genes that change tend not to change in only one condition. Three-way comparisons, as in Figure 8, between wildtype and the remaining combinations of *bcd* depletion, *bcd* overexpression, *hp* depletion, and *zld* depletion. The wildtype-mutant comparison is indicated on each axis, and the color indicates the ΔD score between the two mutants. G20 indicates the bicoid overexpression line, line #20 from 52.

[Click here to access the data](#)

Supplemental Figure 4: Only a handful of genes change differently between the different conditions. In the WT vs *bcd*- vs *zld*- three-way comparison, only 13 genes had a ΔD score above 20%. Thumbnails indicate wild-type pattern in blue, *bcd*- pattern in both replicates in pink, and *zld*- pattern in orange. All expression is scaled to the highest in each individual.

[Click here to access the data](#)

Supplemental Table 1: TF binding is enriched near differentially changing genes across all three-way comparisons. χ^2 test results for TF binding within 10kb of the TSS for the indicated three-way comparison. We examined the top 50 genes by ΔD , compared to the 200 genes closest to the median ΔD of genes that change in response to both mutations. Base frequency indicates the fraction of genes with at least one ChIP peak for that TF and that are expressed at this time point in all three conditions.

[Click here to access the data](#)

References

- Berleth T, Burri M, Thoma G, *et al.*: **The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo.** *EMBO J.* 1988; 7(6): 1749–1756.
[PubMed Abstract](#) | [Free Full Text](#)
- Driever W, Nüsslein-Volhard C: **The bicoid protein is a positive regulator of hunchback transcription in the early *Drosophila* embryo.** *Nature.* 1989; 337(6203): 138–143.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kraut R, Levine M: **Spatial regulation of the gap gene *giant* during *Drosophila* development.** *Development.* 1991; 111(2): 601–609.
[PubMed Abstract](#)
- Small S, Kraut R, Hoey T, *et al.*: **Transcriptional regulation of a pair-rule stripe in *Drosophila*.** *Genes Dev.* 1991; 5(5): 827–839.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sun Y, Nien CY, Chen K, *et al.*: **Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation.** *Genome Res.* 2015; 25(11): 1703–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xu Z, Chen H, Ling J, *et al.*: **Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in *Drosophila*.** *Genes Dev.* 2014; 28(6): 608–621.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li XY, Harrison MM, Villalta JE, *et al.*: **Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition.** *eLife.* 2014; 3: e03737.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Satija R, Bradley RK: **The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the *Drosophila* embryo.** *Genome Res.* 2012; 22(4): 656–665.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kanodia JS, Liang HL, Kim Y, *et al.*: **Pattern formation by graded and uniform signals in the early *Drosophila* embryo.** *Biophys J.* 2012; 102(3): 427–433.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harrison MM, Li XY, Kaplan T, *et al.*: **Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition.** *PLoS Genet.* 2011; 7(10): e1002266.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harrison MM, Botchan MR, Cline TW: **Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed *Drosophila* genes.** *Dev Biol.* 2010; 345(2): 248–255.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liang HL, Nien CY, Liu HY, *et al.*: **The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*.** *Nature.* 2008; 456(7220): 400–403.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wimmer EA, Carleton A, Harjes P, *et al.*: **Bicoid-independent formation of thoracic segments in *Drosophila*.** *Science.* 2000; 287(5462): 2476–2479.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jaeger J: **The gap gene network.** *Cell Mol Life Sci.* 2011; 68(2): 243–274.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Frohnhofer HG, Nüsslein-Volhard C: **Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid.** *Cell.* 1986; 324(2): 120–125.
[Publisher Full Text](#)
- Staller MV, Fowlkes CC, Bragdon MD, *et al.*: **A gene expression atlas of a bicoid-depleted *Drosophila* embryo reveals early canalization of cell fate.** *Development.* 2015; 142(3): 587–596.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Small S, Blair A, Levine M: **Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo.** *Dev Biol.* 1996; 175(2): 314–324.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tadros W, Lipshitz HD: **The maternal-to-zygotic transition: a play in two acts.** *Development.* 2009; 136(18): 3033–3042.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lott SE, Villalta JE, Schroth GP, *et al.*: **Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq.** *PLoS Biol.* 2011; 9(2): e1000590.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

