Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch

Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch

Dimitrios K. Papadopoulos[1,2,*], Kassiani Skouloudaki[1], Ylva Engström[3], Lars Terenius[4], Rudolf Rigler[5,6], Christoph Zechner[1,7], Vladana Vukojević[4] and Pavel Tomancak[1,*]

[1]Max-Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[2]Current address: MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
[3]Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, 10691 Stockholm, Sweden
[4]Center for Molecular Medicine (CMM), Department of Clinical Neuroscience, Karolinska Institutet, 17176 Stockholm, Sweden
[5]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden
[6]Laboratory of Biomedical Optics, Swiss Federal Institute of Technology, 1015 Lausanne, Switzerland
[7]Center for Systems Biology Dresden, 01307, Dresden, Germany

[*]Correspondence: dpapado2@ed.ac.uk and tomancak@mpi-cbg.de

Key words

Auto-regulation, Fluorescence Correlation Spectroscopy, Hox genes, Protein noise, Transcription factors, Variability

Summary statement

Preferentially repressing and activating isoforms of the Hox transcription factor Antennapedia elicit a developmental regulatory switch from auto-activation to auto-repression which increases concentration and suppresses cell-to-cell variability over time.

Abstract

The variability in transcription factor concentration among cells is an important developmental determinant, yet how variability is controlled remains poorly understood. Studies of variability have focused predominantly on monitoring mRNA production noise. Little information exists about transcription factor protein variability, since this requires the use of quantitative methods with single-molecule sensitivity. Using Fluorescence Correlation Spectroscopy (FCS), we characterized the concentration and variability of 14 endogenously tagged TFs in live Drosophila imaginal discs. For the Hox TF Antennapedia we investigated whether protein variability results from random stochastic events or is developmentally regulated. We

found that Antennapedia transitioned from low concentration/high variability early, to high concentration/low variability later, in development. FCS and temporally resolved genetic studies uncovered that Antennapedia itself is necessary and sufficient to drive a developmental regulatory switch from auto-activation to auto-repression, thereby reducing variability. This switch is controlled by progressive changes in relative concentrations of preferentially activating and repressing Antennapedia isoforms, which bind chromatin with different affinities. Mathematical modelling demonstrated that the experimentally supported auto-regulatory circuit can explain the increase of Antennapedia concentration and suppression of variability over time.

Introduction

In order to understand the mechanisms that control pattern formation and cell fate specification in developing organisms, the intranuclear concentration, DNA-binding kinetics and cell-to-cell variability of relevant transcription factors (TFs) need to be quantified. TF concentration variability at the tissue level is thought to arise from diverse processes, including mRNA transcription, translation and protein degradation. Intrinsic noise is due to stochastic binding and interactions of proteins involved in transcriptional activation of a specific gene (Blake et al., 2003; Elowitz et al., 2002). Extrinsic noise arises from inter-cellular differences in abundance of the transcriptional and post-transcriptional machinery (Swain et al., 2002).

In undifferentiated tissue or cells, TF cell-to-cell variability can be the driving force for differentiation. For example, progressive establishment of a Nanog salt-and-pepper expression pattern leads to the formation of primitive endoderm in the mouse preimplantation embryo, whereas loss of the variability results in embryos lacking primitive endoderm entirely (Kang et al., 2013).

Conversely, in already differentiated tissue or cells, TF expression variability among cells may need to be counteracted to ensure homogeneity of gene expression patterns and robustness of commitment to a certain transcriptional regime. Examples are the Snail (Sna) TF, which is required for the invagination of the mesoderm during Drosophila gastrulation (Boettiger and Levine, 2013), or the Bicoid (Bcd) and Hunchback (Hb) TFs during early embryogenesis (Gregor et al., 2007a; Gregor et al., 2007b; Little et al., 2013).

In addition, differential cell fates within the same developmental territory may be specified by TFs deploying different DNA-binding dynamics despite the existence of very similar concentrations (i.e. low variability). For example, studies on the Oct4 TF in early mouse embryos have shown that differential kinetic behavior of DNA binding, despite equal Oct4 concentration among blastomeres, ultimately dictates an early developmental bias towards lineage segregation (Kaur et al., 2013; Plachta et al., 2011).

So far, studies of gene expression variability have focused predominantly on monitoring the noise of mRNA production (Holloway et al., 2011; Holloway and Spirov, 2015; Little et al., 2013; Lucas et al., 2013; Pare et al., 2009). Little information exists about TF variability at the protein level within a tissue. Such studies require the use of quantitative methods with single-molecule sensitivity.

We have previously used Fluorescence Correlation Spectroscopy (FCS), to quantitatively characterize Hox TF interactions with nuclear DNA in living salivary gland cells (Papadopoulos et al., 2015; Vukojevic et al., 2010). FCS is instrumental for quantifying TF dynamics in living cells or tissue (Clark et al., 2016; Kaur et al., 2013; Lam et al., 2012; Mistri et al., 2015; Papadopoulos et al., 2015; Perez-Camps et al., 2016; Szaloki et al., 2015; Tiwari et al., 2013; Tsutsumi et al., 2016). However, in these

studies, mobility has only been measured for overexpressed proteins. However, to understand TF behavior *in vivo*, proteins need to be quantified at endogenous levels (Lo et al., 2015).

In this study, we take advantage of the availability of fly toolkits, in which TFs have been endogenously tagged by different methodologies: fosmid (Baumgartner et al., 1996), BAC (deposition of lines of Rebecca Spokony and Kevin White to Flybase and the Bloomington Stock Center), FlyTrap (Buszczak et al., 2007; Kelso et al., 2004; Morin et al., 2001; Quinones-Coello et al., 2007) and MiMIC lines (Nagarkar-Jaiswal et al., 2015; Venken et al., 2011), to measure the intranuclear concentration of various TFs *in vivo* by FCS, and their cell-to-cell variability in fly imaginal discs. Imaginal discs are flat, single-layered epithelia comprised of small diploid cells and many TFs are expressed in defined regions within these tissues during development.

Results

Characterization of average protein concentrations and cell-to-cell variability of *Drosophila* TFs

Average concentrations of TFs in neighboring nuclei of third instar imaginal discs were measured by FCS (Fig. 1A-J and Supplemental Fig. S1A-P). FCS is a non-invasive method with single molecule sensitivity, in which a confocal arrangement of optical elements is used to generate a small (sub-femtoliter) detection volume inside living cells, from which fluorescence is being detected (Fig. 1C,D; green ellipsoid). Fluorescent molecules diffuse through this observation volume, yielding fluorescence intensity fluctuations that are recorded over time by detectors with single-photon sensitivity (Fig. 1E). These fluctuations are subsequently subjected to temporal autocorrelation analysis, yielding temporal autocorrelation curves (henceforth referred to as FCS curves, Fig. 1F), which are then fitted with selected models to extract quantitative information about the dynamic processes underlying the generation of the recorded fluctuations. In the case of molecular movement of TFs (Supplement 1), information can be obtained regarding: a) the absolute TF concentrations (Fig. 1F), (b) TF dynamic properties, such as: diffusion times, differences in their interactions with chromatin and fractions of free-diffusing *versus* chromatin-bound TFs (Fig. 1G); and c) cell-to-cell TF concentration variability (Fig. 1H).

We selected 14 TF based on the availability of homozygous, endogenously tagged transgenes and generation of robust fluorescence in distinct patterns in various imaginal discs. For the 14 TFs, we measured average concentrations ranging about two orders of magnitude among different TFs, from $\sim 30\ nM$ to $\sim 1.1\ \mu M$ (~400 to 15500 molecules per nucleus, respectively) (Fig. 1I, Supplemental Fig. S1A-Q and Supplement 2). Various diffusion times and fractions of slow and fast diffusing TF molecules (Fig. 1J) indicated differential mobility and degree of DNA-binding among different TFs (Vukojevic et al., 2010). Comparison of the y-axis amplitudes at the zero lag time of the FCS curves, which are inversely proportional to the concentration of fluorescent molecules (Fig. 1F), gives information about concentration variability (heterogeneity) among different cell nuclei, i.e. reflects heterogeneity of protein concentration at the tissue level (Fig. 1H). For all 14 TFs studied, the variability, expressed as the variance over the mean squared, $CV^2 = \frac{s^2}{m^2}$, was determined to be in the range $7 - 37\%$ (Fig. 1K and Supplemental Fig. S1Q).

In biological systems, the Fano factor, expressed as the variance over the mean ($F_f = \frac{s^2}{m}$, in concentration units), is a commonly used index to quantify variability. It has

been proposed that Fano factor values that increase with average concentrations indicate that the underlying transcriptional processes cannot be sufficiently explained by a simple one-step promoter configuration with purely intrinsic Poissonian noise and that extrinsic noise is likely to contribute significantly to the overall variability (Newman et al., 2006; Schwanhausser et al., 2011; Taniguchi et al., 2010). For all TFs measured, Fano factor values from 0 to 20 were obtained (Supplemental Fig. S1R), in line with Fano factor values of other TFs determined previously to lie between 0 and 30 (Sanchez et al., 2011). Moreover, the majority of TFs examined show Fano factor values, $\mathrm{F_f} > 1$, suggesting that transcriptional bursting is likely to be a significant source of the observed cell-to-cell variability. ~~$\mathrm{F_f} \approx 1 \mathrm{F_f} < 1$~~We used this dataset as a starting point for studying the control of variability during imaginal disc development.

The average concentration and variability of the investigated TFs showed no obvious interdependence (Fig. 1K), suggesting that if variability is controlled, there is not one control mechanism that is common to all investigated TFs. Among the studied TFs, the Hox protein Antennapedia (Antp), showed low variability ($CV^2 < 0.2$) in high average concentrations, in particular in the leg disc (Fig. 1K). Since low variability at the tissue level is likely to be achieved through regulatory mechanisms, we investigated Antp variability further by FCS. Because FCS performs best at low to moderate expression levels (Supplement 1), we performed this analysis in the wing disc where expression levels are lower than in the leg disc (Fig. 1K,L). We first established that the observed fluorescence intensity fluctuations were caused by diffusion of TF molecules through the detection confocal volume (Supplemental Fig. S2 and Supplement 1). FCS showed that different clusters of neighboring cells along the Antp expression domain in the wing disc display different average expression levels (Fig. 1L). Moreover, FCS showed that Antp cell-to-cell variability decreased with increasing Antp concentration (Fig. 1M) whereas the Fano factor increased (Supplemental Fig. S1R). Such behavior is indicative of complex transcriptional regulatory processes (Franz et al., 2011; Smolander et al., 2011) that we further investigated using the powerful *Drosophila* genetic toolkit.

