Supplementary Information for "A model of spatially restricted transcription in opposing gradients of activators and repressors"

Michael A. White, Davis S. Parker, Scott Barolo, Barak A. Cohen

Table of Contents

Model results with relaxed assumptions

To derive equation 5 in the main text, we made several simplifying assumptions: 1) the enhancer contains two binding sites, 2) activators and repressors bind sites with equal affinity, and 3) repressors are cooperative but activators are not. Here we show that the results described in the text still hold when these assumptions are relaxed.

Additional binding sites:

We have been unable to discover simple and exact analytical expressions for $K_{threshold}$, and for the boundaries of the middle gradient zone for enhancers containing three or more binding sites. This is because the occupancy equations for enhancers with more than two sites are higher order polynomial functions of *K*. However, the occupancy equations for enhancers with three or more sites can be used to numerically determine $K_{threshold}$ and to identify the middle zone of an OARG, in which gene expression boundaries depend on enhancer binding site affinity. These numerical simulations show that the results described for a two-site enhancer in the main text also apply to enhancers with more than two sites.

We used the occupancy equations for enhancers with three and four sites (Sherman and Cohen, 2012; Cantor and Schimmel, 1980) to examine activator and repressor occupancy as a function of *K* at different positions in the gradient (Supplementary Figure 1; compare with Figure 3 in the main text). These results show that enhancers with more than two sites exhibit the same behavior within an OARG as a two-site enhancer. All enhancers, regardless of TF binding site affinity, are repressed when [R]>[A] (Supplementary Figure 1, C and F). There is a middle gradient zone in which the switch from activation to repression depends on binding site affinity (Supplementary Figure 1, B and E), and in regions of highest activator concentration all enhancers are activated (Supplementary Figure 1, A and D). The numerical simulations show that the boundary beyond which all enhancers are activated is approximately at $[A]^n = [R]^n\omega^{n-1}$. As with two-site enhancers, high-affinity sites produce more restricted gene expression relative to low-affinity sites. Regardless of the number of TF binding sites in the enhancer, the qualitative behavior of that enhancer within an OARG will conform with the main results presented in the text, due to the inherently steep binding curve of cooperative repressors relative to non-cooperative activators.

Unequal affinities for activators and repressors

In the case where the association constant for repressors is a multiple of the association constant *K* for activators, then the association constant for repressors can be written as *K* times a constant γ, and then the occupancy equation for repressor is written:

$$\mathrm{occR} = \frac{2[\mathrm{R}]\gamma K + 2[\mathrm{A}][\mathrm{R}]\gamma K^2 + 2[\mathrm{R}]^2\omega\gamma^2 K^2}{Z} \qquad \text{(eq. S1)}$$

If we replace equation 2 in the main text with equation S1, and proceed with the derivation, we obtain, instead of equation 5, the following result which defines the boundaries of gene expression when activators and repressors exhibit different DNA binding affinities:

$$\frac{[\mathrm{A}]-[\mathrm{R}]}{[\mathrm{R}]^2\omega\gamma^2-[\mathrm{A}]^2}=K \qquad \text{(eq. S2)}$$

Cooperative activators

If activators and repressors are equally cooperative, then the resulting model is identical to the non-cooperative model, and binding affinity does not determine the boundaries of gene expression. However, in cases of asymmetric cooperativity, when activators and repressors both act cooperatively, but repressors exhibit stronger cooperativity, then the middle gradient zone exists, and affinity-based gene expression boundaries can occur.

If repressors exhibit cooperative interaction energy ω, and activators exhibit a cooperative interaction energy ε, then the activator occupancy is written:

$$\mathrm{occA}=\frac{2[\mathrm{A}]K+2[\mathrm{A}][\mathrm{R}]K^2+2[\mathrm{A}]^2\epsilon K^2}{Z} \qquad \text{(eq. S3)}$$

Replacing equation 1 in the main text with equation S3, and proceeding with the derivation, we obtain the following:

$$\frac{[\mathrm{A}]-[\mathrm{R}]}{[\mathrm{R}]^2\omega-[\mathrm{A}]^2\epsilon}=K \qquad \text{(eq. S4)}$$

When $\omega > \varepsilon$, *i.e.*, when repressors are more strongly cooperative than activators, the left side of equation S4 will be positive, and thus will hold true only when $[A] > [R]$ and when $[A]^2\varepsilon < [R]^2\omega$. In this case, there is again a middle gradient zone analogous to that produced by the model discussed in the main text (equation 5). When repressors and activators are both equally cooperative (when $\omega = \varepsilon$), equation S4 becomes equivalent to the non-cooperative model ($\omega = 1$) described in the main text, and no affinity-based expression boundaries are possible.

