

A model of spatially restricted transcription in opposing gradients of activators and repressors

Michael A White¹, Davis S Parker^{2,3}, Scott Barolo² and Barak A Cohen^{1,*}

¹ Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, MO, USA and ² Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA

³ Present address: Department of Cell Biology, Duke University Medical Center, Durham, NC 27710, USA

* Corresponding author. Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, MO 63108, USA. Tel.: +1 314 362 3674; Fax: +1 314 362 2156; E-mail: cohen@genetics.wustl.edu

Received 4.5.12; accepted 20.8.12

Morphogens control patterns of transcription in development, often by establishing concentration gradients of a single transcriptional activator. However, many morphogens, including Hedgehog, create opposing activator and repressor gradients (OARGs). In contrast to single activator gradients, it is not well understood how OARGs control transcriptional patterns. We present a general thermodynamic model that explains how spatial patterns of gene expression are established within OARGs. The model predicts that differences in enhancer binding site affinities for morphogen-responsive transcription factors (TFs) produce discrete transcriptional boundaries, but only when either activators or repressors bind cooperatively. This model quantitatively predicts the boundaries of gene expression within OARGs. When trained on experimental data, our model accounts for the counterintuitive observation that increasing the affinity of binding sites in enhancers of Hedgehog target genes produces more restricted transcription within Hedgehog gradients in *Drosophila*. Because our model is general, it may explain the role of low-affinity binding sites in many contexts, including mammalian Hedgehog gradients.

Molecular Systems Biology 8: 614; published online 25 September 2012; doi:10.1038/msb.2012.48

Subject Categories: simulation and data analysis; development

Keywords: cooperativity; morphogen gradient; thermodynamic model; transcription

Introduction

Gradients of morphogens convey spatial information in developing embryos by acting through morphogen-responsive transcription factors (TFs). Some morphogens generate a single gradient of transcriptional activator (Figure 1A), and different target genes respond to different levels of this transcriptional activator. The expression pattern of target genes in these single activator gradients is strongly influenced by the affinity of activator binding sites (reviewed in Ashe and Briscoe, 2006 and Rogers and Schier, 2011). The role of binding site affinity can be explained by an activator threshold model: high-affinity binding sites are bound at low concentrations of activator, and thus enable a gene to respond broadly within the activator gradient. Conversely, the expression of genes regulated by low-affinity sites is more restricted because they respond only to higher concentrations of activator. This has been demonstrated in a variety of systems by directly modulating binding site affinity (Driever *et al.*, 1989; Jiang and Levine, 1993; Arnosti *et al.*, 1996; Wharton *et al.*, 2004; Rowan *et al.*, 2010).

An alternate way in which morphogen gradients act is to form opposing activator and repressor gradients (OARGs; Figure 1B), such as the reciprocal gradients of Gli activator and

Gli repressor produced by the Hedgehog (Hh) morphogen (Méthot and Basler, 1999). How OARGs produce spatially patterned gene expression is not well understood. Because systems of Gli OARGs include an activator gradient, the activator threshold model has, in some cases, been proposed to explain the role of DNA binding site affinity in the transcriptional response within OARGs (Wang and Holmgren, 1999; Lum and Beachy, 2004). However, in direct opposition to the predictions of the activator threshold model, we previously observed that low-affinity Gli sites drive broad expression within the Hh gradient of the *Drosophila* wing imaginal disc, while high-affinity Gli sites restrict expression to regions of highest Gli activator concentration (Parker *et al.*, 2011). This suggests that at least some OARGs are interpreted by a fundamentally different regulatory logic than that used to interpret single gradients of activators.

Proper interpretation of the OARG produced by Hh within the wing disc depends on cooperative interactions between Gli repressors (Parker *et al.*, 2011). Gli activators and repressors compete for common DNA binding sites, and cooperative interactions give Gli repressors a competitive advantage over non-cooperative Gli activators at high-affinity sites, restricting gene expression to regions of highest activator concentration.

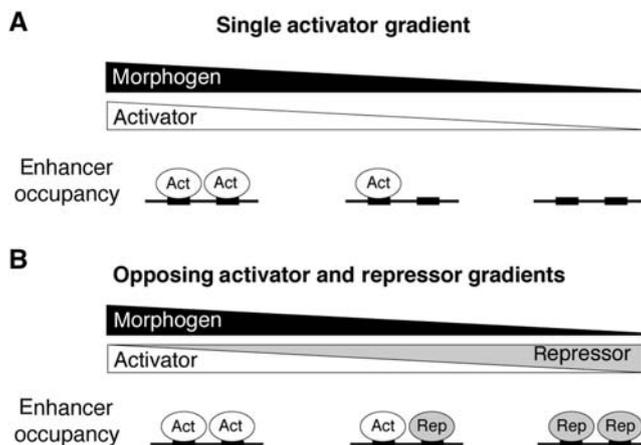


Figure 1 Morphogen gradients establish single activator gradients or opposing gradients of activators and repressors (OARGs). (A) A single transcriptional activator gradient. Activators occupy target enhancers in proportion to their concentration within the gradient. (B) OARGs. In OARGs, activators and repressors compete for enhancer binding.

This competitive advantage of Gli repressors does not extend to low-affinity sites, as low-affinity sites drive transcription broadly within the gradient. It is unclear why cooperatively interacting repressors lose their competitive advantage over activators at low-affinity sites and thus allow low-affinity sites to drive spatially broad expression. To address this question, we constructed a general quantitative framework for studying the relationships between binding site affinity, cooperativity, and gene expression boundaries in opposing gradients of activators and repressors.

Here, we derive from thermodynamic principles general features of any *cis*-regulatory system that responds to OARGs, in which activators and repressors compete for common binding sites. We show that binding site affinity determines the spatial boundaries of gene expression only in a differentially cooperative system; conversely, in a system in which both activators and repressors act non-cooperatively, discrete gene expression boundaries cannot be established by differences in enhancer binding site affinity. We show that when repressors are cooperative and activators are not, the effects of binding site number and affinity on the boundary of gene expression are the opposite of that seen in single activator gradients: multiple high-affinity sites will always produce more spatially restricted expression than low-affinity sites within OARGs. We explain why cooperative repressors out-compete activators at high-affinity sites, but not at low-affinity sites. We test our model on the Hh target gene *wingless* (*wg*), and show that the model successfully explains the expression pattern produced by low- and high-affinity versions of the *wg* enhancer in Hh OARGs of *Drosophila* embryonic parasegments. The enhancers of mammalian Hh target genes contain conserved, low-affinity Gli binding sites, which suggest that the proper response to Hh may depend on low-affinity Gli interactions in mammals as well as in *Drosophila*. Therefore, our model may explain the *cis*-regulatory logic governing the transcriptional response to Hh in a variety of species contexts, as well as the response to other morphogens that operate through OARGs.