Control of Antp concentration by transcriptional auto-regulation

One mechanism by which genes control their expression level variability is auto-regulation (Becskei and Serrano, 2000; Dublanche et al., 2006; Gronlund et al., 2013; Nevozhay et al., 2009; Shimoga et al., 2013; Thattai and van Oudenaarden, 2001). To test whether Antp can regulate its own protein levels, we monitored the concentration of endogenous Antp protein upon overexpression of *Antp* from a transgene. To distinguish between overexpressed and endogenous protein, we used synthetic Antp (SynthAntp) transgenes fused to eGFP (SynthAntp-eGFP). These transgenes encode the Antp protein (amino acids 278-378), which includes the homeodomain, the conserved YPWM motif and the C terminus (but lack the long and non-conserved N terminus of the protein, against which widely used Antp antibodies have been raised) and they harbor Antp-specific homeotic function (Papadopoulos et al., 2011). Clonal overexpression of *SynthAntp-eGFP* in the wing disc notum (Fig. 2A,B',D and controls in Supplemental Fig. S3D,D') repressed the endogenous Antp protein, indicating that Antp is indeed able to regulate its own protein levels.

Since Antp is a TF, we next asked whether the auto-repression occurs at the transcriptional level. The *Antp* locus is subject to complex transcriptional regulation, involving a distal and a proximal promoter (P1 and P2 promoters, respectively), spanning more than 100 kb of regulatory sequences. We established that the P1 promoter (rather than the P2 promoter) is predominantly required to drive expression

of Antp in the wing disc notum (Supplemental Fig. S3A-C'), in line with previous observations ((Engstrom et al., 1992; Jorgensen and Garber, 1987; Zink et al., 1991) and Materials and Methods). Moreover, mitotic recombination experiments in regions of the wing disc unique to P2 transcription have shown no function of the P2 promoter transcripts in wing disc development (Abbott and Kaufman, 1986). Thus, the P1 Antp reporter serves as a suitable reporter of the *Antp* locus transcriptional activity in this context.

Clonal overexpression of SynthAntp-eGFP in the wing disc repressed the *Antp* P1 transcriptional reporter (Fig. 2C-D and controls in Supplemental Fig. S3E,E'). To rule out putative dominant negative activity of the small SynthAntp-eGFP peptide, we also performed these experiments with the full-length Antp protein (Supplemental Fig. S3F,F') and found them to also repress the reporter. We conclude that the Antp protein is able to repress its own transcription from the P1 promoter (directly or indirectly), suggesting a possible mechanism of suppressing cell-to-cell variability of *Antp* expression levels (Fig. 2E).

In the course of these experiments, we noticed that ectopic overexpression of *SynthAntp-eGFP* or the full-length Antp protein from the *Distal-less* (*Dll*) (MD23) enhancer resulted in activation of the *Antp* P1 reporter in distal compartments of the wing disc, such as the wing pouch, where Antp is normally not detected (Fig. 2F-H and controls in Supplemental Fig. S3G-H'). This suggests that next to its auto-repressing function, Antp is also capable of activating its own transcription (Fig. 2I).

To exclude that the auto-activation and repression of Antp are artifacts of overexpression, we used FCS to measure the concentration of Antp triggered by different Gal4-drivers (Supplemental Fig. S4A-E). We observed indistinguishable DNA-binding behavior by FCS, not only across the whole concentration range examined (Supplemental Fig. S4F), but also between endogenous and overexpressed Antp (Supplemental Fig. S5A,B). Importantly, the auto-activating and auto-repressing capacity of Antp was preserved even with the weak Gal4-driver *69B* (Supplemental Fig. S4K,L) that triggered concentrations of Antp lower than its normal concentration in the leg disc ($473\ nM$ versus $1110\ nM$), indicating that auto-activation and auto-repression of Antp take place at endogenous protein concentrations.

We conclude that Antp is able to repress and activate its own transcription (Fig. 2E,I) and hypothesize that this auto-regulatory circuit sets the "correct" concentration of Antp protein in imaginal discs.

<u>A temporal switch controls the transition of *Antp* from a state of auto-activation to a state of auto-repression</u>

To further investigate the mechanism by which the Antp auto-regulatory circuit sets the precise Antp expression levels, we next asked whether the seemingly opposing auto-regulatory activities of Antp are separated in time during development. To that end, we induced gain-of-function clones of full-length untagged *Antp* either at 26 h (first larval instar – henceforth referred to as "early" stage) or at 60 h (late second larval instar – henceforth referred to as "late" stage) of development and analyzed the clones in late third instar wing imaginal discs (Fig. 3). We chose these time points based on Antp expression being widespread during first instar disc development and therefore possibly amenable to auto-activation before becoming confined to the proximal disc regions, whereas in the late second instar it is restricted to proximal only regions (Emerald and Cohen, 2004). As a pre-requisite for this analysis, we established that the Antp-eGFP homozygous viable MiMIC allele recapitulates the endogenous Antp pattern in the embryo and all thoracic imaginal discs and therefore

can be used to monitor endogenous Antp protein (Supplemental Fig. S6). Clonal induction of full-length untagged *Antp* in early development triggered strong auto-activation of Antp-eGFP (Fig. 3A,B,B' and controls in Supplemental Fig. S7A-C'). As before, we confirmed that early auto-activation of Antp is transcriptional and similar for both full-length and SynthAntp proteins (Supplemental Fig. S7D-E' and controls in F-G'). Early auto-activation was further supported by a loss-of-function experiment, where *RNAi*-mediated early knockdown of Antp resulted in downregulation of the Antp reporter (Fig. 3C,C' and controls in Supplemental Fig. S7H,H'). The loss and gain-of-function analysis together suggest that during early disc development Antp is required for sustaining its own expression.

In contrast, clonal induction during the late second instar stage (Fig. 3F) repressed Antp-eGFP (Fig. H,H') and, reciprocally, the clonal knockdown by *RNAi* triggered auto-activation of Antp transcription (Fig. 3I,I'). Hence, in contrast to early development, Antp represses its own expression in third instar discs.

While the gain-of-function experiments show that Antp is sufficient to execute auto-regulation, loss-of-function analysis indicates that it is also necessary for both repression and activation at the transcriptional level.

Together, these results revealed the existence of a switch in Antp auto-regulatory capacity on its own transcription during development. Starting from a preferentially auto-activating state early in development (Fig. 3D), Antp changes to an auto-inhibitory mode at later developmental stages (Fig. 3J).

<u>During development Antp switches from a low-concentration/high-variability to a high-concentration/low-variability state</u>

If the *Antp* auto-repressive state limits the variability of Antp protein concentration among neighboring cells late in development, we expected that the variability would be higher during earlier stages, when auto-repression does not operate. We, therefore, used FCS to characterize the endogenous expression levels and cell-to-cell variability of Antp concentration in nuclei of second instar wing and leg discs. We observed significantly lower average concentrations of Antp protein in second versus third instar wing and leg discs and the inverse was true for concentration variability (Fig. 3E and Supplemental Fig. S8A,A',C), indicating that the developmental increase in concentration is accompanied by suppression of concentration variability. In addition, FCS revealed a notable change in Antp characteristic decay times (signifying molecular diffusion, limited by chromatin-binding) at early versus late stages (Supplemental Fig. S8B). This behavior indicates that endogenous Antp is initially moving fast in the nucleus, as it undergoes considerably fewer interactions with chromatin compared to later stages where its interactions with chromatin are more frequent and longer lasting.

Taken together, our measurements show that *Antp* is expressed at relatively low and highly variable levels in early developing discs, when genetic evidence indicates auto-activation capacity on its own transcription. Later in development, when Antp has reached a state of higher average concentrations, auto-repression kicks in, resulting in considerably lower variability among neighboring cells.

<u>Dynamic control of Antp auto-regulation by different Antp isoforms</u>

The changing binding behavior of Antp on chromatin from second to third instar discs and the developmental transition from an auto-activating to an auto-repressing state suggested a causal relationship between the two phenomena. We, therefore, sought to identify molecular mechanisms that could link the observed changes in Antp

chromatin-binding to Antp auto-activation and repression. It is well established that the Antp mRNA contains an alternative splice site in exon 7 immediately upstream of the homeobox-containing exon 8, and generates Antp isoforms differing in as little as 4 amino acids in the linker between the YPWM motif (a cofactor-interacting motif) and the homeodomain (Fig. 4A) (Stroeher et al., 1988). Our previous observation that long linker isoforms favor transcriptional activation of Antp target genes, whereas short linker isoforms favor repression of Antp targets (Papadopoulos et al., 2011), prompted us to examine whether the linker length is also responsible for differences in auto-regulation.

Ectopic expression of SynthAntp-eGFP peptides featuring a long linker displayed significantly weaker repression capacity on endogenous Antp, as compared to their short linker counterparts (Fig. 4B,B',F,F' and quantified in D,H, see also Materials and Methods). We confirmed that, also in this case, the repression was at the transcriptional level (Supplemental Fig. S9I-J'). Inversely, long linker Antp isoforms exhibited stronger activation of Antp reporter, as compared to short linker isoforms (Fig. 4C,C',G,G' and quantified in D,H; see also Materials and Methods). We, additionally, validated that short linker isoforms encoded by full-length or SynthAntp cDNAs behaved as weaker auto-activating and stronger auto-repressing Antp species in all our previous experiments using the endogenous Antp protein and the P1 reporter (Supplemental Fig. S9A-H'). We conclude that, also in the case of Antp auto-regulation, short linker isoforms function as more potent repressors, whereas long linker ones operate as more potent activators.

Since the Antp P1 promoter undergoes a swich from preferential auto-activation to auto-repression, and short and long linker Antp isoforms function as preferential auto-repressors and auto-activators, respectively, it appeared possible that the switch in Antp regulation is executed at the level of transcript variant abundance of these isoforms. Therefore, we next quantified the relative abundance of long and short linker transcript variants in the embryo, second and third instar discs (Fig. 4D,H). The data showed that the abundance of the long linker variant decreased, whereas the abundance of the short linker variant increased over time in development, in line with previous observations (Stroeher et al., 1988). Thus, as hypothesized, this finding suggested that relative transcript variant abundance may underlie the switch between auto-activation and auto-repression (without excluding additional mechanisms, such as changes in the chromatin modifications between early and later disc development, or the participation of different cofactors).