When activators are more cooperative than repressors ($\varepsilon > \omega$), the effect of binding site affinity will be the opposite of that described for the cooperative repressor model in the main text. The left side of equation S4 is positive only when both $[R] > [A]$ and $[A]^2\varepsilon > [R]^2\omega$, which is the opposite of cooperative repressor model (see equations 6 and 7 in the main text). This creates a middle gradient zone that extends in the direction opposite that of the middle zone of the cooperative repressor model, and the effect of binding site affinity on the spatial pattern of expression is also the opposite. When activators are more cooperative than repressors, low-affinity sites will produce more restricted expression within the gradient, while high-affinity sites will produce broader activation.

Enhancers with both low- and high-affinity sites

Many enhancers in the genome contain mixtures of high- and low- affinity sites. We used the thermodynamic model to determine the gene expression boundary of an enhancer with one low- and one high-affinity site. Because the sites in the enhancer have varying affinities, there is no single value for the association constant *K*, and therefore it is not possible to derive a simple analytical expression for the mixed site enhancer analogous to equation 5 in the main text. We there for used thermodynamic occupancy equations (similar to equations 1 and 2 in the main

text) to model gene expression boundaries for an enhancer with one low- and one high-affinity site.

The activator (occA) and repressor (occR) occupancy equations for the mixed enhancer are:

$$\mathrm{occA} = \frac{[\mathrm{A}]K_{\mathrm{low}} + [\mathrm{A}]K_{\mathrm{high}} + 2[\mathrm{A}][\mathrm{R}]K_{\mathrm{low}}K_{\mathrm{high}} + 2[\mathrm{A}]^2K_{\mathrm{low}}K_{\mathrm{high}}}{Z} \qquad \text{(eq. S5)}$$

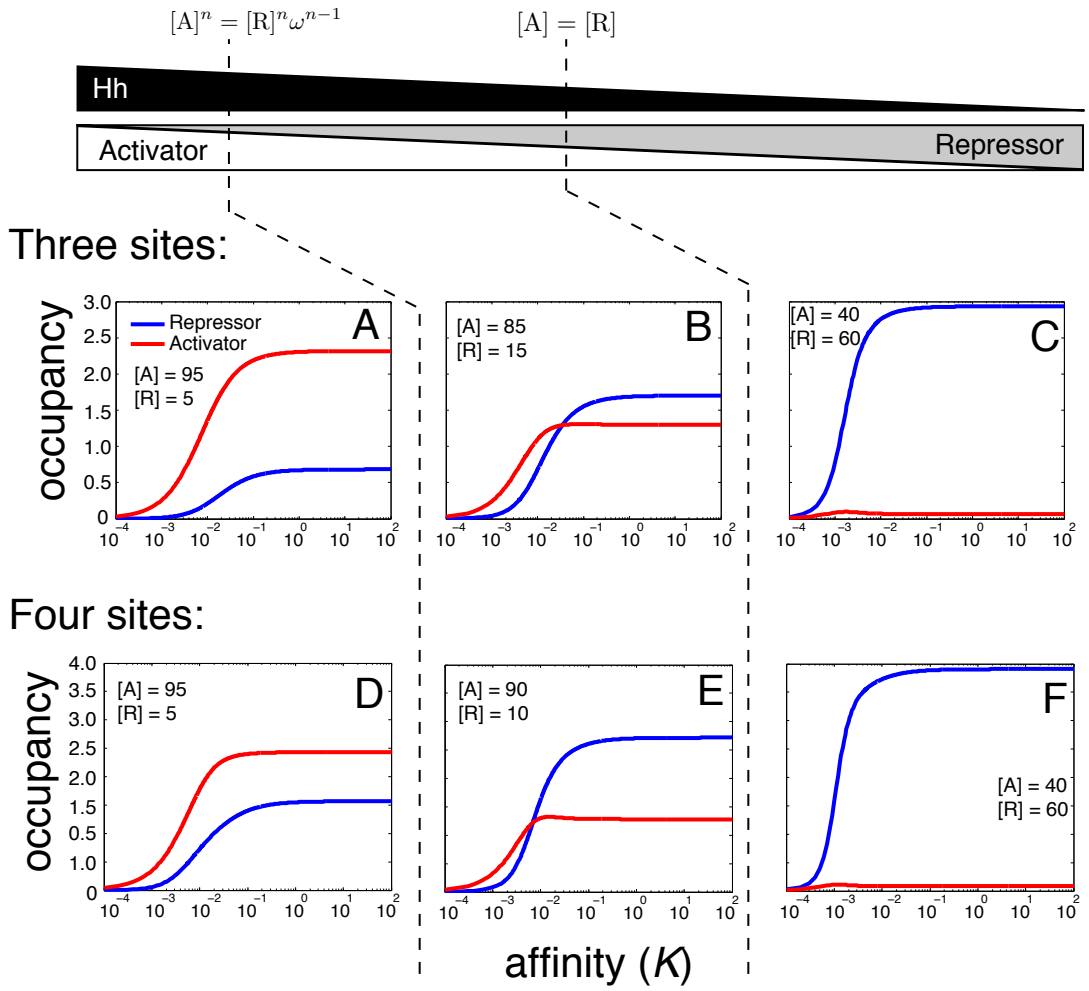
$$\mathrm{occR} = \frac{[\mathrm{R}]K_{\mathrm{low}} + [\mathrm{R}]K_{\mathrm{high}} + 2[\mathrm{A}][\mathrm{R}]K_{\mathrm{low}}K_{\mathrm{high}} + 2[\mathrm{R}]^2\omega K_{\mathrm{low}}K_{\mathrm{high}}}{Z} \qquad \text{(eq. S6)}$$