20. Rubner Y, Tomasi C, Guibas LJ: **A metric for distributions with applications to image databases.** In: *Computer Vision, 1998. Sixth International Conference on 1998*; 59–66.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Chen H, Xu Z, Mei C, *et al.*: **A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes.** *Cell.* 2012; **149**(3): 618–629.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Ochoa-Espinosa A, Yu cel G, Kaplan L, *et al.*: **The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*.** *Proc Natl Acad Sci U S A.* 2005; **102**(14): 4960–4965.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Schroeder MD, Pearce M, Fak J, *et al.*: **Transcriptional control in the segmentation gene network of *Drosophila*.** *PLoS Biol.* 2004; **2**(9): E271.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Biemar F, Zinzen R, Ronshaugen M, *et al.*: **Spatial regulation of microRNA gene expression in the *Drosophila* embryo.** *Proc Natl Acad Sci U S A.* 2005; **102**(44): 15907–15911.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Hartmann B, Reichert H, Walldorf U: **Interaction of gap genes in the *Drosophila* head: *tailless* regulates expression of *empty spiracles* in early embryonic patterning and brain development.** *Mech Dev.* 2001; **109**(2): 161–172.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Kantorovitz MR, Kazemian M, Kinston S, *et al.*: **Motif-blind, genome-wide discovery of cis-regulatory modules in *Drosophila* and mouse.** *Dev Cell.* 2009; **17**(4): 568–579.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Riddihough G, Ish-Horowitz D: **Individual stripe regulatory elements in the *Drosophila* hairy promoter respond to maternal, gap, and pair-rule genes.** *Genes Dev.* 1991; **5**(5): 840–854.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. MacArthur S, Li XY, Li J, *et al.*: **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.* 2009; **10**(7): R80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Gregor T, Tank DW, Wieschaus EF, *et al.*: **Probing the limits to positional information.** *Cell.* 2007; **130**(1): 153–164.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. La Rosée A, Häder T, Taubert H, *et al.*: **Mechanism and Bicoid-dependent control of *hairy* stripe 7 expression in the posterior region of the *Drosophila* embryo.** *EMBO J.* 1997; **16**(14): 4403–4411.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Driever W, Nüsslein-Volhard C: **The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner.** *Cell.* 1988; **54**(1): 95–104.
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Graur D, Zheng Y, Price N, *et al.*: **On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE.** *Genome Biol Evol.* 2013; **5**(3): 578–590.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Teytelman L, Thurtle DM, Rine J, *et al.*: **Highly expressed loci are vulnerable to misleading CHIP localization of multiple unrelated proteins.** *Proc Natl Acad Sci U S A.* 2013; **110**(46): 18602–18607.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Frasch M, Levine M: **Complementary patterns of *even-skipped* and *fushi tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*.** *Genes Dev.* 1987; **1**(9): 981–995.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Frasch M, Warrior R, Tugwood J, *et al.*: **Molecular analysis of *even-skipped* mutants in *Drosophila* development.** *Genes Dev.* 1988; **2**(12B): 1824–1838.
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Andrioli LP, Vasishth V, Theodosopoulou E, *et al.*: **Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms.** *Development.* 2002; **129**(21): 4931–4940.
[PubMed Abstract](#)
37. Junker JP, Noël ES, Guryev V, *et al.*: **Genome-wide RNA Tomography in the zebrafish embryo.** *Cell.* 2014; **159**(3): 662–675.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Satija R, Farrell JA, Gennert D, *et al.*: **Spatial reconstruction of single-cell gene expression data.** *Nat Biotechnol.* 2015; **33**(5): 495–502.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Achim K, Pettit JB, Saraiva LR, *et al.*: **High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin.** *Nat Biotechnol.* 2015; **33**(5): 503–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Lee JH, Daugharthy ER, Scheiman J, *et al.*: **Highly multiplexed subcellular RNA sequencing *in situ*.** *Science.* 2014; **343**(6177): 1360–1363.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Chen KH, Boettiger AN, Moffitt JR, *et al.*: **RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells.** *Science.* 2015; **348**(6233): aaa6090.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Buenrostro JD, Giresi PG, Zaba LC, *et al.*: **Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.** *Nat Methods.* 2013; **10**(12): 1213–1218.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Staller MV, Yan D, Randklev S, *et al.*: **Depleting gene activities in early *Drosophila* embryos with the “maternal-Gal4-shRNA” system.** *Genetics.* 2013; **193**(1): 51–61.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Combs PA, Eisen MB: **Sequencing mRNA from cryo-sliced *Drosophila* embryos to determine genome-wide spatial patterns of gene expression.** *PLoS One.* 2013; **8**(8): e71820.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Picelli S, Faridani OR, Björklund AK, *et al.*: **Full-length RNA-seq from single cells using Smart-seq2.** *Nat Protoc.* 2014; **9**(1): 171–181.
[PubMed Abstract](#) | [Publisher Full Text](#)
46. Combs PA, Eisen MB: **Low-cost, low-input RNA-seq protocols perform nearly as well as high-input protocols.** *PeerJ.* 2015; **3**: e869.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Dobin A, Davis CA, Schlesinger F, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013; **29**(1): 15–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Trapnell C, Roberts A, Goff L, *et al.*: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc.* 2012; **7**(3): 562–578.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Jones E, Oliphant T, Peterson P, *et al.*: **SciPy: Open source scientific tools for Python.** 2001.
[Reference Source](#)
50. Hunter JD: **Matplotlib: A 2D graphics environment.** In: *Computing In Science & Engineering.* 2007; **9**(3): 90–95.
[Publisher Full Text](#)
51. McKinney W: **Data Structures for Statistical Computing in Python.** In: *Proceedings of the 9th Python in Science Conference.* Ed. by Stéfan van der Walt and Jarrod Millman. 2010; 51–56.
[Reference Source](#)
52. Liu F, Morrison AH, Gregor T: **Dynamic interpretation of maternal inputs by the *Drosophila* segmentation gene network.** *Proc Natl Acad Sci U S A.* 2013; **110**(17): 6724–6729.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Nien CY, Liang HL, Butcher S, *et al.*: **Temporal coordination of gene networks by *Zelda* in the early *Drosophila* embryo.** *PLoS Genet.* 2011; **7**(10): e1002339.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 22 February 2017