Relative changes in Antp transcript variant abundance (Fig. 4D,H), differential efficiency of their encoded isoforms to repress or activate the *Antp* gene (Fig. 4B-D,F-H), the developmental switch of *Antp* from auto-activation to repression (Fig. 3) and the different mobility of Antp between second and third instar imaginal discs (Fig. 3E) all pointed towards the hypothesis that the two isoforms have different modes of interaction with chromatin. To investigate this, we expressed the two isoforms from the *69B* enhancer in third instar wing and antennal discs. This results in Antp concentrations close to (if not below) endogenous levels (Supplemental Fig. S4A-J). FCS measurements revealed that the short linker isoform displayed longer characteristic decay times and a higher fraction of DNA-bound molecules, suggesting stronger and more pronounced binding to chromatin than its long linker counterpart (Fig. 4D,H and Supplemental Fig. S10A,B). With chromatin (and therefore Antp binding sites configuration), as well as the presence of cofactor proteins, being identical between the two instances (short and long linker isoforms examined in third instar wing and antennal imaginal discs of the same age), we were able to directly

compare the apparent equilibrium dissociation constants for the two isoforms (Supplement 3). We found that the affinity of binding to chromatin ($K_d{}^{-1}$) of the repressing short linker isoform is at least 2.3 times higher compared to the activating long linker isoform ($\frac{K_{d,Antp}^{long\ linker\ isof.}}{K_{d,Antp}^{short\ linker\ isof.}} > 2.3$) (Fig. 4D,H and Supplemental Fig. S10C-D'). To corroborate these findings, we also performed gel-shift experiments to test how full-length recombinant Antp isoforms, bearing a short and a long linker, bind previously characterized Antp binding sites. We found that equal amounts of Antp long linker isoform binds *Antp* binding sites weaker than its short linker counterpart (Supplemental Fig. S11). Collectively, these experiments support the notion that differences in Antp regulation during disc development can be largely attributed to differences in the affinity of the investigated Antp isoforms.

Taken together, the switch of Antp from an auto-activating to an auto-repressing state and the alteration of its DNA-binding behavior during disc development can be largely explained by a temporal developmental regulation of the relative concentrations of preferentially auto-activating and auto-repressing Antp protein isoforms, which themselves display distinct properties in their modes of interaction with chromatin (Fig. 4E,I).

Robustness of Antp auto-regulation

The mechanism of developmental Antp auto-regulation offered a possible explanation for the observed increase in Antp concentration from second to third instar discs, as well as the suppression of variability. What remained an open question is the functional significance of suppression of Antp variability in development. To test this, we require to manipulate variability, yet this is currently not possible to achieve at the endogenous locus. However, since average concentration and variability are interdependent, we used an ectopic expression system to progressively dampen Antp variability by manipulating its concentration. To this end, we expressed SynthAntp ectopically in the antennal disc, devoid of endogenous Antp expression, and monitored the extent (strength) of homeotic transformations induced by different Gal4 drivers corresponding to different SynthAntp concentrations (as measured by FCS previously in Supplemental Fig. S4A-D). In this experiment, expression of SynthAntp is controlled by the Gal4 driver, independently of the Antp locus, therefore the phenotypic output does not depend on Antp auto-regulation. We observed that partial transformations of antennae to tarsi could be obtained with drivers expressing Antp at close to endogenous concentration (*ptc*-Gal4, *Dll*-Gal4 (MD713) and *69B*-Gal4 drivers, Fig. 5B-D and Supplemental Fig. S4B-D). Therefore, Antp can repress the antennal and launch the leg developmental program in the antennal disc at endogenous concentrations, although not robustly across the tissue (Supplement 5). As expected, the three weak transformation phenotypes, elicited by *ptc*-, *Dll* (MD713)- and *69B*-Gal4 (Fig. 5B-D) were accompanied by high variability of SynthAntp concentration in developing discs (Fig. 5E,F). In contrast, strong expression of SynthAntp from the *Dll*-Gal4 (MD23) enhancer resulted in robust homeotic transformation to a complete tarsus (Fig. 5A), accompanied by low cell to cell variability (Fig. 5F). This condition resembled most closely the endogenous Antp variability in the leg disc ($CV^2 = 0.103$). Importantly, endogenous Antp and Antp overexpressed by any of the Gal4 drivers showed indistinguishable chromatin-binding behavior by FCS (Supplemental Fig. S4F and Supplemental Fig. S5A,B).Therefore, robust Antp homeotic function can be achieved at concentrations that are accompanied by low variability.

In order to further substantiate the qualitative model of Antp auto-regulation suggested by our findings and examine its impact on protein variability, we developed a simple mathematical model of stochastic *Antp* expression (Supplement 4). This model tests whether positive and negative auto-regulation of Antp through distinct isoforms is sufficient to explain the increase in protein concentration and decrease in nucleus-to-nucleus variability from early to late stages. The model consists of a dynamic promoter, which drives transcription of *Antp* followed by a splicing step, yielding either the auto-repressing or the auto-activating isoform of Antp. Since the repressing isoform has higher abundance at later stages, we assumed that splicing is more likely to generate this isoform than the activating isoform. The initial imbalance of Antp towards the activating isoform (Fig. 4D,H) is modeled through appropriate initial concentrations of each isoform.

Since Antp copy numbers per nucleus are in the thousands at both early and late stages, intrinsic noise of gene expression is likely to explain only a certain portion of the overall variability in Antp concentrations (Elowitz et al., 2002; Taniguchi et al., 2010). The remaining extrinsic variability is due to cell-to-cell differences in certain factors affecting gene expression such as the ribosomal or ATP abundances. To check whether extrinsic variability significantly affects Antp expression, we expressed nuclear mRFP1 constitutively, alongside with endogenous Antp-eGFP, and measured their abundances (Supplemental Fig. S12). With extrinsic factors affecting both genes similarly, we expected a correlation between the concentration of nuclear mRFP1 and Antp-eGFP. Our data showed a statistically significant correlation between mRFP1 and Antp (Supplemental Fig. S11C, $r = 0.524$ and $p = 9.77 \cdot 10^{-5}$). Correspondingly, we accounted for extrinsic variability also in our model by allowing gene expression rates to randomly vary between cells (Zechner et al., 2012).

The promoter itself is modeled as a Markov chain with three distinct transcriptional states. In the absence of Antp, the promoter is inactive and transcription cannot take place (state “U” in Fig. 5G). It can switch into a highly expressing state “A” at a rate that is assumed to be proportional to the concentration of the auto-activating isoform (Antp-A, Fig. 5G). This resembles the positive auto-regulatory function of Antp. Conversely, the promoter can be repressed by recruitment of the auto-repressing isoform, state “R” in the model (Antp-R, Fig. 5G). Since the auto-repressing isoform of Antp can also activate the promoter, albeit significantly weaker than the auto-activating isoform, and vice versa, we allow the promoter to switch between states “A” and “R”.

In this promoter model, it remained unclear whether the two isoforms compete for the same binding sites on the P1 promoter. In this case, an increase in concentration of repressing Antp species enhances the probability to reach state “R” only if the promoter is in state “U” (Fig. 5G). In the absence of competitive binding, the rate of switching between “A” and “R” also depends on the concentration of repressing isoforms of Antp (Fig. 5G, compared to Supplemental Fig. S13A). We analyzed both model variants by forward simulation and found that both of them can explain the switch-like increase in average Antp concentration between early and late stages (Fig. 5J, compared to Supplemental Fig.S13D) and the relative fraction of repressing and activating isoforms (Fig. 5I, compared to Supplemental Fig. S13C). However, only the non-competitive binding model (Fig. 5G) can explain the substantial reduction of total Antp variability between early and late stages (Fig. 5J, Supplemental Fig. S13D). Simulation trajectories of individual nuclei indicated an initial increase and a subsequent stabilization of concentration, whereas in the competitive model, or in the absence of the negative feedback, this is not achieved (Fig. 5H, compared to Supplemental Fig. S13B,F). Additionally, we established that the negative feedback is

required for suppression of variability (Supplemental Fig. S13E,H), as otherwise no suppression of variability is conferred (Supplemental Fig. S13H). Thus, the model suggested that auto-repression is required and that isoforms do not compete for binding to the P1 promoter.

To further validate this model, we analyzed how Antp variability scales with average concentrations, compared to our experimental measurements. To generate different average concentrations, we varied the gene expression rates over three orders of magnitude. The model predicted a decrease in variability as a function of total Antp concentration and an increase in the Fano factor. These findings are in good agreement with the experimental data (compare Fig. 5K to K' and L to L').

We next analyzed the model behavior under different genetic perturbations. Increase of Antp concentration by overexpressing SynthAntp transgenes (bearing either a long or a short linker isoform) from the Antp P1 promoter (Antp P1-Gal4>UAS-SynthAntp-eGFP long or short linker) resulted in 100% embryonic lethality, rendering the analysis of concentration and variability in imaginal discs impossible. This indicated that indiscriminate increase of the dosage of either Antp variant from early embryonic development onwards cannot be tolerated or buffered by the auto-regulatory circuit.

However, overexpression from a *Dll* enhancer [*Dll*-Gal4 (MD23)] in the leg discs or in the notum (*MS243*-Gal4), which overlaps with the endogenous Antp expression pattern only during first instar disc development (Emerald and Cohen, 2004), resulted in normal adult leg and notum structures. Flies overexpressing either the SynthAntp auto-activating or the auto-repressing isoform in distal appendages (Fig. 6A,B) or the notum (Supplemental Fig. S14A) displayed the wild type morphology, indicative of normal Antp function, regardless of which isoform (activating or repressing) was overexpressed. We further measured by FCS the concentration and variability of the total Antp protein (endogenous Antp-eGFP and overexpressed SynthAntp-eGFP) in proximal regions of the leg disc at second and third instar stages (Fig. 6C,C'). We found that the concentration remained high at both stages due to the overexpression but variability was reduced to endogenous levels at late stages. Also, the reduced Antp variability does not seem to depend on Antp concentration alone, because for high concentrations at both early and late stages, variability is high only in the early stage but reduced in the late stage. Together, the phenotypic analysis and FCS measurements indicate that Antp auto-regulation is able to reduce variability, even at high levels of expression of either isoform, ensuring proper leg development.

The experimental data were corroborated by the model, which predicted that more than three-fold overexpression of either auto-activating or auto-repressing Antp isoforms (Fig. 6E,H) will nevertheless equilibrate to normal expression levels at later stages (Fig. 6D,F,G,I). Specifically, we have measured by FCS roughly 15400 molecules in the wild type leg disc, and the model is in good quantitative agreement with this measurement upon overexpression of the activating or repressing isoform. In addition, there is no negative effect on the noise suppressing property of the circuit (Fig. 6F,I). Thus, both the model and experimental data indicate that transient high levels of either isoform early during disc development can be tolerated and that the concentration and cell-to-cell variability of the endogenous Antp protein is restored at later stages.