Results

The counterintuitive role of enhancer binding site affinity in OARGs depends on cooperativity

Our goal was to understand the *cis*-regulatory logic of Gli OARGs that causes TF binding site affinity in enhancers to play a role that is the opposite of what is observed in single activator gradients. Therefore, we created a quantitative framework that describes the relationship between TF binding site affinity, cooperativity, and boundaries of gene expression in OARGs. To simplify the analysis, we began with a model of a hypothetical enhancer with two Gli binding sites, the minimum number required for cooperativity, and we also assumed that activators and repressors bind these sites with equal affinity. Because cooperative interactions between repressors play an important role in the interpretation of Gli OARGs in *Drosophila* (Parker *et al*, 2011), we assumed that repressors interact cooperatively, while activators do not. In Supplementary information and Supplementary Figure 1, we show that the results described below still apply when there are more than two binding sites, when activators and repressors do not bind with equal affinity, and when activators also exhibit some cooperativity, but less than repressors. Corresponding results hold when activators are cooperative and repressors are non-cooperative (Supplementary information).

We used a statistical thermodynamic model (Shea and Ackers, 1985; Buchler *et al*, 2003; Janssens *et al*, 2006; Gertz *et al*, 2008; Segal *et al*, 2008; Gertz and Cohen, 2009; Fakhouri *et al*, 2010; He *et al*, 2010; Parker *et al*, 2011) to compute the occupancy of TFs at a hypothetical enhancer with two Gli sites. With this model we calculate the occupancy of Gli activators and Gli repressors at the enhancer for each position in a Gli OARG. In regions of the OARG where the enhancer occupancy of activators (*occA*) is greater than the occupancy of repressors (*occR*), the target gene is activated. In regions where repressor occupancy is greater than activator occupancy, the gene is repressed. By modeling activator and repressor occupancy along the OARG, we determined the spatial boundaries of gene

expression driven by enhancers with TF binding sites of various affinities.

Using statistical thermodynamic principles (Cantor and Schimmel, 1980; Sherman and Cohen, 2012), we can write the $occA$ and $occR$ at a two-site enhancer as:

$$occA = \frac{2[A]K + 2[A][R]K^2 + 2[A]^2K^2}{Z} \quad (1)$$

$$occR = \frac{2[R]K + 2[A][R]K^2 + 2[R]^2\omega K^2}{Z} \quad (2)$$

Equations (1) and (2) give the average activator and repressor occupancy of the enhancer at a single position within the OARG (i.e., at given concentrations of activator [A] and repressor [R]), as a function of binding site affinity (K) and the strength of cooperativity between repressors (ω). The individual terms in the numerator of Equations (1) and (2) indicate the probability of specific enhancer occupancy states (e.g., one activator bound, one activator and one repressor bound, etc.), weighted by the number of activators or repressors bound in that state. These summed probabilities are normalized by the partition function Z , which is a sum over the probabilities of all possible occupied and unoccupied states of the enhancer:

$$Z = 1 + 2[R]K + 2[A]K + 2[A][R]K^2 + [R]^2\omega K^2 + [A]^2K^2 \quad (3)$$

In order to understand how enhancer binding site affinity establishes gene expression boundaries in a Gli OARG, we sought to find a relationship between K , ω and the position in the reciprocal gradients at which the switch from activation to repression occurs. The switch from activation to repression occurs at that point in the OARG when activators and repressors are exactly balanced on the enhancer, i.e., when activator occupancy equals repressor occupancy. Therefore, using the occupancy Equations (1) and (2), the switch from activation to repression occurs at a position in the OARG when [A] and [R] are such that:

$$\frac{2[A]K + 2[A][R]K^2 + 2[A]^2K^2}{Z} = \frac{2[R]K + 2[A][R]K^2 + 2[R]^2\omega K^2}{Z} \quad (4)$$

This relationship can be simplified to:

$$\frac{[A] - [R]}{[R]^2\omega - [A]^2} = K \quad (5)$$

Equation (5) defines where in the gradient (i.e., at what [A] and [R]) activator occupancy will exactly balance repressor occupancy, and therefore where a gene will switch from being activated to being repressed. Specifically, a target gene regulated by an enhancer with binding sites of affinity K will switch from activation to repression at that position in the gradient where [A] and [R] are such that the left term of Equation (5) equals K .

Since by definition K must always take a positive value, Equation (5) can only hold true when the left side of Equation (5) is positive. The left side of Equation (5) is

positive only when the following conditions hold:

$$[R] < [A] \quad (6)$$

$$[A]^2 < [R]^2\omega \quad (7)$$

Equations (6) and (7) define the boundaries of a middle zone of the OARG in which differences in enhancer TF binding site affinity will produce different boundaries of gene expression (Figure 2A). For each position in this middle zone, [A] and [R] make the left hand side of Equation (5) positive, and thus an enhancer with binding sites of affinity K which will switch from activation to repression at that position. Enhancers with TF binding sites of affinity greater than K have higher repressor occupancy and are repressed, while enhancers with binding sites of affinity lower than K are preferentially occupied by activators and drive gene expression (Figure 2A). Differences in TF binding site affinity thus produce distinct gene expression boundaries, but only within the middle zone of the gradient (Figure 2B). Outside of this middle region Equation (5) is not true, because [A] and [R] are such that the left half of Equation (5) is always negative, and all genes are either repressed (when $[R] > [A]$) or activated (when $[A]^2 > [R]^2\omega$) (Figure 2A). Therefore, outside the middle region of the OARG, differences in TF binding site affinity cannot produce different gene expression boundaries.