doi:10.5256/f1000research.10478.r19328



Angela H. DePace

Department of Systems Biology, Harvard University, Boston, MA, USA

This paper by Combs and Eisen examines spatially patterned gene expression in *Drosophila* embryos in response to mutation of key developmental transcription factors. The unique contribution of this work is the scale of their measurement — mRNA expression assessed genome-wide in slices along the anterior/posterior axis of individual embryos. Typically, genome-wide measurements are not spatially resolved; conversely spatially resolved measurements are typically only made for a handful of genes at once. This technique was described and validated in a previous paper. Here they apply it to characterize the response to mutations of bicoid, zelda and hunchback.

The dataset is valuable and the underlying data collection is well described. There results also point to a number of interesting conclusions. In particular, it's notable that despite being bound widely across the genome, knocking down these TFs produces only modestly affects patterning on average. It is also notable that Bicoid knockdown can affect patterning in the posterior, where it is not expressed. My suggestions for improving the paper are below.

1. The introduction does not introduce a biological question that they will be able to address with their method, but instead focuses on the novelty of the measurement. They state that “Given the crucial roles of each of these factors in spatial patterning, we expected that perturbing their levels would lead to widespread direct and indirect effects on patterned genes.” It would be useful to expand on this hypothesis in terms specific to the data they will collect and how they will analyze it. Which features of the data would confirm or refute this hypothesis? What will they find interesting or surprising? For example, they do some of this in the discussion. “Our initial expectation was that there would be many more ways to fail to properly pattern expression, and that different mutations would have different average effects from each other.” Can their expectation be made more specific given the diversity of patterns they can detect in WT embryos? Why did they have this expectation about the average effects when so many genes are unpatterned? Bringing some of the more interesting points in the discussion to the introduction to frame the paper would motivate the reader to understand the various analyses.
2. The quality, reproducibility and structure of the data should be more explicitly discussed, and used to explain their choice of analytical strategy. For example, they justify their use of the Earth Mover's Distance to examine expression patterns: “Because there are a different number of slices both between the wild type and mutant flies, and between replicates of the mutant flies...” Does this sentence mean that there are more slices per embryo for a WT versus a mutant embryo, or that there are in aggregate more mutant slices than WT slices (because they measured more mutants)? How does this impact their choice of analyses? Given the structure of the data and its

variability, what are the limits of analysing effects on single genes (which would be more familiar to developmental biologists)? Simple language could help more readers appreciate the utility and limits of their data for future analyses.

In addition, a narrative description of the EMD would be helpful. The authors state : "...we decided to use Earth Mover's Distance (EMD) to compare patterns[20]. This metric captures intuitive notions about what kinds of patterns are dissimilar, yielding higher distances for dissimilar distributions of RNA, and zero for identical distributions." This states that the range will go from 0 (identical) to "higher" (dissimilar) but does not justify their use of this metric. It would be useful to explain the value of this choice of metric over available alternatives.

3. The figure legends and titles could be more informative. In general, the legends state only what is depicted and does not point the reader to the most relevant comparisons. While this isn't strictly necessary, it can help readers who are unfamiliar with the particular type of plot used. Figure 6 refers to panels A - D, which are not labeled in the figure. The legend of Figure 7 is a single sentence stating "Each embryo is normalized separately." In some cases, the grammar of the titles is difficult to parse. For example, the title of Figure 4 is: "Distributions of patterning differences show that mutants have wide-spread subtle patterning effects and more genes with large patterning differences than replicates." The clause of this sentence is unclear. The title of Figure 5 is similarly difficult to parse.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 20 February 2017

doi:[10.5256/f1000research.10478.r19327](https://doi.org/10.5256/f1000research.10478.r19327)