In contrast, overexpression of an exogenous repressor, such as Sex combs reduced (Scr), which can repress Antp at the transcriptional level, but can neither activate it nor activate its own transcription (Supplemental Fig. S14E-J'), resulted in abnormal leg (Fig. 6J) and notum (Supplemental Fig. S14B) development. These flies

died as pharate adults with malformed legs, in line with Antp being required for proper leg development in all ventral thoracic discs (legs). FCS measurements in the corresponding proximal leg disc cell nuclei of second and third instar leg discs overexpressing mCherry-SynthScr revealed pronounced reduction in Antp concentration and remarkable increase in variability (Fig. 6K). In agreement, the model predicted a similar block of transcription and correspondingly severe effects on Antp dynamics (Fig. 6L-N). In both the measurements and the model prediction, the high increase in variability was triggered by the fact that a majority of the cells do not manage to switch into the highly-expressing state since too few long-linker Antp molecules are present to establish the positive auto-regulation. Since splicing favors the short-linker isoforms at later stages, these cells never “recover” from Scr repression after restriction of the Antp overexpression domain to proximal regions of the leg disc (Fig. 6L).

Taken together, the minimal model of Antp auto-regulatory genetic circuit is able to explain the experimentally observed differences in Antp concentration and cell-to-cell variability at early and late developmental stages.

Discussion

In this work, we found that *Antp* auto-regulates its expression levels development, starting from a preferentially auto-activating state early and transitioning to a preferentially auto-repressing state later. The early state is characterized by lower average Antp concentrations and high variability, whereas the opposite is true for the later repressing state. Without excluding other mechanisms, such as chromatin configuration, accessibility of Hox binding sites to Antp, the differential abundance of cofactors among developmental stages, or different modes of interactions with different Antp isoforms, we have shown that differential expression of Antp isoforms is one contributing mechanism for the observed regulatory switch. These isoforms have preferentially activating or repressing activities on the Antp promoter, bind chromatin with different affinities and are themselves expressed in different relative amounts during development. A loss-of-function analysis of the isoforms *in vivo* will be required to provide a definitive answer on the relative contribution of the Antp isoform-mediated auto-regulatory circuit towards observed suppression of variability. CRISPR/Cas9-mediated genome manipulation, in principle, allows the generation of Antp loci that express only one or the other isoform. However, it is not clear whether these flies can reach the larval developmental stages, given the Antp embryonic functions and, in fact, strong biases towards only the activating or repressing isoform introduced by Antp-Gal4-mediated expression of either Antp isoform resulted in embryonic lethality. In the absence of such direct evidence, we turned to mathematical modelling and derived, based on our experimental data, a simple kinetic model of *Antp* auto-regulation that confirmed the plausibility of the proposed mechanism. In addition, the model generated predictions that could be verified by introducing genetic perturbations.

Negative auto-regulation has been identified as a frequently deployed mechanism for the reduction of noise (cell-to-cell variability) and the increase of regulatory robustness in various systems (Becskei and Serrano, 2000; Dublanche et al., 2006; Gronlund et al., 2013; Nevozhay et al., 2009; Shimoga et al., 2013; Thattai and van Oudenaarden, 2001). Auto-repression has been described for the Hox gene *Ultrabithorax* (*Ubx*) in haltere specification and as a mechanism of controlling Ubx levels against genetic variation (Crickmore et al., 2009; Garaulet et al., 2008), as well as in *Ubx* promoter regulation in Drosophila S2 cells (Krasnow et al., 1989). In contrast,

an auto-activating mechanism is responsible for the maintenance of *Deformed* expression in the embryo (Kuziora and McGinnis, 1988). These experiments suggest similar mechanisms for establishing (auto-activation) or limiting (auto-repression) Hox TF levels and variability in different developmental contexts.

Our data suggest that the developmental switch from auto-activation to auto-repression is, at least in part, mediated by molecularly distinct Antp linker isoforms. Differences in affinities of different Hox TF isoforms, based on their linker between the YPWM motif and the homeodomain, have also been identified for the Hox TF Ubx. Interestingly, its linker is also subject to alternative splicing at the RNA level (Reed et al., 2010). In a similar way to Antp, the long linker Ubx isoform displays four to five fold lower affinity of DNA binding, as compared to short linker isoforms, and the two isoforms are not functionally interchangeable in *in vivo* assays. Finally, the Ubx linker also affects the strength of its interaction with the Hox cofactor Extradenticle (Exd), underscoring the functional importance of linker length in Hox TF function (Saadaoui et al., 2011).

Mathematical modeling predicts that the Antp auto-regulatory circuit is robust with respect to initial conditions and extrinsic noise by suppressing cell-to-cell concentration variability even at high concentrations of any of the two Antp isoforms (auto-repressing or auto-activating). This “buffering” capacity on cell-to-cell variability is reflected in the ability of flies to tolerate more than 3-fold overexpression of Antp without dramatic changes in endogenous Antp levels or generation of abnormal phenotypes. Therefore, two different isoforms produced from the same gene with opposing roles in transcriptional regulation and different auto-regulatory binding sites on the gene’s promoter seem to suffice to create a robust gene expression circuit that is able to “buffer” perturbations of the starting conditions.

So far, we have only been able to indiscriminately increase or decrease Antp concentration at the tissue level and record the phenotypic outcome of these boundary states. It will be interesting to test whether controlled perturbations of TF variability at the tissue level that render TF concentration patterns less, or more, noisy among neighboring cells, while maintaining similar mean protein concentrations, lead to abnormal phenotypes. The technology to selectively manipulate expression variability of specific TF in a developing tissue is yet to be established.

Materials and Methods

Fly stocks used

The Antp-eGFP MiMIC line has been a kind gift from Hugo J. Bellen (Bloomington stock 59790). The *atonal* (VDRC ID 318959), *brinker* (VDRC ID 318246), *spalt major* (VDRC ID 318068), *yorkie* (VDRC ID 318237), *senseless* (VDRC ID 318017) and *Sex combs reduced* (VDRC ID 318441) fosmid lines are available from the Vienna Drosophila Resource Center (VDRC) and have been generated recently in our laboratory (Sarov et al., 2016). The *fork head* (stock 43951), *grainy head* (stock 42272), *Abdominal B* (stock 38625), *eyeless,* (stock 42271), *spineless* (transcript variant A, stock 42289), and *grain* (stock 58483) tagged BACs were generated by Rebecca Spokony and Kevin P. White and are available at the Bloomington Stock Center. For the *scalloped* gene, a GFP-trap line was used (Buszczak et al., 2007), a kind gift from Allan C. Spradling laboratory (line CA07575), with which genome-wide chromatin immunoprecipitation experiments have been performed (Slattery et al., 2013). For the *spineless* gene, Bloomington stock 42676, which tags isoforms C and D of the Spineless protein has been also tried in

fluorescence imaging and FCS experiments, but did not yield detectable fluorescence in the antennal disc, rendering it inappropriate to be used in our analysis. Therefore, we resided to stock 42289, which tags the A isoform of the protein. For the *eyeless* gene, the FlyFos015860(pRedFlp-Hgr)(ey13630::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT line (VDRC ID 318018) has been tried also in fluorescence imaging and FCS experiments, but did not yield detectable fluorescence in the eye disc for it to be used in our analysis. The *act5C*-FRT-yellow-FRT-Gal4 (Ay-Gal4) line used for clonal overexpression or *RNAi* knockdown has been described (Ito et al., 1997). The UAS-Antp lines (synthetic and full-length), as well as UAS-SynthScr constructs have been previously described (Papadopoulos et al., 2011; Papadopoulos et al., 2010). The *Dll*-Gal4 (MD23) line has been a kind gift of Ginés Morata (Calleja et al., 1996). *69B*-Gal4 and *ptc*-Gal4 have been obtained from the Bloomington Stock Center. The *Antp* P1-*lacZ* and P2-*lacZ* have been previously described (Engstrom et al., 1992; Zink et al., 1991). The P1 reporter construct spans the region between 9.4 kb upstream of the P1 promoter transcription initiation site and 7.8 kb downstream into the first intron, including the first exon sequences and thus comprising 17.2 kb of Antp regulatory sequences (pAPT 1.8). The line used has been an insertion of the pAPT 1.8 vector bearing the P1 promoter regulatory sequences upstream of an *actin-lacZ* cytoplasmic reporter and has been inserted in cytogenetic location 99F on the right chromosomal arm of chromosome 3. The *Antp-RNAi* line has been from VDRC, line KK101774. UAS-eGFP stock was a kind gift of Konrad Basler. We are indebted to Sebastian Dunst for generating the *ubi*-FRT-mCherry(stop)-FRT-Gal4(VK37)/CyO line, which drives clonal overexpression upon flippase excision, while simultaneously marking cells by the loss of mCherry. For red-color labeling of clones the *act5C*-FRT-CD2-FRT-Gal4, UAS-mRFP1(NLS)/TM3 stock 30558 from the Bloomington Stock Center has been used. For marking the ectopic expression domain of untagged Antp proteins the UAS-mRFP1(NLS)/TM3 stock 31417 from the Bloomington Stock Center has been used. The *MS243*-Gal4; UAS-GFP/CyO line was a kind gift from the laboratory of Ernesto Sánchez-Herrero.