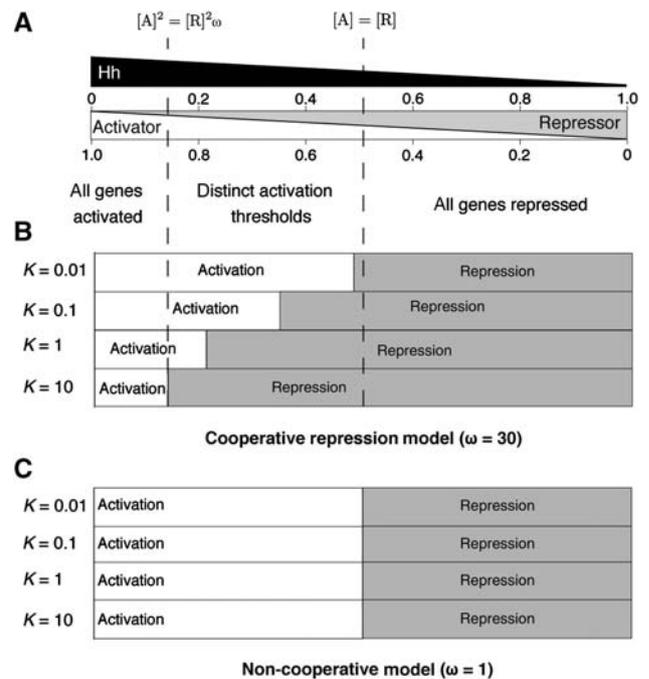


Figure 2 Repressor cooperativity establishes a middle gradient zone in which distinct gene expression boundaries are possible. (A) Hh gradient and the corresponding Gli OARG. [A] and [R] range from 0 to 100 in arbitrary units, with fraction activator and repressor shown on the scale. Dashed lines indicate the boundaries of the middle gradient zone, as defined by Equations (6) and (7). (B) Under a cooperative model ($\omega = 30$), different enhancer affinities define distinct boundaries of gene activation. As enhancer affinity increases (as K increases), the region of gene activation is more restricted. (C) Under a non-cooperative model there is no middle zone in which multiple gene expression boundaries are possible, and all genes are activated when $[A] > [R]$, regardless of affinity.

The results above show that differential cooperativity is essential for achieving distinct, affinity-based patterns of gene expression in an OARG. When both activators and repressors are non-cooperative (when $\omega = 1$), the gradient boundaries defined by Equations (6) and (7) are identical. There is then no middle zone in which the left half of Equation (5) is positive and distinct patterns of gene expression are possible. In this non-cooperative case, all genes, regardless of TF binding site affinity, are activated when $[A] > [R]$, and all genes are repressed when $[R] > [A]$ (Figure 2C). The same result holds when both activators and repressors are equally cooperative (Supplementary information).

Our results show the fundamental differences between the *cis*-regulatory logic of Gli OARGs and single activator gradients. In single activator gradients, increasing binding site affinity renders an enhancer sensitive to lower concentrations of activator, and therefore broadens gene expression within the gradient. Our model shows that the opposite response is expected in a Gli OARG, where repressors act cooperatively: higher affinity enhancers produce more restricted expression within the OARG (Figure 2B). We thus provide a potentially general explanation for the roles of cooperativity, TF binding site affinity, and activator–repressor competition in the transcriptional response to Hh, or to any morphogen that establishes OARGs.

Steepness of cooperative binding curves explains the relationship between affinity and gene expression boundaries in OARGs

The model presented above predicts that within Gli OARGs, in contrast to single activator gradients, high-affinity sites will

restrict expression within the gradient, while low-affinity sites will drive broad expression. This is the behavior observed at the Gli-regulated *dpp* enhancer in the Hh gradient of the *Drosophila* wing disc (Parker *et al*, 2011). The model indicates that this behavior occurs because, within the middle region of the OARG, cooperatively interacting repressors out-compete activators for enhancer binding at high-affinity sites, but not at low-affinity sites. However, it remains unclear from inspection of the above equations why cooperative repressors should lose their advantage over activators as the affinity of the DNA binding sites decreases. Why do low-affinity enhancers permit broad expression in the OARG, while high-affinity enhancers restrict expression?

To understand how cooperative repressors can lose their advantage over activators at low-affinity sites, we examined what occurs at a single position in an OARG as the affinity of binding sites in an enhancer is decreased. We used our model to plot activator and repressor occupancy as a function of affinity K , at specific positions within the Gli OARG (Figure 3). The result is a ligand binding curve showing the response to changes in binding site affinity at fixed ligand concentration. In the region of the gradient with high levels of activator, activator occupancy is always greater than repressor occupancy, even as total enhancer occupancy decreases with affinity (Figure 3A). In the gradient region where repressor concentration is greater than activator concentration, repressor occupancy always exceeds activator occupancy (Figure 3D). In the middle zone of the gradient, with boundaries defined by Equations (6) and (7), concentrations of activator and repressor are at intermediate levels, but activator concentration is always greater than repressor concentration (Figure 3B and C). Within this OARG region, cooperatively binding repressors out-compete activators for