The partition function *Z* is:

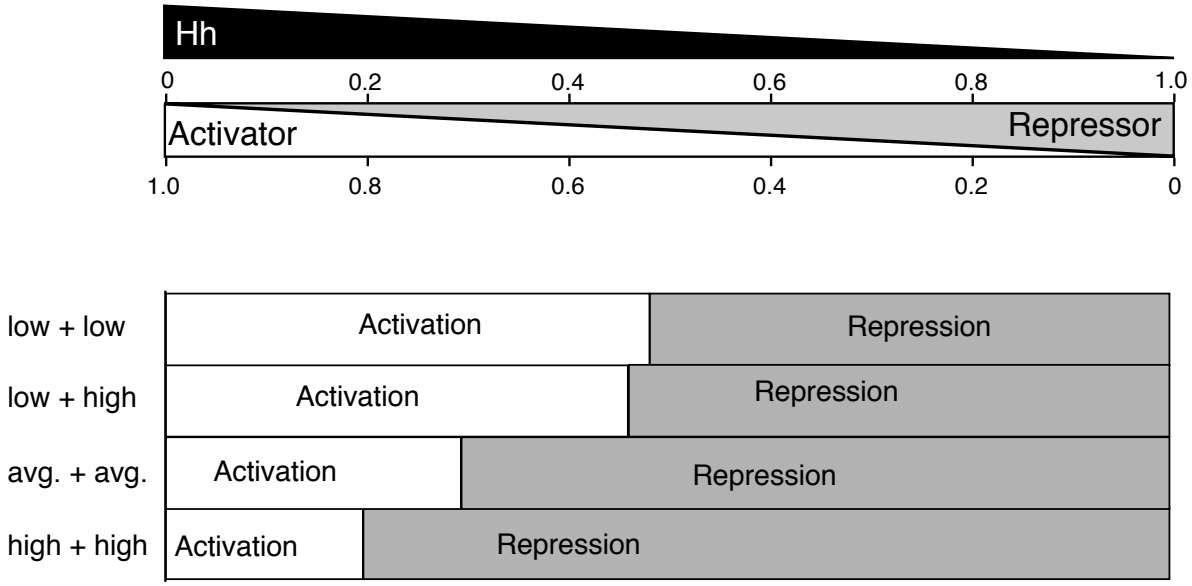
$$Z = [\mathrm{A}]K_{\mathrm{low}} + [\mathrm{A}]K_{\mathrm{high}} + [\mathrm{R}]K_{\mathrm{low}} + [\mathrm{R}]K_{\mathrm{high}} + 2[\mathrm{A}][\mathrm{R}]K_{\mathrm{low}}K_{\mathrm{high}} + [\mathrm{A}]^2K_{\mathrm{low}}K_{\mathrm{high}} + [\mathrm{R}]^2\omega K_{\mathrm{low}}K_{\mathrm{high}} \qquad \text{(eq. S7)}$$