<u>Fly genotypes corresponding to fluorescence images</u>

Supplemental Fig. S1A: FlyFos018487(pRedFlp-Hgr)(ato37785::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1B: FlyFos024884(pRedFlp-Hgr)(brk25146::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1C: FlyFos030836(pRedFlp-Hgr)(salm30926::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1: FlyFos029681(pRedFlp-Hgr)(yki19975::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1E: w^{1118}; PBac(fkh-GFP.FPTB)VK00037/SM5
Supplemental Fig. S1F: *sd*-eGFP (FlyTrap, homozygous)
Supplemental Fig. S1G: w^{1118}; PBac(grh-GFP.FPTB)VK00033
Supplemental Fig. S1H: FlyFos018974(pRedFlp-Hgr)(Scr19370::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1I: FlyFos015942(pRedFlp-Hgr)(sens31022::2XTY1-SGFP-V5-preTEV-BLRP-3XFLAG)dFRT
Supplemental Fig. S1J,K: Antp-eGFP (MiMIC) homozygous (line MI02272, converted to an artificial exon)
Supplemental Fig. S1L: w^{1118}; PBac(Abd-B-EGFP.S)VK00037/SM5
Supplemental Fig. S1M: w^{1118}; PBac(ey-GFP.FPTB)VK00033

Supplemental Fig. S1N: w^{1118}; PBac(ss-GFP.A.FPTB)VK00037
Supplemental Fig. S1O,P: w^{1118}; PBac(grn-GFP.FPTB)VK00037
Fig. 2B,B': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Fig. 2C,C': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; UAS-Antp long linker (full-length, untagged)/+
Fig. 2G,G': *Dll*-Gal4 (MD23)/+; UAS-SynthAntp-eGFP/*Antp* P1-*lacZ*
Supplemental Fig. S3A,A': *Antp* P1-*lacZ*/TM3
Supplemental Fig. S3B,B': *Antp* P2-*lacZ*/CyO
Supplemental Fig. S3C,C': wild type
Supplemental Fig. S3D,D': *hs*-flp; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP
Supplemental Fig. S3E,E': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; *Antp* P1-*lacZ*/+
Supplemental Fig. S3F,F': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; UAS-Antp long linker (full-length, untagged)/*Antp* P1-*lacZ*
Supplemental Fig. S3G,G': *Dll*-Gal4 (MD23)/+; UAS-Antp long linker (full-length, untagged), UAS-mRFP1(NLS)/ *Antp* P1-*lacZ*
Supplemental Fig. S3H,H': *Dll*-Gal4 (MD23)/+; UAS-mRFP1(NLS)/ *Antp* P1-*lacZ*
Supplemental Fig. S4A: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp long linker-eGFP/+
Supplemental Fig. S4B: *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Supplemental Fig. S4C: *Dll*-Gal4 (MD713)/+; UAS-SynthAntp long linker-eGFP/+
Supplemental Fig. S4D,G,H,K: *69B*-Gal4/UAS-SynthAntp long linker-eGFP
Supplemental Fig. S4I,J,L: *69B*-Gal4/UAS- eGFP
Fig. 3B,B',G,G': *hs*-flp/+; *ubi*-FRT-mChery-FRT-Gal4/+; Antp-eGFP (MiMIC)/UAS-Antp long linker (full-length, untagged)
Fig. 3C,C': *hs*-flp/+; UAS-Antp^{RNAi}/+; *Antp* P1-*lacZ*/*act5C*-FRT-CD2-FRT-Gal4, UAS-mRFP1(NLS)
Fig. 3H,H': *hs*-flp/+; UAS-Antp^{RNAi}/*act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP; *Antp* P1-*lacZ*/+
Supplemental Fig. S6B: *Antp* P1-*lacZ*/TM6B
Supplemental Fig. S7A,A': *hs*-flp/+; *ubi*-FRT-mChery-FRT-Gal4/+; Antp-eGFP (MiMIC)/UAS-Antp long linker (full-length, untagged)
Supplemental Fig. S7B-C': *hs*-flp/+; *ubi*-FRT-mChery-FRT-Gal4/+; Antp-eGFP (MiMIC)/+
Supplemental Fig. S7D,D': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; *Antp* P1-*lacZ*/UAS-Antp long linker (full-length, untagged)
Supplemental Fig. SE,E': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Supplemental Fig. S7F,F': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+
Supplemental Fig. S7G,G': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4, UAS-eGFP/+; *Antp* P1-*lacZ*/+
Supplemental Fig. S7H,H': *hs*-flp/+; UAS-Antp^{RNAi}/+; Antp-eGFP (MiMIC)/*act5C*-FRT-CD2-FRT-Gal4, UAS-mRFP1(NLS)
Fig. 4B,B': *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Fig. 4C,C': *Dll*-Gal4 (MD23)/+; UAS-SynthAntp long linker-eGFP/*Antp* P1-*lacZ*
Fig. 4F,F': *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Fig. 4G,G': *Dll*-Gal4 (MD23)/+; UAS-SynthAntp short linker-eGFP/*Antp* P1-*lacZ*
Supplemental Fig. S9A,A': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp short linker-eGFP/+

Supplemental Fig. S9B,B',G,G': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-SynthAntp short linker-eGFP/*Antp* P1-*lacZ*
Supplemental Fig. S9C,C',H,H': *hs*-flp/+; *act5C*-FRT-yellow-FRT-Gal4/+; UAS-Antp short linker (full-length, untagged)/*Antp* P1-*lacZ*
Supplemental Fig. S9D,D': *hs*-flp/+; *Dll*-Gal4 (MD23)/+; UAS-Antp short linker (full-length, untagged), UAS-mRFP1(NLS)/*Antp* P1-*lacZ*
Supplemental Fig. S9E-F': *hs*-flp/+; *ubi*-FRT-mChery-FRT-Gal4/+; Antp-eGFP (MiMIC)/UAS-Antp short linker (full-length, untagged)
Supplemental Fig. S9I,I': *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/*Antp* P1-*lacZ*
Supplemental Fig. S9J,J': *ptc*-Gal4/+; UAS-SynthAntp short linker-eGFP/*Antp* P1-*lacZ*
Fig. 5A: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp long linker-eGFP/+
Fig. 5B: *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Fig. 5C: *Dll*-Gal4 (MD713)/+; UAS-SynthAntp long linker-eGFP/+
Fig. 5D: *69B*-Gal4/UAS-SynthAntp long linker-eGFP
Supplemental Fig. S12A-B': *ubi*-mRFP1(NLS)/+ or y; Antp-eGFP (MiMIC)/+
Supplemental Fig. S13B,C: *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/+
Fig. 6A,B: *Dll*-Gal4 (MD23)/+; UAS-SynthAntp long linker-eGFP/+ or *Dll*-Gal4 (MD23)/+; UAS-SynthAntp short linker-eGFP/+
Fig. 6J: *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/+
Supplemental Fig. S14A,D,D': *MS243*-Gal4/+; UAS-SynthAntp long linker-eGFP/*Dr* or *MS243*-Gal4/+; UAS-SynthAntp short linker-eGFP/*Dr*
Supplemental Fig. S14B,F,F': *MS243*-Gal4/+; UAS-mCitrine-SynthScr/+
Supplemental Fig. S14C,I,I': *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/+
Supplemental Fig. S14E,E': *ptc*-Gal4/+; UAS-SynthAntp long linker-eGFP/+
Supplemental Fig. S14F,F': *MS243*-Gal4/+; UAS- mCitrine-SynthScr/+
Supplemental Fig. S14G,G': *ptc*-Gal4/+; UAS- mCitrine-SynthScr/*Antp* P1-*lacZ*
Supplemental Fig. S14H,H': *Dll*-Gal4 (MD23)/+; UAS-mCitrine-SynthScr/*Antp* P1-*lacZ*
Supplemental Fig. S14J,J': *MS243*-Gal4/+; UAS-eGFP/+

Preparation of second and third instar imaginal discs for FCS measurements

For FCS measurements, imaginal discs (eye-antennal, wing, leg, humeral and genital) and salivary glands were dissected from third instar wandering larvae, or wing and leg discs from second instar larvae, in Grace's insect tissue culture medium (ThermoFisher Scientific, 11595030) and transferred to 8-well chambered coverglass (Nunc® Lab-Tek™, 155411) containing PBS just prior to imaging or FCS measurements. Floating imaginal discs or salivary glands were sunk to the bottom of the well using forceps.

Immunostainings in larval imaginal discs

Larval imaginal discs were stained according to (Papadopoulos et al., 2010). Stainings for the endogenous Antp protein have been performed using a mouse anti-Antp antibody (Developmental Studies Hybridoma Bank, University of Iowa, anti-Antp 4C3) in a dilution of 1:250 for embryos and 1:500 for imaginal discs. eGFP, or eGFP-tagged proteins have been stained using mouse or rabbit anti-GFP antibodies from ThermoFisher Scientific in a dilution of 1:500 in imaginal discs and 1:250 in embryos. mRFP1 was stained using a Chromotek rat anti-RFP antibody. For *Antp* P1 promoter stainings in imaginal discs we used the mouse anti-β-galactosidase 40-1a antibody from Developmental Studies Hybridoma Bank, University of Iowa in a dilution of 1:50. The rabbit anti-Scr antibody was used in a dilution of 1:300 (LeMotte et al., 1989). Confocal images of antibody stainings represent predominatly Z-projections and Zeiss

LSM510, Zeiss LSM700 or Zeiss LSM880 Airyscan confocal laser scanning microscopy systems with an inverted stand Axio Observer microscope were used for imaging. Image processing and quantifications have been performed in Fiji (Schindelin et al., 2012). For optimal spectral separation, secondary antibodies coupled to Alexa405, Alexa488, Alexa594 and Cy5 (ThermoFischer Scientific) were used.

Colocalization of wild type and eGFP-tagged MiMIC Antp alleles in imaginal discs

To examine whether the pattern of the MiMIC Antp-eGFP fusion protein recapitulates the Antp wild type expression pattern in both embryo and larval imaginal discs, we performed immunostainings of heterozygous Antp-eGFP and wild type flies to visualize the embryonic (stage 13) and larval expression of *Antp* and eGFP. In this experiment, we 1) visualized the overlap between eGFP and *Antp* (the eGFP pattern reflects the protein encoded by the MiMIC allele, whereas the *Antp* pattern reflects the sum of protein produced by the MiMIC allele and the allele of the balancer chromosome) and 2) compared the eGFP expression pattern to the Antp expression pattern in wild type discs and embryos.

Induction of early and late overexpression and *RNAi*-knockdown clones in imaginal discs

Genetic crosses with approximately 100 virgin female and 100 male flies were set up in bottles and the flies were allowed to mate for 2 days. Then, they were transferred to new bottles and embryos were collected for 6 hours at 25°C. Flies were then transferred to fresh bottles and kept until the next collection at 18°C. To asses Antp auto-activation, the collected eggs were allowed to grow at 25°C for 26 h from the midpoint of collection, when they were subjected to heat-shock by submersion to a water-bath of 38°C for 30 min and then placed back at 25°C until they reached the stage of third instar wandering larvae, when they were collected for dissection, fixation and staining with antibodies. To assess Antp auto-repression, the same procedure was followed, except that the heat-shock was performed at 60 h of development after the midpoint of embryo collection. Whenever necessary, larval genotypes were selected under a dissection stereomicroscope with green and red fluorescence filters on the basis of *deformed* (*dfd*)-YFP bearing balancer chromosomes (Le et al., 2006) and visual inspection of fluorescence in imaginal discs.