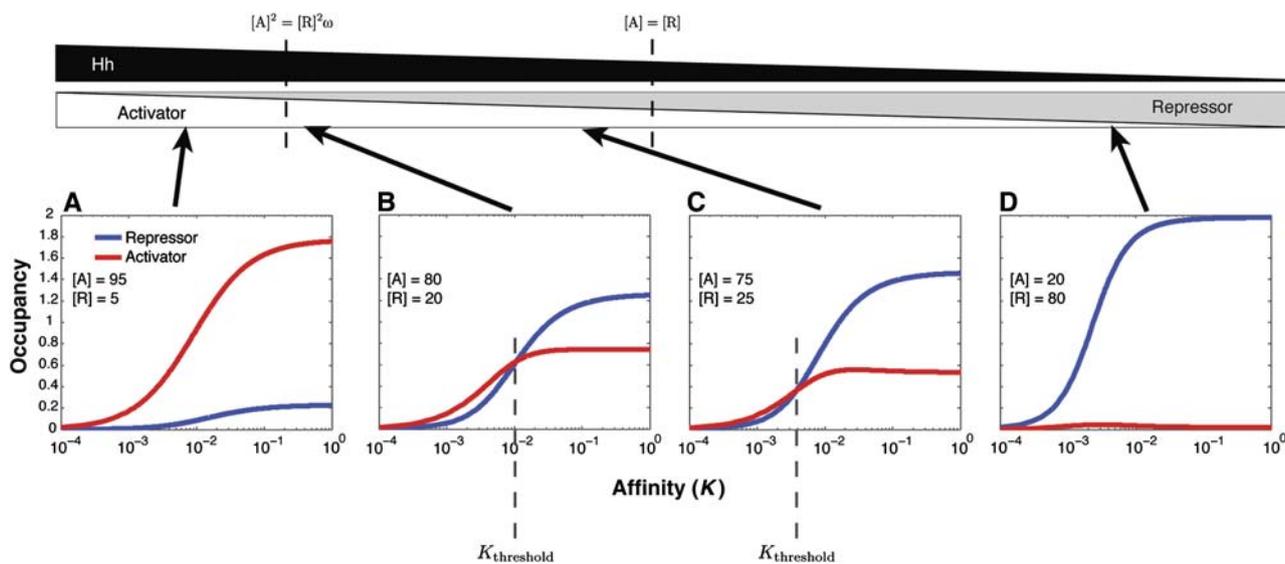


Figure 3 Repressor and activator occupancy curves demonstrate why repressors out-compete activators at high-affinity sites but not at low-affinity sites. Occupancy (y axis) of activator (red) and repressor (blue) shown as a function of the association constant, K (x axis), for a single position within the gradient (fixed $[A]$ and $[R]$). $[A]$ and $[R]$ in the gradient range between 1 and 100, in arbitrary units, and cooperativity ω is fixed at 30. (A) At high levels of activator, activator occupancy always exceeds repressor occupancy, and all genes are activated regardless of enhancer affinity. (B, C) Within the intermediate zone of the gradient, repressor and activator occupancy curves intersect, due to the inherent steepness of the cooperative repressor binding curve. The curves intersect at a threshold affinity $K_{\text{threshold}}$. Higher affinity enhancers (larger K) have higher repressor occupancy and are repressed, while lower affinity enhancers have higher activator occupancy and are activated. (D) When $[R] > [A]$, repressor occupancy always exceeds activator occupancy, and all genes are repressed regardless of enhancer affinity. These results hold for enhancers with more than two sites (Supplementary Figure 1).

occupancy at high-affinity sites, despite the greater concentration of activators. As binding site affinity decreases, overall occupancy of the enhancer decreases, and, as expected for a cooperatively binding ligand, the occupancy of cooperative repressors decreases more rapidly than the occupancy of non-cooperative activators (Figure 3B and C). Cooperative repressors lose their advantage over activators at low-affinity sites due to the increased steepness inherent in a cooperative ligand binding curve, compared with a non-cooperative binding curve.

The activator and repressor binding curves in Figure 3 explain why the role of binding site affinity in Gli OARGs is the opposite of what is observed in single activator gradients. These results also explain why distinct, affinity-based boundaries of gene expression occur in OARGs. At any given position within the middle region of the OARG, the affinity at which the activator and repressor binding curves intersect defines a threshold affinity $K_{\text{threshold}}$ (Figure 3B and C). At an individual gradient position, enhancers with binding sites of affinity greater than $K_{\text{threshold}}$ are dominated by repressors, while enhancers with binding sites of lower affinity than $K_{\text{threshold}}$ are dominated by activators. The value of $K_{\text{threshold}}$ varies with each position in the middle zone of the gradient. $K_{\text{threshold}}$ is high at higher concentrations of activator, and decreases with $[A]$ along middle region of the OARG. The result is that genes switch from activation to repression at the point in the gradient where TF binding site affinity equals $K_{\text{threshold}}$. Outside of the middle zone of the OARG, the activator and repressor occupancy curves never intersect, and there is thus no $K_{\text{threshold}}$. This is why all enhancers are either activated (Figure 3A) or repressed (Figure 3D) outside the middle of the OARG. These results demonstrate why cooperativity is necessary to produce affinity-based gene expression boundaries within OARGs. The inherent steepness of a cooperative binding curve, relative to a non-cooperative binding curve, makes it possible for both activation and repression to occur at enhancers with different binding site affinities, at a single position within the gradient.

Increasing Gli affinity in the *wg* enhancer restricts expression in *Drosophila* embryos

Our results show how the *cis*-regulatory logic of OARGs is fundamentally different from that of single activator gradients. Our Gli OARG model also successfully accounts for the unexpected role of Gli site affinity in the *dpp* wing disc enhancer. Wild-type, low-affinity Gli sites drive broad *dpp* expression within the wing disc Hh gradient. When the affinity of these sites is increased, *dpp* expression is restricted to the region of highest Gli activator concentration (Parker *et al*, 2011). Consistent with our cooperative repression OARG model, the response of *dpp* within the Gli OARG depends on the presence of multiple sites. While three high-affinity sites produce repression in the middle region of the Hh gradient, a single high-affinity site, where cooperative interactions are not possible, produces activation (Parker *et al*, 2011).

To determine whether this behavior occurs in response to other Gli OARGs, we examined the role of Gli binding site affinity in another Hh target gene, *wingless* (*wg*), which is expressed in a different tissue and developmental time point in *Drosophila*. Like *dpp* in the wing disc, the expression of *wg* in

the Hh gradients of embryonic parasegments is driven by conserved, low-affinity Gli sites (Figure 4A; Von Ohlen and Hooper, 1997). We tested the role of Gli binding site affinity by using low- and high-affinity versions of the *wg* enhancer to drive the expression of a GFP reporter gene in the parasegments of *Drosophila* embryos (Figure 4B). Low- and high-affinity versions of the *wg* enhancer were compared to a 'basal', Gli-independent *wg* enhancer in which the Gli sites were abolished. This basal enhancer revealed the effects of other factors acting on the *wg* enhancer, and activation and repression by Gli were defined relative to the basal levels of expression across each parasegment. Reporter gene expression above basal expression indicated Gli-mediated activation, while expression below basal levels indicated Gli-mediated repression.

We again observed results that contradict what would be expected in a single activator gradient, but are consistent with the predictions of our general OARG model. The low-affinity Gli sites produced activation while high-affinity sites produced repression in the posterior region of the gradient (Figure 4C). The observation that the high-affinity version of the *wg* enhancer never produces activation relative to basal suggests that much of the parasegment Hh gradient corresponds to the middle gradient zone (Figure 2A).

To make a quantitative comparison between our model and the *wg* results, we built a specific model of the *wg* enhancer and trained it on the GFP reporter data. This model was applied following the approach described in Parker *et al* (2011). In this model, non-cooperative activators and cooperative repressors compete for binding to four high- or low-affinity Gli sites in the *wg* enhancer, and bound activators recruit RNA polymerase (RNAP). Gene expression was taken to be proportional to the probability of RNAP binding to the target gene. In the model, RNAP represents a factor whose recruitment is necessary and rate limiting for target gene transcription, and which interacts with Gli proteins bound to the enhancer. The basal transcription rate was determined directly by the GFP expression from the basal enhancer. This basal expression incorporated the effects of all other regulators, and was the baseline against which Gli-mediated activation and repression were defined. The model has five free parameters that represent interactions of Gli activators and repressors with the DNA, with polymerase, and between repressors, and two free parameters that determine the steepness and height of the Hh gradient. These parameters were estimated by fitting the model to the reporter data (see Materials and methods).