We found that an enhancer with one low- and one high-affinity site exhibited a gene expression boundary lying between the boundaries of the enhancer with high-affinity sites and the enhancer with low-affinity sites, but closer to the enhancer with low-affinity sites (Supplementary Figure 2). The gene expression boundary of the mixed enhancer was clearly not a simple function of the average affinity of the two binding sites, because an enhancer with two identical sites of average affinity exhibited more restricted activation than the mixed enhancer. These results suggest that mixtures of sites of different affinities provide an additional means of tuning gene expression boundaries in an OARG. Mixtures of high- and low-affinity sites may also provide a means of semi-independently tuning gene expression boundary and level of expression. The mixed enhancer shown in Supplementary Figure 2 exhibits a gene expression boundary similar to that of the enhancer with two low-affinity sites. But because the mixed site enhancer has one high-affinity site, it exhibits a substantially higher TF occupancy than the enhancer with two low-affinity sites (not shown). The result is that, when activated, the mixed enhancer is predicted to drive higher levels of gene expression compared with the enhancer containing only low-affinity sites.

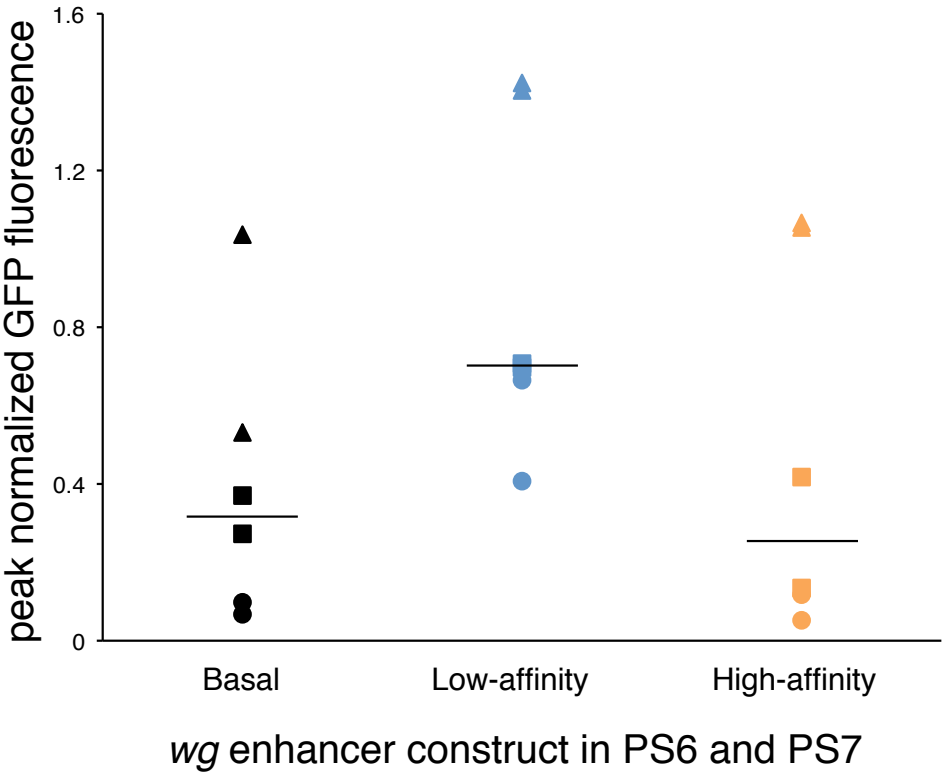
Supplementary Figures



Supplementary Figure 1: High-affinity sites restrict expression from enhancers with more than two 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. (Compare with Figure 3 in the text.) **A and D)** At high levels of activator, activator occupancy always exceeds repressor occupancy, and all gene are activated regardless of enhancer binding site affinity. **B and E)** Within the intermediate zone of the gradient, repressor and activator occupancy curves intersect, due to the inherent steepness of the cooperative repressor binding curve. Enhancers with higher affinity binding sites (larger *K*) have higher repressor occupancy and are repressed, while enhancers with lower affinity sites have higher activator occupancy and are activated. **C and F** When [R] > [A], repressor occupancy always exceeds activator occupancy, and all genes are repressed regardless of TF binding site affinity.