Thomas Gregor

Department of Physics, Princeton University, Princeton, NJ, USA

The paper by Combs and Eisen extends the previously published dataset of RNA-seq measurements in cryo-sliced *Drosophila* embryos to mutants of patterning genes. As the authors state, most common methods for gene expression measurements are either limited in throughput or average over spatial coordinates. The development and application of an experiential approach addressing this gap is therefore of high interest. Datasets produced with this approach are an extremely valuable resource to diverse studies in the community. In the presented application the author go beyond the characterization of wild type embryos to the examination of *Zelda*, *Hunchback* and *Bicoid* mutants. This study thus offers the opportunity to examine the contribution of these key factors to the formation of spatial patterns of gene expression in the early embryo. That said, the description of the analyses carried out and the presentation of the results could be improved to further facilitate the usefulness of this dataset and of the approach.

While the analyses focuses on comparison between samples (and here a single slice from a single embryo at a single time point can be considered a sample) the authors provide little information on how these comparisons are done. It would be beneficial to specifically note what type of assumptions are made in the analysis, how samples are normalized with respect to one another and how these

normalizations influence the quantitative nature of the statements that can be made. For example, does the analysis performed implicitly assume that the total number of RNA produced (and captured) in each slice is the same? Can the authors address more extensively how they deal with differences in the exact number and size of slices. As patterns are then normalized to their maximal level prior to comparison between conditions, it seems genes are not assumed to maintain the same maximal activity yet claims about absolute changes in level cannot be made; can the authors clarify if this is indeed the case. Given these issues, it would be useful to include a more explicit description how a “uniform” vs a “patterned” gene is defined. Maybe the authors could show this graphically using an average plot of all of the heat map data for the genes classified this way. A clarification of the quantitative meaning of phrases as a “greatly perturbed pattern”, “Higher expression everywhere” or “same response in absolute level”, would help elucidate the insights obtained from the data.

It would also be useful if the authors provided a more extensive explanation of their usage of EMD, a method that was not employed in their analysis in their 2013 paper. This can be included in the methods/supplemental information. How would the similarity in the changes between genes/mutants ascertained with this measure compare to other measures (e.g. correlating patterns or changes in patterns). In Figure 10, could the summation of the EMD scores mask some of the signal? (is the result the same for each EMD score?). What is the initial diversity in patterns among genes that belong to the same set in this analysis (i.e. sharing a similar ChIP binding pattern)? The authors present figure 10 as an investigation of how genes with binding sites for the same TFs respond to perturbation (assuming “patterned responses” means responses to mutation), but only discuss their wild-type patterns. This is a good example of the heatmaps having too much information to be illustrative. It is difficult to relate the EMD to physical changes in patterns of target genes. It's not obvious, based on the barplot and the heatmaps in the figure, what it is about these expression patterns that resulted in these two groups of genes being significantly different from each other. Maybe a better explanation of the EMD would clarify this. It is not directly clear mathematically how EMD measures how quantitatively different two groups are. For example, only two cohorts of TFs showed a significant difference in the EMD of the genes in their group; the histograms show the distributions of EMDs in genes with each cohort of TF binding sites vs. the EMDs in the total dataset; it looks like the two significantly different groups are both skewed toward higher EMDs than the total dataset, but it is unclear what makes them different from *each other*, based on the figure and the explanations provided.

Additionally, despite showing the replicates in most figures, I still find it hard to assess to what extent are patterns variable between replicates (due to both biological and technical reasons) and what is the magnitude of these differences compared to those observed in the mutants. For instance in Figure 4A, do the *zld* replicates show bigger differences than most of the other presented comparisons? How is the significance of the differences between the presented distributions should be assessed?

While the large heatmaps are useful for initially appreciating the scope of the method and for showing overall patterns, it is difficult to assess the change in a pattern for a specific gene (or subset of genes). Possibly presetting the data in each mutant genotype as the change from wild-type would help in that regard?

The legend of the figures can be more elaborate and clearer. For example I am not sure I understand the analyses presented in Figure 6. Is the pattern or the change in pattern averaged and presented? The replicates in this figure seem quite different from one another, and it is hard to see how the “changes recapitulate known TF localization and function”, as stated in the title. Isn't averaging over the 100 genes with largest changes masking the patterning changes (unless they are highly similar among all of these genes)? Only some of the colors presented are mentioned in the legend. There is a reference to panels

A-D but these do not appear in the figure. Some figures include “G20”, yet only in a supplementary figure legend it is specified that it indicates the bicoid overexpression line.

Finally, as the method is still rather new, and it is very hard to assess experimental error, can the authors pick one of the more surprising observations, like a gene with increased expression in the posterior in *bcd* mutant, and provide a control experiment with an alternative method (e.g. *in situ*) demonstrating at least a qualitative agreement?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