We found that the model successfully explained the observed pattern of activation and repression (Figure 4C). Our model of the *wg* enhancer also largely captured the magnitude of activation and repression across the parasegments. These results, together with our earlier observations of the behavior of *dpp* (Parker *et al*, 2011), demonstrate that our model of OARG *cis*-regulatory logic successfully explains the spatial boundaries of activation and repression established by Hh in different developmental contexts.

Discussion

The behavior of *dpp* and *wg* in Hh gradients in *Drosophila* shows that OARGs work by a regulatory logic that is

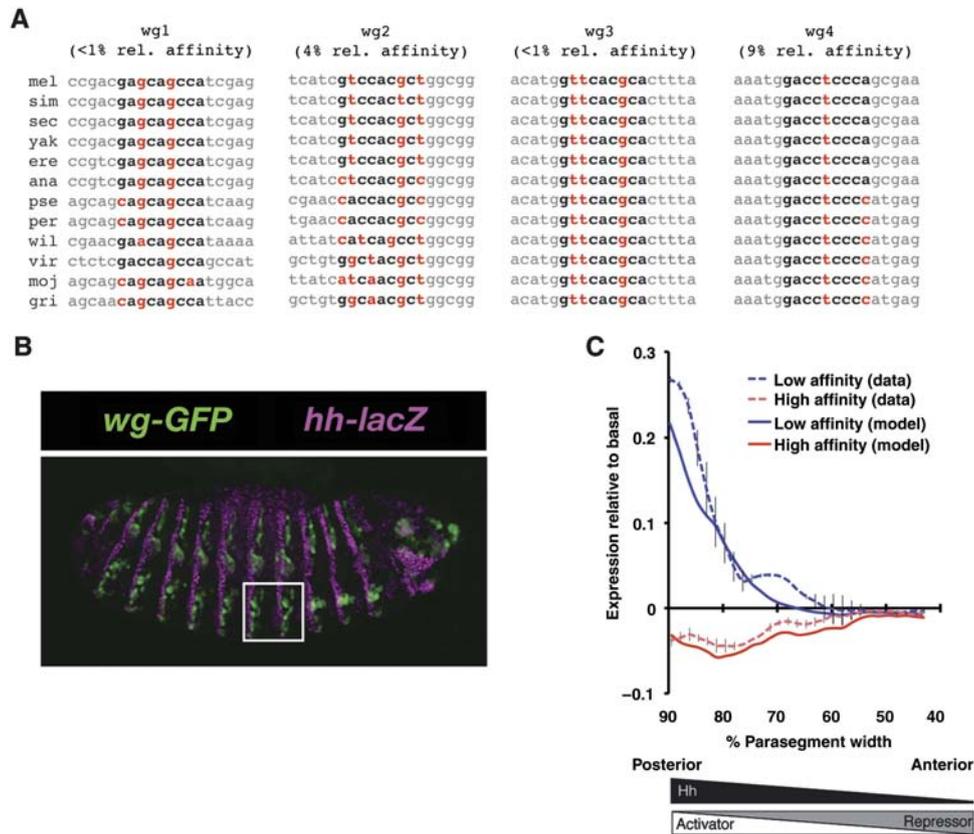


Figure 4 High-affinity Gli sites in the *wg* enhancer repress expression within Hh gradients of embryonic parasegments. (A) Four low-affinity Gli sites in the *wg* embryonic ectoderm enhancer (Von Ohlen and Hooper, 1997) are conserved in 12 *Drosophila* species. The Gli binding site is shown in bold, and mismatches to the consensus sequence (gaccacca) are shown in red. Numbers in parentheses indicate the measured affinity of the site relative to the consensus sequence (see Materials and methods). Species abbreviations: *mel*, *D. melanogaster*; *sim*, *D. simulans*; *sec*, *D. sechellia*; *yak*, *D. yakuba*; *ere*, *D. erecta*; *ana*, *D. ananasae*; *pse*, *D. pseudoobscura*; *per*, *D. persimilis*; *wil*, *D. willistoni*; *vir*, *D. virilis*; *moj*, *D. mojavensis*; *gri*, *D. grimshawi*. (B) GFP reporter gene expression in embryonic parasegments driven by an Hh-responsive *wg* enhancer. Magenta indicates Hh-*lacZ* (see Materials and methods). The white box indicates the ventral half of parasegments 6 and 7 that were quantified. (C) Increasing the affinity of Gli sites produces repression. Normalized GFP expression from low-affinity (dashed blue line) and high-affinity (dashed red line) versions of the *wg* enhancer shows mean Gli-mediated activation (positive y axis) or repression (negative y axis) along the Hh gradient (x axis) of parasegments 6 and 7. Vertical bars indicate standard error. GFP expression is defined relative to GFP expression from a basal, Gli-independent enhancer in which the Gli sites have been abolished. The cooperative repression model (solid lines) trained on these data accurately identifies regions where low-affinity sites produce activation and high-affinity sites produce repression. Wedges indicate the orientation of the Hedgehog gradient and opposing gradients of Gli activators and repressors. Reproducibility of *wg* reporter measurements in three different transgenic lines is shown in Supplementary Figure 3. Source data is available for this figure in the Supplementary Information.

fundamentally different from the logic of single activator gradients. We have presented a model of OARG *cis*-regulatory logic which can account for the surprising observation that increasing the affinity of Gli sites restricts expression in Hh gradients, and that multiple Gli sites produce repression in regions of the gradient where single sites activate (Figure 4C; Parker *et al*, 2011). Our model of OARG response explains how cooperative interactions and binding site affinity modulate the competition between activators and repressors to produce distinct gene expression boundaries within OARGs. Cooperativity is required to make discrete gene expression patterns possible, and the strength of cooperativity determines the size of the middle gradient region in which distinct expression boundaries are possible. The affinities of the enhancer TF binding sites then determine where the expression boundary of a given gene occurs. Enhancers with high-affinity sites produce spatially restricted expression, while enhancers with low-affinity sites produce broad expression. In Supplementary information (Supplementary Figure 2), we show that a mixed enhancer with one high- and one low-affinity site produces an

intermediate expression boundary. The general features of this model are not dependent on the specific composition of particular enhancers, and thus our OARG model may apply broadly to many different Hedgehog-responsive genes, as well as to targets of other morphogens that act through OARGs, such as Wnt and Dpp (Ashe and Briscoe, 2006).