Supplementary Figure 2: Enhancers with mixed high- and low-affinity sites produce intermediate expression boundaries in a cooperative repression model. The Hh gradient and the corresponding Gli OARG are shown above, and the modeled boundaries of gene expression for different two-site enhancers are shown below. Boundaries were modeled for enhancers with two low-affinity sites (low + low), one low- and one high-affinity site (low + high), two sites of average affinity (avg. + avg.), and two high affinity sites (high + high). Low-affinity sites were assigned K = 0.01, high-affinity sites were assigned K = 1, and average sites were assigned K = 0.505. Average affinity was determined by taking the average of one low- and one high-affinity site, *i.e.*, (1 + 0.01)/2. Activators were non-cooperative, and repressors were cooperative ($\omega = 30$).

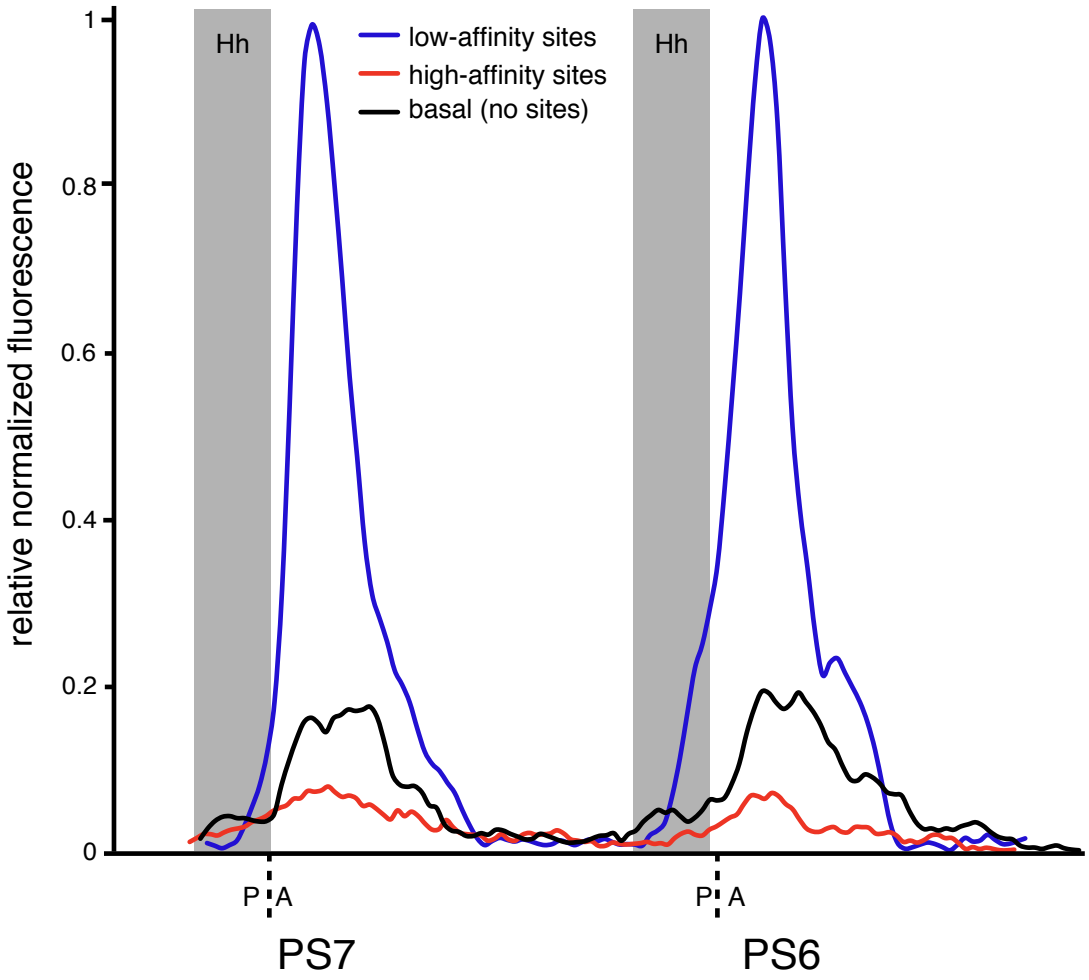


Supplementary Figure 3: Reproducibility of *wg* reporter expression in different transgenic lines. Data points indicate peak normalized GFP reporter expression (see Materials and Methods) driven by a Hh-responsive enhancer of *wg* in parasegments (PS) 6 and 7. Each *wg* enhancer construct was tested in three independent embryos, each from an independent transgenic *Drosophila* line. Fluorescence intensity was measured for two parasegments (PS6 and PS7) within each embryo and different data point shapes (circles, triangles, squares) within each category indicate measurements from different embryos. Horizontal bars indicate median peak GFP fluorescence.