Measurement of Antp transcript variant abundance

The linker between the Antp YPWM motif and the homeodomain contains the sequence RSQFGKCQE. Short linker isoforms encode the sequence RSQFE, whereas long linker isoforms are generated by alternative splicing of a 12 base pair sequence encoding the four amino acid sequence GKCQ into the mRNA. We initially designed primer pairs for RT-qPCR experiments to distinguish between the short and long linker mRNA variants. For the short linker variant, we used nucleotide sequences corresponding to RSQFERKR (with RKR being the first 3 amino acids of the homeodomain). For detection of the long linker variant we designed primers either corresponding to the RSQFGKCQ sequence, or GKCQERKR. We observed in control PCRs (using plasmid DNA harboring either a long or a short linker cDNA) that primers designed for the short linker variant still amplified the long linker one. Moreover, with linker sequences differing in only four amino acids, encoded by 12 base pars, primer pairs flanking the linker could also not be used, since, due to very similar sizes, both variants would be amplified in RT-qPCR experiments with almost equal efficiencies. Therefore, we used primer pairs flanking the linker region to indiscriminately amplify

short and long linker variants, using non-saturating PCR (18 cycles) on total cDNA generated from total RNA. We then resolved and assessed the relative amounts of long and short linker amplicons in a second step using Fragment Analyzer (Advanced Analytical). RNA was extracted from stage 13 embryos, second instar larvae at 60 h of development, and leg or wing discs from third instar wandering larvae using the Trizol® reagent (ThermoFischer Scientific), following the manufacturer's instructions. Total RNA amounts were measured by NanoDrop and equal amounts were used to synthesize cDNA using High-Capacity RNA-to-cDNA™ Kit (ThermoFischer Scientific), following the manufacturer's instructions. Total cDNA yields were measured by NanoDrop and equal amounts were used in PCR, using in-house produced Taq polymerase. 10 ng of plasmid DNA, bearing either a long or a short transcript cDNA were used as a control. PCR product abundance was analyzed both by agarose gel electrophoresis and using Fragment Analyzer (Advanced Analytical).

The quantification of the transcript variant concentration (Fig. 4 D and H) has been made considering 100% (value equal to 1 on the y axis) as the sum of long and short isoforms at each developmental stage, whereas the quantification of the relative activation and repression efficiency has been performed considering the short linker variant as having 100% repression and the long linker variant as having 100% activation (values equal to 1 on the y-axis) efficiency.

Quantification of the relative repressing and activating efficiencies of different Antp isoforms

Quantification of the relative efficiency of Antp activating and repressing isoforms (Fig. 4D,H) were performed in Fiji (Schindelin et al., 2012) by outlining the total region of repression or activation of Antp protein or P1 reporter staining and quantification of the relative fluorescence intensity of the selected regions. From the calculated values, we have subtracted the values obtaining by outlining and calculating Antp protein or reporter beta-galactosidase staining background in the region of expression of an eGFP transgene alone (negative control). 5-7 imaginal disc images per investigated genotype were used for analysis. For the repression assay the obtained values have been normalized over the intensity of Antp protein calculated in the region of overlap between an eGFP expressing transgene and Antp (negative control). In both cases (repression and activation), the highest efficiency per transcript variant (for repression, the short linker isoform; for activation the long linker isoform) have been set to 100%.

Fluorescence Microscopy Imaging of live imaginal discs and FCS

Fluorescence imaging and FCS measurements were performed on two uniquely modified confocal laser scanning microscopy systems, both comprised of the ConfoCor3 system (Carl Zeiss, Jena, Germany) and consisting of either an inverted microscope for transmitted light and epifluorescence (Axiovert 200 M); a VIS-laser module comprising the Ar/ArKr (458, 477, 488 and 514 nm), HeNe 543 nm and HeNe 633 nm lasers and the scanning module LSM510 META; or a Zeiss LSM780 inverted setup, comprising Diode 405 nm, Ar multiline 458, 488 and 514 nm, DPSS 561 nm and HeNe 633 nm lasers. Both instruments were modified to enable detection using silicon Avalanche Photo Detectors (SPCM-AQR-1X; PerkinElmer, USA) for imaging and FCS. Images were recorded at a 512X512 pixel resolution. C-Apochromat 40x/1.2 W UV-VIS-IR objectives were used throughout. Fluorescence intensity fluctuations were recorded in arrays of 10 consecutive measurements, each measurement lasting 10 s. Averaged curves were analyzed using the software for online data analysis or

exported and fitted offline using the OriginPro 8 data analysis software (OriginLab Corporation, Northampton, MA). In either case, the nonlinear least square fitting of the autocorrelation curve was performed using the Levenberg–Marquardt algorithm. Quality of the fitting was evaluated by visual inspection and by residuals analysis. Control FCS measurements to asses the detection volume were routinely performed prior to data acquisition, using dilute solutions of known concentration of Rhodamine 6G and Alexa488 dyes. The variability between independent measurements reflects variabilitys between cells, rather than imprecision of FCS measurements. For more details on Fluorescence Microscopy Imaging and FCS, refer to Supplement 1.

In Figure 1A-H the workflow of FCS measurements is schematically represented. Live imaging of imaginal discs, expressing endogenously-tagged TFs, visualized by fluorescence microscopy and neighboring cells, expressing TFs at different levels, selected for FCS measurements (Fig. 1A-B). FCS measurements are performed by placing the focal point of the laser light into the nucleus (Fig 1C-D) and recording fluorescence intensity fluctuations (Fig. 1E), generated by the increase or decrease of the fluorescence intensity, caused by the arrival or departure of fast- and slowly-diffusing TF molecules into or out of the confocal detection volume (Fig. 1D). The recorded fluctuations are subjected to temporal autocorrelation analysis, which generates temporal autocorrelation curves (henceforth referred to as FCS curves), which by fitting with an appropriate model (Supplement 1), yield information about the absolute concentration of fluorescent molecules (F) and, after normalization to the same amplitude, their corresponding diffusion times, as well as the fraction of fast- and slowly-diffusing TF molecules (Fig. 1G). The concentration of molecules is inversely proportional to the y-axis amplitude at the origin of the FCS curve (Fig. 1F). Processes that slow down the diffusion of TF molecules, such as binding to very large molecules (e.g. chromosomal DNA), are visible by a shift of the FCS curves to longer characteristic times (Fig. 1G). Measurements in a collection of neighboring cell nuclei also allow the calculation of protein concentration variability at the live tissue level (Fig. 1H).

Sample size, biological and technical replicates

For the measurement of TF molecular numbers and variability (Fig. 1 and Supplemental Fig. S1), 7-10 larvae of each fly strain were dissected, yielding at least 15 imaginal discs, which were used in FCS analysis. For the Fkh TF, 7 pairs of salivary glands were analyzed and for AbdB, 12 genital discs were dissected from 12 larvae. More than 50 FCS measurements were performed in patches of neighboring cells of these dissected discs, in the regions of expression indicated in Supplemental Fig. S1 by arrows. Imaginal discs from the same fly strain (expressing a given endogenously-tagged TF) were analyzed on at least 3 independent instances (FCS sessions), taking place on different days (biological replicates) and for Antp, which was further analyzed in this study, more than 20 independent FCS sessions were used. As routinely done with FCS measurements in live cells, these measurements were evaluated during acquisition and subsequent analysis and, based on their quality (high counts per molecule and second, low photobleaching), were included in the calculation of concentration and variability. In Supplemental Fig. S1Q, *n* denotes the number of FCS measurements included in the calculations.

For experiments involving immunostainings in imaginal discs to investigate the auto-regulatory behavior of Antp (Figs. 2-5 and supplements thereof, except for the temporally-resolved auto-activating and repressing study of Antp in Fig. 3, as discussed above), 14-20 male and female flies were mated in bottles and 10 larvae

were selected by means of fluorescent balancers and processed downstream. Up to 20 imaginal discs were visualized by fluorescence microscopy and high resolution Z-stacks were acquired for 3-5 representative discs or disc regions of interest per experiment. All experiments were performed in triplicate, except for the temporal analysis of Antp auto-regulatory behavior in Fig. 3 (and Supplemental Figs. thereof), which was performed 6 times and the quantification of repression efficiency of short and long linker Antp isoforms in Fig. 4 (and Supplemental Figs. thereof), which was performed 5 times.

For the quantification of transcript variant abundance in Fig. 4D,H, RNA and thus cDNA was prepared from each stage 3 independent times (biological replicates) and the transcript abundance per RNA/cDNA sample was also analyzed 3 times.
For the experiments involving perturbations in Antp expression whereby the proper development of the leg and the notum have been assessed in Fig. 5, more than 100 adult flies have been analyzed and this experiment has been performed more than 10 times independently.

Statistical significance

Fig. 2D: Statistical significance was determined using a two-tailed Student's T-test (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{repression\ clone\ vs\ surrounding}_{Antp\ protein} = 1.36 \cdot 10^{-15}$ and $p^{repression\ clone\ vs\ surrounding}_{Antp\ reporter} = 3.17 \cdot 10^{-16}$).
Fig. 2I: Statistical significance was determined using a two-tailed Student's T-test (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{Dll\ expression\ domain\ vs\ surrounding}_{Antp\ reporter} = 1.55 \cdot 10^{-17}$).
Fig. 3E: Statistical significance was determined using a two-tailed Student's T-test (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{early\ activation\ clone\ vs\ surrounding}_{Antp\ protein} = 6.23 \cdot 10^{-13}$ and $p^{early\ knockdown\ (RNAi)\ clone\ vs\ surrounding}_{Antp\ reporter} = 2.98 \cdot 10^{-9}$).
Fig. 3F: Statistical significance was determined using a two-tailed Student's T-test (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{2nd\ vs\ 3rd\ instar}_{concentration} = 2.04 \cdot 10^{-20}$, $p^{2nd\ vs\ 3rd\ instar}_{\tau D2} = 7.2 \cdot 10^{-4}$ and $p^{2nd\ vs\ 3rd\ instar}_{variation} = 3.4 \cdot 10^{-2}$).
Fig. 3K: Statistical significance was determined using a two-tailed Student's T-test (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{late\ repression\ clone\ vs\ surrounding}_{Antp\ protein} = 3.98 \cdot 10^{-17}$ and $p^{late\ knockdown\ (RNAi)\ clone\ vs\ surrounding}_{Antp\ reporter} = 1.16 \cdot 10^{-21}$).
Fig. 4D, H: Statistical significance was determined using a two-tailed Student's T-test between measurements performed with the long linker (auto-activating) isoform (Fig. 4D) and the short linker (auto-repressing) isoform (Fig. 4H) (***, $p < 0.001$ and *, $p < 0.05$, namely $p^{embryo\ long\ vs\ short}_{concentration} = 3.16 \cdot 10^{-5}$, $p^{2nd\ instar\ long\ vs\ short}_{concentration} = 1.16 \cdot 10^{-4}$, $p^{3rd\ instar\ long\ vs\ short}_{concentration} = 2.85 \cdot 10^{-6}$, $p^{long\ vs\ short}_{relative\ activation} = 4.1 \cdot 10^{-3}$, $p^{long\ vs\ short}_{relative\ repression} = 2.4 \cdot 10^{-4}$ and $p^{long\ vs\ short}_{DNA-bound\ fraction\ (FCS)} = 5.6 \cdot 10^{-10}$).
Fig. 6C-C': Statistical significance was determined using a two-tailed Student's T-test ($p^{early\ vs\ late\ conc.}_{leg\ disc,\ o/e\ activator} = 0.679$ and $p^{early\ vs\ late\ conc.}_{leg\ disc,\ o/e\ repressor} = 0.454$).
Fig. 6K: Statistical significance was determined using a two-tailed Student's T-test ($p^{early\ vs\ late\ conc.}_{leg\ disc,\ o/e\ exog.\ repr.} = 0.892$).