Our model makes several testable predictions. In the case of an OARG with cooperative repressors, such as Gli OARGs in *Drosophila*, our model predicts that high-affinity sites should restrict expression, while low-affinity sites should broaden expression within the gradient. This is what we observed in experiments that directly test the role of Gli binding site affinity in the *dpp* and *wg* enhancers. Since low-affinity Gli binding sites play key roles in the enhancers of mammalian Hh target genes (Winklmayr *et al*, 2010), the *cis*-regulatory logic of OARG response that we have described here may apply to mammalian Hh gradients. In many cases, the sequence of functional Gli binding sites is conserved as low-affinity variants. This suggests that the role of affinity in mammalian Gli OARGs may be similar to what we observed in *Drosophila*.

Winklmayr *et al* (2010) observed that functional, non-consensus Gli sites are often required for the proper expression of mammalian Hedgehog target genes. Many of these low-affinity Gli sites are conserved between mouse and human (Eichberger *et al*, 2004, 2008; Ikram *et al*, 2004; Regl *et al*, 2004; Kasper *et al*, 2006). Such low-affinity binding sites are often missed in genome-wide scans because they fall below the selected significance threshold, yet they can affect the response to cooperatively acting TFs (Burz *et al*, 1998; Gertz *et al*, 2008). Our model suggests that these low-affinity sites are evolutionarily conserved because low affinity is essential for a broad response to Hh; mutations that increase the affinity of these sites would restrict the expression of Hh target genes, and are presumably deleterious.

Our model predicts that, because cooperativity is necessary for distinct transcriptional boundaries, the proper transcriptional response to OARGs will frequently depend on the presence of multiple binding sites. This is true of the *dpp* wing disc enhancer, where the presence of three high-affinity binding sites produces restricted expression within the Hh gradient, while a single high-affinity binding site drives broad expression (Parker *et al*, 2011). Similar results have been observed in mammals. A study of Gli binding sites in the enhancer of the mouse gene *FGF15* showed that a pair of closely spaced Gli sites produces repression (Komada *et al*, 2008). This pair of Gli sites in the *FGF15* promoter drives low expression. When one of these sites is abolished, transcription increases, suggesting that repression by Gli depends on cooperative interactions. Cooperative repression appears to control the Hh response of *Follistatin* (*FST*). The *FST* promoter also contains a pair of closely spaced Gli sites. One of these sites is necessary for activation, and cannot be abolished without impairing transcription, but deletion of the second site causes de-repression (Eichberger *et al*, 2008). These results support the hypothesis that cooperative repression is necessary for the proper response to Hh in mammals. We predict that many Hh target genes driven by multiple Gli sites will be subject to cooperative repression, and that reducing the number of Gli sites will increase the transcriptional response to Hh.

It is clear that spatial boundaries of transcription within morphogen gradients are determined by a variety of influences in addition to enhancer binding site affinity, including feedback loops, temporal adaptation of signaling pathways, and combinatorial regulation by context-specific TFs (Wharton *et al*, 2004; Ochoa-Espinosa *et al*, 2005; Vokes *et al*, 2008; Dessaud *et al*, 2010; Balaskas *et al*, 2012). Although TF binding site affinity influences gene expression within the context of these additional influences, whenever the role of affinity has been directly tested by modification of binding sites within single activator gradients, higher affinity sites have always broadened transcription within the gradient (Driever *et al*, 1989; Jiang and Levine, 1993; Arnosti *et al*, 1996; Wharton *et al*, 2004). The opposite result is observed when binding site affinity is increased within Gli OARGs in *Drosophila* (Parker *et al*, 2011). Together, these results indicate that binding site affinity is important even in the presence of feedback loops and combinatorial regulation, and that low-affinity binding sites are likely to be especially important within OARGs.

Our model suggests that a simple system is available for tuning spatial patterns of gene expression in OARGs.

Evolutionary changes that modify the physical properties of morphogen-responsive TFs are likely to have complex, global effects on downstream target genes, and therefore such changes are more likely to produce negative effects. Under our model, changes in specific gene expression boundaries can be achieved by small changes to the affinity of *cis*-regulatory sites. Such changes require few mutations and act locally on the downstream target gene. Thus, our model has potentially broad implications for our understanding of the evolutionary changes that modify developmental patterns.

Materials and methods

Wingless enhancer constructs and *Drosophila* transgenic strains

Cloning and P-element transformation were performed as previously described (Parker *et al*, 2011). An updated protocol is available at <http://sitemaker.umich.edu/barolo/protocols>. *w¹¹¹⁸* flies were used for transgenesis. To control for the effects of genomic insertion site on gene expression, three embryos from independent transgenic lines were examined for each *wg* enhancer construct, and representative results from individual embryos are shown in Figure 4C. The reproducibility of peak GFP expression in different embryos is shown in Supplementary Figure 3. Hh-lacZ (Lee *et al*, 1992) was provided by the Bloomington *Drosophila* Stock Center (Bloomington, IN). The *wg* enhancer sequences are given in Supplementary Figure 4.

Whole mount staining and microscopy

In all, 0–24 h embryos were dechorionated with bleach and fixed in 4% paraformaldehyde. Following devitellinization in methanol/heptane, embryos were blocked in 5% fetal bovine serum/PBS and immunostained. Mouse anti-LacZ (1:200) was from Promega. Rabbit anti-GFP (1:100), anti-rabbit Alexa Fluor 488 (1:1000), and anti-mouse Alexa Fluor 568 (1:1000) were obtained from Invitrogen. Embryos were mounted in ProLong Gold with DAPI (Invitrogen). Fluorescent images were captured on an Olympus FluoView 500 Laser Scanning Confocal Microscope mounted on an Olympus IX-71 inverted microscope. In order to capture the epidermis of half of each embryo (which curves in the Z dimension), we took z-stacks of 29 images spaced at 1.5 μ m. The image presented in Figure 4B is a maximum-intensity image from a stack. To minimize staining differences between samples, all embryos were treated with antibody simultaneously in a single experiment, using the same batches of reagents. All images analyzed were collected in one microscopy session, using the same confocal settings.