***wg* embryonic ectoderm enhancer (1026 bp)**
-4830 to -3805 upstream of *wg* (2L:7,302,331-7,303,356)

```
GATATCTGATTAATATCATTTTTTCATTGAAACACATTACCAAGTAATATCGACCCCTTTCCTTCCATTTGCCT
GATATTTACCATTTCGGTGCCATTTCCCAGTGAACCCATTATGGATCTTGTTACCATAATGTCAGCACTAAAAT
GGCTTCCTCCGCTGCCGAGAAGAGATCGCCACCACCACTCTACTCTTTTGCACATGCCACCACCACCACCACCA
                                                Ci1
CACGCTCCGGTGGTTCCCCTTTCGCCCCGCCATGTTTCCTCGATGGCTGCTCGTCGGATCGCACTTCCGGAG
                                      Ciptc: ---G--G--
                                       CiKO: -----T---
            Ci2
CGCGGACCGCCAGCGTGGACGATGATAATGCATTATGCATTCCGCCGCCTGGGTTTTAATTGTTATTAACTCGC
    Ciptc: T-T----T-
     CiKO: -----T---
GATAAGGAGAACTTTAACACCTATAGTTAGGCCGGGCCAGAAAATGTAGCCGAGCATGCTAAGATATAATTTCA
ACATTTTATAAGCACTTAACAACCCAAAAACCGCTGGTAATGCACCGAAAAAAAGTAGTTAGAAGTAATGTGAG
TAATAATAGAATATTAGTTTTTAAATAGTATAGAAAAAAAACATTAAATGACAAGAAAGTTCAAAGAAAACGTT
TACAAATGCATTTTTATCCAATTACAAAATATCTTCCTGATATTGAGCGATTTAAATGTAGTTGCATTCTTTTC
CCTCTGTGCATAAATTTTTCGGTAACAAAAACTGGCTGCCTTTTCCTTTCGCTGGCGGGGGAACGGGAAATATA
AATTGTCGGACCCCGAAATTCTACACTGTCATCAAAGCGTAACTAAACGCCCATTTTTCTCAACTTTTTGGAGT
                                                        Ci3
CTCTTTCGGACCGGACAACCATTTTCGGTAGTTATTAGTGGCATATTTTGGCCTAAAGTGCGTGAACCATGTCG
                                                  Ciptc: --G---GT-
                                                  CiKO:  -----T---
                                                           Ci4
AACAGCCAGCGCAGGCAACAAAATTTATTGTTTAGTAGTGGAGTGAAGTTAACAACAAAATGGACCTCCCAGCG
                                                       Ciptc: --G---GT-
                                                        CiKO: ---T-----
AAAGAGAAAGACACGCAACGATCCCAACGCGGACCTGGCCAGAAAAAAATATTAACGCCTCGAG
```

Supplementary Figure 4: Sequences of variants of the *wg* embryonic ectoderm enhancer used in reporter constructs. The four Ci (*Drosophila* Gli) binding sites are indicated (Von Ohlen and Hooper, 1997). Wild-type, low-affinity sequences are shown in green, high-affinity, consensus sequences from the *ptc* promoter (*Ciptc*) are shown in red, and the mutant, non-functional Ci sites (*CiKO*) in the basal enhancer are shown in blue.



Supplementary Figure 5: Normalized anti-GFP fluorescence from different versions of the *wg* enhancer. Anti-GFP fluorescence was measured in embryonic parasegments 6 (PS6) and 7 (PS7) in individual embryos harboring different versions of the *wg* enhancer carrying low-affinity, high-affinity, or abolished (basal) Gli sites. Hedgehog (Hh) signal and the anterior (A)/posterior (P) boundary positions were determined from the peak anti-Hh-lacZ signal. Anti-GFP signal was normalized against anti-Hh-lacZ signal and rescaled relative to peak expression from the low-affinity version of the *wg* enhancer, as described in Materials and Methods.