Acknowledgements

We are deeply saddened by the unexpected passing of Prof. Walter J. Gehring, at the very inception of this work, when the project was still in the planning and

preliminary data gathering stage. Prof. Gehring was an extraordinary human being and a scientific giant, whose work will continue to educate and inspire generations to come. DKP would like to express his gratitude to PT for outstanding scientific, and uninterrupted financial, support. DKP would like to acknowledge Markus Burkhardt, Sylke Winkler, Aliona Bogdanova and the Light Microscopy facility of MPI-CBG. DKP is also grateful to Konstantinos Papadopoulos for advice on the mathematical analysis. The authors thank Wendy Bickmore, Jan Brugues and Thomas M. Schultheiss for critical comments on the manuscript.

Competing interests

No competing interests declared.

Funding

DKP has been supported by a long-term fellowship from the Swiss National Science Foundation (PBBSP-138700), a long-term fellowship from the Federation of European Biochemical Societies (FEBS), the Max-Planck society (through funds to PT) and a Chancellor's Fellowship from the University of Edinburgh. VV has been supported by the Knut and Alice Wallenberg foundation and Karolinska Institute Research Funds.

References

Abbott, M. K. and Kaufman, T. C. (1986). The relationship between the functional complexity and the molecular organization of the Antennapedia locus of Drosophila melanogaster. *Genetics* **114**, 919-942.

Baumgartner, S., Littleton, J. T., Broadie, K., Bhat, M. A., Harbecke, R., Lengyel, J. A., Chiquet-Ehrismann, R., Prokop, A. and Bellen, H. J. (1996). A Drosophila neurexin is required for septate junction and blood-nerve barrier formation and function. *Cell* **87**, 1059-1068.

Becskei, A. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593.

Blake, W. J., M, K. A., Cantor, C. R. and Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature* **422**, 633-637.

Boettiger, A. N. and Levine, M. (2013). Rapid transcription fosters coordinate snail expression in the Drosophila embryo. *Cell reports* **3**, 8-15.

Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., Skora, A. D., Nystul, T. G., Ohlstein, B., Allen, A., et al. (2007). The carnegie protein trap library: a versatile tool for Drosophila developmental studies. *Genetics* **175**, 1505-1531.

Calleja, M., Moreno, E., Pelaz, S. and Morata, G. (1996). Visualization of gene expression in living adult Drosophila. *Science* **274**, 252-255.

Clark, N. M., Hinde, E., Winter, C. M., Fisher, A. P., Crosti, G., Blilou, I., Gratton, E., Benfey, P. N. and Sozzani, R. (2016). Tracking transcription factor mobility and interaction in Arabidopsis roots with fluorescence correlation spectroscopy. *eLife* **5**.

Crickmore, M. A., Ranade, V. and Mann, R. S. (2009). Regulation of Ubx expression by epigenetic enhancer silencing in response to Ubx levels and genetic variation. *PLoS genetics* **5**, e1000633.

Dublanche, Y., Michalodimitrakis, K., Kummerer, N., Foglierini, M. and Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol* **2**, 41.

Dworkin, I., Lee, W., McCloskey, F. and Larsen, E. (2007). Complex genetic interactions govern the temporal effects of Antennapedia on antenna-to-leg transformations in Drosophila melanogaster. *J Genet* **86**, 111-123.

Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* **297**, 1183-1186.

Emerald, B. S. and Cohen, S. M. (2004). Spatial and temporal regulation of the homeotic selector gene Antennapedia is required for the establishment of leg identity in Drosophila. *Dev Biol* **267**, 462-472.

Engstrom, Y., Schneuwly, S. and Gehring, W. J. (1992). Spatial and Temporal Expression of an Antennapedia Lac Z Gene Construct Integrated into the Endogenous Antennapedia Gene of Drosophila-Melanogaster. *Dev Genes Evol* **201**, 65-80.

Franz, K., Singh, A. and Weinberger, L. S. (2011). Lentiviral vectors to study stochastic noise in gene expression. *Methods in enzymology* **497**, 603-622.

Garaulet, D. L., Foronda, D., Calleja, M. and Sanchez-Herrero, E. (2008). Polycomb-dependent Ultrabithorax Hox gene silencing induced by high Ultrabithorax levels in Drosophila. *Development* **135**, 3219-3228.

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* **58**, 35-55.

Gregor, T., Tank, D. W., Wieschaus, E. F. and Bialek, W. (2007a). Probing the limits to positional information. *Cell* **130**, 153-164.

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. and Tank, D. W. (2007b). Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141-152.

Gronlund, A., Lotstedt, P. and Elf, J. (2013). Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat Commun* **4**, 1864.

Halford, S. E. and Marko, J. F. (2004). How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res* **32**, 3040-3052.

Holloway, D. M., Lopes, F. J., da Fontoura Costa, L., Travencolo, B. A., Golyandina, N., Usevich, K. and Spirov, A. V. (2011). Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation. *PLoS computational biology* **7**, e1001069.

Holloway, D. M. and Spirov, A. V. (2015). Mid-embryo patterning and precision in Drosophila segmentation: Kruppel dual regulation of hunchback. *PLoS One* **10**, e0118450.

Ito, K., Awano, W., Suzuki, K., Hiromi, Y. and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124**, 761-771.

Jorgensen, E. M. and Garber, R. L. (1987). Function and misfunction of the two promoters of the Drosophila Antennapedia gene. *Genes Dev* **1**, 544-555.

Kang, M., Piliszek, A., Artus, J. and Hadjantonakis, A. K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development* **140**, 267-279.

Kaur, G., Costa, M. W., Nefzger, C. M., Silva, J., Fierro-Gonzalez, J. C., Polo, J. M., Bell, T. D. and Plachta, N. (2013). Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy. *Nat Commun* **4**, 1637.

Kelso, R. J., Buszczak, M., Quinones, A. T., Castiblanco, C., Mazzalupo, S. and Cooley, L. (2004). Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. *Nucleic Acids Res* **32**, D418-420.

Krasnow, M. A., Saffman, E. E., Kornfeld, K. and Hogness, D. S. (1989). Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. *Cell* **57**, 1031-1043.

Kuziora, M. A. and McGinnis, W. (1988). Autoregulation of a Drosophila homeotic selector gene. *Cell* **55**, 477-485.

Lam, C. S., Mistri, T. K., Foo, Y. H., Sudhaharan, T., Gan, H. T., Rodda, D., Lim, L. H., Chou, C., Robson, P., Wohland, T., et al. (2012). DNA-dependent Oct4-Sox2 interaction and diffusion properties characteristic of the pluripotent cell state revealed by fluorescence spectroscopy. *The Biochemical journal* **448**, 21-33.

Le, T., Liang, Z., Patel, H., Yu, M. H., Sivasubramaniam, G., Slovitt, M., Tanentzapf, G., Mohanty, N., Paul, S. M., Wu, V. M., et al. (2006). A new family of Drosophila balancer chromosomes with a w- dfd-GMR yellow fluorescent protein marker. *Genetics* **174**, 2255-2257.

LeMotte, P. K., Kuroiwa, A., Fessler, L. I. and Gehring, W. J. (1989). The homeotic gene Sex Combs Reduced of Drosophila: gene structure and embryonic expression. *EMBO J* **8**, 219-227.

Little, S. C., Tikhonov, M. and Gregor, T. (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* **154**, 789-800.

Lo, C. A., Kays, I., Emran, F., Lin, T. J., Cvetkovska, V. and Chen, B. E. (2015). Quantification of Protein Levels in Single Living Cells. *Cell reports* **13**, 2634-2644.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A. M., Coppey, M. and Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in Drosophila embryos. *Current biology : CB* **23**, 2135-2139.

Mistri, T. K., Devasia, A. G., Chu, L. T., Ng, W. P., Halbritter, F., Colby, D., Martynoga, B., Tomlinson, S. R., Chambers, I., Robson, P., et al. (2015). Selective influence of Sox2 on POU transcription factor binding in embryonic and neural stem cells. *EMBO reports* **16**, 1177-1191.

Morin, X., Daneman, R., Zavortink, M. and Chia, W. (2001). A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila. *Proc Natl Acad Sci U S A* **98**, 15050-15055.

Muller, C. B., Loman, A., Pacheco, V., Koberling, F., Willbold, D., Richtering, W. and Enderlein, J. (2008). Precise measurement of diffusion by multi-color dual-focus fluorescence correlation spectroscopy. *Epl-Europhys Lett* **83**.

Nagarkar-Jaiswal, S., Lee, P. T., Campbell, M. E., Chen, K., Anguiano-Zarate, S., Gutierrez, M. C., Busby, T., Lin, W. W., He, Y., Schulze, K. L., et al. (2015). A library of MiMICs allows tagging of genes and reversible, spatial and temporal knockdown of proteins in Drosophila. *eLife* **4**.

Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K. and Balazsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc Natl Acad Sci U S A* **106**, 5123-5128.

Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. and Weissman, J. S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846.

Papadopoulos, D. K., Krmpot, A. J., Nikolic, S. N., Krautz, R., Terenius, L., Tomancak, P., Rigler, R., Gehring, W. J. and Vukojevic, V. (2015). Probing the kinetic landscape of Hox transcription factor-DNA binding in live cells by massively parallel Fluorescence Correlation Spectroscopy. *Mechanisms of development* **138 Pt 2**, 218-225.

Papadopoulos, D. K., Resendez-Perez, D., Cardenas-Chavez, D. L., Villanueva-Segura, K., Canales-del-Castillo, R., Felix, D. A., Funfschilling, R. and Gehring, W. J. (2011). Functional synthetic Antennapedia genes and the dual roles of YPWM motif and linker size in transcriptional activation and repression. *Proc Natl Acad Sci U S A* **108**, 11959-11964.