Quantitation of transgenic reporter expression data

Quantitation was performed with ImageJ software (National Institutes of Health, Bethesda, MD, USA) as previously described (Parker *et al*, 2011). For embryo quantitation experiments, we imaged the most vertical parasegments (6 and 7) from stage 13–14 embryos. All data were collected from images of the same size and pixel resolution, and images were quantified by summing all pixel intensities at the same A/P position. For each parasegment, the peak of the anti-Hh-lacZ signal was used as the reference point on the x axis to align data, and data were plotted as percentage distance between the peaks of anti-Hh-lacZ along the x axis. To correct for variable reagent penetration from embryo to embryo, which is the most significant source of variability in this type of experiment, anti-GFP immunofluorescence intensity was normalized by dividing it by the peak anti-Hh-lacZ immunofluorescence signal. We determined relative activation and repression from the low- and high-affinity *wg* enhancers by subtracting basal, normalized GFP expression from the low- or high-affinity normalized GFP expression. Plots of the data before subtraction of basal expression are shown in Supplementary Figure 5. To compare the *wg*

reporter data with the thermodynamic model, peak signal from the low-affinity version of the *wg* enhancer was assigned an expression value of 1, and all other reporter data were rescaled accordingly.

EMSA competition assays

wg binding site affinity was measured *in vitro* as described in Parker *et al* (2011). Affinities in Figure 4A are given relative to the consensus sequence Gli sites within the *ptc1* promoter. The sequences of the oligonucleotides are as follows: ptc1, ttggtagggGACCACCA-catcgcttg; wg1, gcgatccgacGAGCAGCCAtcgaggaaac; wg2, cattat-catcGTCCACGCTggcggctccgc; wg3, gttcgacatgTTCACGCActtaggcca; wg4, caacaaatgGACCTCCCAgcgaaagaga.

Thermodynamic model of hypothetical two-site enhancers

To obtain the results shown in Figures 2 and 3 and Supplementary Figures 1 and 2, we used Equations (1)–(3) and Supplementary Equations S5–S7, and the corresponding equations for three- and four-site enhancers to directly calculate occupancy, given hypothetical concentrations of activators and repressors, and affinities. The equations were implemented in Matlab (The Mathworks, Inc.) and the Matlab code is available as Supplementary information. For Figure 2 and Supplementary Figure 2, regions of activation were determined by identifying regions where activators exhibited higher occupancy than repressors at the relevant enhancer.

Thermodynamic model of the *wg* enhancer

The *wg* enhancer model was constructed and applied using the approach described previously (Parker *et al*, 2011). The Hh gradient is implemented in the model as a gradient of activator. Activator concentration is maximal near the Hh source, and decays exponentially in the anterior direction. Total Gli concentration is assumed to be constant over the gradient, and any Gli not in activator form is assumed to be in repressor form. As described previously (Parker *et al*, 2011), the assumption of constant total Gli levels is not essential as long as total Gli levels change slowly relative to the steepness of the Hh gradient, and total Gli occupancy of high-affinity sites remains high relative to low-affinity sites.

The model has seven free parameters: (1) the affinity of Gli activators and repressors for DNA binding sites (K), assumed to be equal in this model because allowing for differences in activator and repressor affinities did not improve the model fit; (2) the cooperative interaction energy between repressors (R-R); (3) the favorable interaction energy between activators and RNAP (A-P); (4) the unfavorable interaction energy between repressors and RNAP (R-P); (5) a parameter describing the steepness of the exponential Hh gradient (D); (6) a scaling parameter (h) that determines the total concentration of Gli in any form (held constant along the gradient); (7) the fraction of total Gli that is in activator form at maximum Hh signal within the parasegment gradient (f). We trained this model on the data shown in Supplementary Figure 5, as described in Parker *et al* (2011). The model parameters used to generate Figure 4C are R-R 11.8, A-P 9, R-P -1.3 , $h = 2.38 \times 10^{-8}$, $D = 500$, $f = 0.42$, K (low-affinity model) 4.5×10^3 , K (high-affinity model) 5×10^8 .

Gene expression across the gradient is determined by calculating the probability of RNAP binding at each gradient position. Gradient positions in the model correspond to individual data points taken from the embryo images. To calculate the probability of RNAP binding at a specific position in the gradient, we first calculate a statistical weight W_s for each possible occupancy state s of the *wg* enhancer:

$$W_s = \exp\left(-\sum_i q_i - \sum_j \omega_{ij}\right)$$

The summation is over protein–DNA interaction Gibbs-free energies q for each occupied Gli or RNAP binding site i , and each protein–protein interaction Gibbs-free energy ω between occupied Gli or RNAP sites i and j .

To obtain the absolute probability of RNAP binding, we sum the statistical weights of all RNAP bound states (W_b), and divide by the partition function, which is the sum over all enhancer states (W_s). Gene expression is proportional to the absolute probability of RNAP binding:

$$\text{Expression} = \alpha \left(\frac{\sum W_b}{\sum W_s} \right)$$

For this study, we let the proportionality constant α equals 1, because the transgenic reporter gene measurements as performed were normalized to fall between 0 and 1.

Gli activator or repressor protein–DNA interaction energies q at a specific position in the gradient are calculated from the association constant K and the activator or repressor concentrations [Gli] at that gradient position:

$$q = -\ln K - \ln[\text{Gli}]$$

Gli activator is determined by the position in the gradient:

$$\text{Gli}_{\text{ACT}} = h \times \exp\left(-\frac{x}{\sqrt{D}}\right)$$

where h is a concentration scaling parameter, D determines the steepness of the gradient, and x is the position in the gradient, with $x = 1$ at maximum activator concentration (maximum Hh signal). Total Gli concentration is determined by the maximal concentration of activator (at $x = 1$), divided by the fraction of total Gli in activator form at maximum activator concentration (f):

$$\text{Gli}_{\text{total}} = \frac{h \times \exp\left(-\frac{1}{\sqrt{D}}\right)}{f}$$

Gli repressor concentration at a specific position in the gradient is determined by subtracting $[\text{Gli}_{\text{ACT}}]$ at that position from $[\text{Gli}_{\text{total}}]$.

The intrinsic interaction energy q between RNAP and DNA is not a fit parameter, but is determined directly from the basal *wg* enhancer construct in which the Gli sites have been abolished, as described in Parker *et al* (2011). This term captures the effects of all other factors affecting the ‘basal’ expression of the *wg* enhancer. Activation and repression were defined relative to basal expression, by subtracting the expression levels of the basal *wg* construct from the predicted gene expression levels of the low- or high-affinity versions of the modeled *wg* enhancer. Positive expression values indicate activation, while negative values indicate repression, as seen in Figure 4C.

Genome sequences and alignments

Sequences and multi-species alignments were obtained from the UCSC Genome Browser (<http://genome.ucsc.edu>; Kent *et al*, 2002). *Drosophila* sequences are from the April 2006 assembly (Drosophila 12 Genomes Consortium, 2007; Clark *et al*, 2007).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank R Mitra and members of the Cohen laboratory for helpful comments on the manuscript. This work was supported by grants from the NIH to BAC and SB, including GM078222 and GM076509, American Recovery and Reinvestment Act supplements GM07822203S1 and GM07650903S1, and grant MCB-1157800 from the NSF to SB and BAC.

Author contributions: MAW and BAC developed and analyzed the model, and wrote the paper. DSP and SB conducted the *wg* experiments.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- Arnosti DN, Barolo S, Levine M, Small S (1996) The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**: 205–214
- Ashe HL, Briscoe J (2006) The interpretation of morphogen gradients. *Development* **133**: 385–394
- Balaskas N, Ribeiro A, Panovska J, Dessaud E, Sasai N, Page KM, Briscoe J, Ribes V (2012) Gene regulatory logic for reading the sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* **148**: 273–284
- Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. *Proc Natl Acad Sci USA* **100**: 5136–5141
- Burz DS, Rivera-Pomar R, Jäckle H, Hanes SD (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo. *EMBO J* **17**: 5998–6009
- Cantor C, Schimmel P (1980) *Biophysical Chemistry: The Behavior of Biological Macromolecules (Part III): New York: W.H. Freeman and Company*
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW et al (2007) Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**: 203–218
- Dessaud E, Ribes V, Balaskas N, Yang LL, Pierani A, Kicheva A, Novitsch BG, Briscoe J, Sasai N (2010) Dynamic assignment and maintenance of positional identity in the ventral neural tube by the morphogen sonic Hedgehog. *PLoS Biol* **8**: e1000382
- Driever W, Thoma G, Nüsslein-Volhard C (1989) Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340**: 363–367
- Eichberger T, Kaser A, Pixner C, Schmid C, Klingler S, Winklmayr M, Hauser-Kronberger C, Aberger F, Frischauf AM (2008) GLI2-specific transcriptional activation of the bone morphogenetic protein/activin antagonist follistatin in human epidermal cells. *J Biol Chem* **283**: 12426–12437
- Eichberger T, Regl G, Ikram MS, Neill GW, Philpott MP, Aberger F, Frischauf AM (2004) FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. *J Invest Dermatol* **122**: 1180–1187
- Fakhouri WD, Ay A, Sayal R, Dresch J, Dayringer E, Arnosti DN (2010) Deciphering a transcriptional regulatory code: modeling short-range repression in the Drosophila embryo. *Mol Syst Biol* **6**: 341
- Gertz J, Cohen BA (2009) Environment-specific combinatorial cis-regulation in synthetic promoters. *Mol Syst Biol* **5**: 244
- Gertz J, Siggia ED, Cohen BA (2008) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**: 215
- He X, Samee MAH, Blatti C, Sinha S (2010) Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comp Biol* **6**: pii: e1000935
- Ikram MS, Neill GW, Regl G, Eichberger T, Frischauf AM, Aberger F, Quinn A, Philpott M (2004) GLI2 is expressed in normal human epidermis and BCC and induces GLI1 expression by binding to its promoter. *J Invest Dermatol* **122**: 1503–1509
- Janssens H, Hou S, Jaeger J, Kim A-R, Myasnikova E, Sharp D, Reinitz J (2006) Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even-skipped gene. *Nat Genet* **38**: 1159
- Jiang J, Levine M (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**: 741–752
- Kasper M, Schnidar H, Neill GW, Hanneder M, Klingler S, Blaas L, Schmid C, Hauser-Kronberger C, Regl G, Philpott MP, Aberger F (2006) Selective modulation of Hedgehog/GLI target gene expression by epidermal growth factor signaling in human keratinocytes. *Mol Cell Biol* **26**: 6283–6298
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler AD (2002) The Human Genome Browser at UCSC. *Genome Res* **12**: 996–1006
- Komada M, Saitsu H, Shiota K, Ishibashi M (2008) Expression of Fgf15 is regulated by both activator and repressor forms of Gli2 *in vitro*. *Biochem Biophys Res Commun* **369**: 350–356
- Lee JJ, Kessler von DP, Parks S, Beachy PA (1992) Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog. *Cell* **71**: 33–50
- Lum L, Beachy P (2004) The Hedgehog response network: sensors, switches, and routers. *Science* **304**: 1755
- Méthot N, Basler K (1999) Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**: 819–831
- Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, Oberstein A, Papatsenko D, Small S (2005) The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. *Proc Natl Acad Sci USA* **102**: 4960
- Parker DS, White MA, Ramos AI, Cohen BA, Barolo S (2011) The cis-regulatory logic of hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. *Sci Signal* **4**: ra38
- Regl G, Kasper M, Schnidar H, Eichberger T, Neill GW, Philpott MP, Esterbauer H, Hauser-Kronberger C, Frischauf AM, Aberger F (2004) Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. *Cancer Res* **64**: 7724–7731
- Rogers KW, Schier AF (2011) Morphogen gradients: from generation to interpretation. *Annu Rev Cell Dev Biol* **27**: 377–407
- Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, Maas RL (2010) Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev* **24**: 980–985
- Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U (2008) Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451**: 535
- Shea MA, Ackers GK (1985) The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J Mol Biol* **181**: 211–230
- Sherman MS, Cohen BA (2012) Thermodynamic state ensemble models of cis-regulation. *PLoS Comp Biol* **8**: e1002407
- Vokes SA, Ji H, Wong WH, McMahon AP (2008) A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* **22**: 2651–2663
- Von Ohlen T, Hooper JE (1997) Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer. *Mech Dev* **68**: 149–156
- Wang QT, Holmgren RA (1999) The subcellular localization and activity of Drosophila cubitus interruptus are regulated at multiple levels. *Development* **126**: 5097–5106
- Wharton SJ, Basu SP, Ashe HL (2004) Smad affinity can direct distinct readouts of the embryonic extracellular Dpp gradient in Drosophila. *Curr Biol* **14**: 1550–1558
- Winklmayr M, Schmid C, Laner-Plamberger S, Kaser A, Aberger F, Eichberger T, Frischauf A-M (2010) Non-consensus GLI binding sites in Hedgehog target gene regulation. *BMC Mol Biol* **11**: 2



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.