Papadopoulos, D. K., Vukojevic, V., Adachi, Y., Terenius, L., Rigler, R. and Gehring, W. J. (2010). Function and specificity of synthetic Hox transcription factors in vivo. *Proc Natl Acad Sci U S A* **107**, 4087-4092.

Pare, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y. and McGinnis, W. (2009). Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. *Current biology : CB* **19**, 2037-2042.

Perez-Camps, M., Tian, J., Chng, S. C., Sem, K. P., Sudhaharan, T., Teh, C., Wachsmuth, M., Korzh, V., Ahmed, S. and Reversade, B. (2016). Quantitative imaging reveals real-time Pou5f3-Nanog complexes driving dorsoventral mesendoderm patterning in zebrafish. *eLife* **5**.

Plachta, N., Bollenbach, T., Pease, S., Fraser, S. E. and Pantazis, P. (2011). Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat Cell Biol* **13**, 117-123.

Quinones-Coello, A. T., Petrella, L. N., Ayers, K., Melillo, A., Mazzalupo, S., Hudson, A. M., Wang, S., Castiblanco, C., Buszczak, M., Hoskins, R. A., et al. (2007). Exploring strategies for protein trapping in Drosophila. *Genetics* **175**, 1089-1104.

Rao, C. V. and Arkin, A. P. (2003). Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J Chem Phys* **118**, 4999-5010.

Reed, H. C., Hoare, T., Thomsen, S., Weaver, T. A., White, R. A., Akam, M. and Alonso, C. R. (2010). Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics* **184**, 745-758.

Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., Brena, C., Pradel, J. and Graba, Y. (2011). Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. *Proc Natl Acad Sci U S A* **108**, 2276-2281.

Sanchez, A., Garcia, H. G., Jones, D., Phillips, R. and Kondev, J. (2011). Effect of promoter architecture on the cell-to-cell variability in gene expression. *PLoS computational biology* **7**, e1001100.

Sarov, M., Barz, C., Jambor, H., Hein, M. Y., Schmied, C., Suchold, D., Stender, B., Janosch, S., K, J. V., Krishnan, R. T., et al. (2016). A genome-wide resource for the analysis of protein localisation in Drosophila. *eLife* **5**, e12068.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337-342.

Sela, I. and Lukatsky, D. B. (2011). DNA sequence correlations shape nonspecific transcription factor-DNA binding affinity. *Biophys J* **101**, 160-166.

Shimoga, V., White, J. T., Li, Y., Sontag, E. and Bleris, L. (2013). Synthetic mammalian transgene negative autoregulation. *Mol Syst Biol* **9**, 670.

Slattery, M., Voutev, R., Ma, L., Negre, N., White, K. P. and Mann, R. S. (2013). Divergent transcriptional regulatory logic at the intersection of tissue growth and developmental patterning. *PLoS genetics* **9**, e1003753.

Smolander, O. P., Kandhavelu, M., Mannerstrom, H., Lihavainen, E., Kalaichelvan, S., Healy, S., Yli-Harja, O., Karp, M. and Ribeiro, A. S. (2011). Cell-to-cell diversity in protein levels of a gene driven by a tetracycline inducible promoter. *BMC molecular biology* **12**, 21.

Soltani, M., Bokes, P., Fox, Z. and Singh, A. (2015). Nonspecific transcription factor binding can reduce noise in the expression of downstream proteins. *Phys Biol* **12**, 055002.

Stroeher, V. L., Gaiser, J. C. and Garber, R. L. (1988). Alternative RNA splicing that is spatially regulated: generation of transcripts from the Antennapedia gene of Drosophila melanogaster with different protein-coding regions. *Molecular and cellular biology* **8**, 4143-4154.

Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* **99**, 12795-12800.

Szaloki, N., Krieger, J. W., Komaromi, I., Toth, K. and Vamosi, G. (2015). Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling. *Molecular and cellular biology* **35**, 3785-3798.

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X. S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538.

Thattai, M. and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* **98**, 8614-8619.

Tiwari, M., Mikuni, S., Muto, H. and Kinjo, M. (2013). Determination of dissociation constant of the NFkappaB p50/p65 heterodimer using fluorescence cross-correlation spectroscopy in the living cell. *Biochemical and biophysical research communications* **436**, 430-435.

Tsutsumi, M., Muto, H., Myoba, S., Kimoto, M., Kitamura, A., Kamiya, M., Kikukawa, T., Takiya, S., Demura, M., Kawano, K., et al. (2016). In vivo fluorescence correlation spectroscopy analyses of FMBP-1, a silkworm transcription factor. *FEBS open bio* **6**, 106-125.

Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A., et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat Methods* **8**, 737-743.

Vukojevic, V., Heidkamp, M., Ming, Y., Johansson, B., Terenius, L. and Rigler, R. (2008). Quantitative single-molecule imaging by confocal laser scanning microscopy. *Proc Natl Acad Sci U S A* **105**, 18176-18181.

Vukojevic, V., Papadopoulos, D. K., Terenius, L., Gehring, W. J. and Rigler, R. (2010). Quantitative study of synthetic Hox transcription factor-DNA interactions in live cells. *Proc Natl Acad Sci U S A* **107**, 4093-4098.

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J. and Koeppl, H. (2012). Moment-based inference predicts bimodality in transient gene expression. *Proc Natl Acad Sci U S A* **109**, 8340-8345.

Zink, B., Engstrom, Y., Gehring, W. J. and Paro, R. (1991). Direct interaction of the Polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of Drosophila melanogaster. *EMBO J* **10**, 153-162.

Figure Legends

Figure 1: **Concentration, DNA-binding dynamics and cell-to-cell protein concentration variability of 14 Drosophila TFs.** (A-H) Workflow of the study of TFs by FCS (see Materials and Methods and Supplement 1). (I) Representative average FCS measurements of eight TFs. (J) FCS curves shown in (I), normalized to the same amplitude, $G_n(\tau) = 1$ at $\tau = 10\,\mu s$. (K) Variability of the 14 TFs as a function of concentration. (L) Variability in concentration of endogenous Antp in the wing disc. (M) Variability of Antp concentration in clusters of neighboring cell nuclei as a function of its average concentrations. Error bars in (K) and (N) represent 1 standard deviation.

Figure 2: **Antp activates and represses its own transcription.** (A) Schematic representation of the wing disc region of highest Antp expression (green cells). Antp is highly expressed in the wing disc in the regions of the notum that correspond to the structure of the prescutum in the adult cuticle, as well as in the base of the wing blade, which gives rise to the mesopleura and pteropleura of the adult thoracic cuticle. The black rectangle indicates the region of clonal analysis in (B-C'). (B-B') Clonal overexpression of a *SynthAntp-eGFP* construct. Dashed line in (B) shows a clone in the Antp expression domain. (C-C') Transcriptional auto-repression of Antp using the *Antp* P1-*lacZ*. (D) Quantification of repression of Antp protein and reporter inside the repression clones, as compared to the surrounding tissue. (E) Schematic of Antp transcriptional auto-repression. Repression can be direct or indirect. (F) Wing disc region of ectopic Antp P1 reporter expression in (G-G'). (G-G') Ectopic induction of *Antp* P1-*lacZ* in distal compartments of the wing disc by expression of *SynthAntp-eGFP* using *Dll*-Gal4 (MD23). (H) Quantification of auto-activation of Antp reporter within the *Dll*-Gal4 expression domain, as compared to the surrounding tissue. (I) Schematic representation of Antp auto-activation. Scale bars denote $100\ \mu m$.

Figure 3: **Antp switches from transcriptional auto-activation to auto-repression.** (A, F) Clone induction at 26 h (early) or 60 h (late) with analysis at third instar larval stage (~96-120 h of development). Black rectangles represent the corresponding regions of clonal analysis. (B-B') Early clonal induction of full-length, untagged *Antp*, (mCherry positive) reveals strong auto-activation of endogenous Antp-eGFP (dashed lines in (B)). The cyan line outlines the region of highest endogenous Antp expression. The whole *Antp* expression domain expresses *Antp-eGFP*, but overexpression clones (sub-regions marked by absence of mCherry staining) express *Antp-eGFP* much stronger (B'). (C-C') *Antp* P1 transcription in *Antp RNAi* knockdown clones (early clonal induction, dashed line in (C), marked by nuclear mRFP1. (D) Updated Antp auto-activation model, showing strong auto-activation of Antp at early stages. (E) Concentration, DNA-binding and variability studied by FCS at second instar leg and wing discs (FCS analysis in Supplemental Fig. S8). Low concentration, low degree of DNA-binding and high variability are observed in second instar wing and leg discs, but the opposite is true for third instar discs. (G-G') Late-induced clones (dashed lines in

(G), expressing full-length, untagged *Antp* (mCherry positive). Auto-repression of Antp (dashed lines in (G)) is monitored by the endogenous Antp-eGFP protein. The cyan lines in (G) outline the region of strong endogenous expression. (H-H') *Antp* P1 transcription in late *Antp RNAi* knockdown clones (dashed line in (H), 60 h of development) within the Antp normal expression domain, marked by nuclear mRFP1. The cyan line in (H) outlines the region of strong endogenous expression of the P1 reporter. Cytoplasmic eGFP marks the *Antp* knockdown clone (H'). (I) Updated Antp auto-repression model showing the pronounced auto-repressing capacity of Antp at late stages.

Figure 4: **Antp auto-activation and auto-repression relies on Antp isoforms with different binding affinities to chromatin.** (A) Schematic of the Antp mRNA, generated from the P1 promoter. Exons are represented by grey boxes. Magnified exons 4-7 (drawn to scale, omitting splicing points for simplicity) show the alternative splice site (3' of exon 7), resulting in isoforms featuring a short linker between the YPWM motif and the homeodomain (RSQF, grey box), or a long linker isoform (RSQFGKCQ, white box). (B-B') *SynthAntp-eGFP* bearing a long linker expressed by *ptc*-Gal4 and endogenous Antp protein auto-repression were monitored at the proximal portion of the wing disc. A white dashed line outlines the region of auto-repression that was used for quantification (see Materials and Methods). (C-C') Similar to B-B', except that expression was induced by *Dll* (MD23)-Gal4 distally (yellow dashed line). (D) Abundance of long linker isoform (see Materials and Methods); auto-activation and auto-repression efficiencies (Materials and Methods); DNA-bound fractions, measured by FCS (Supplemental Figure S10); and relative affinity of binding to chromatin, calculated by FCS (Supplemental Figure S10) are presented for comparison with (H). (E) Updated model of *Antp* auto-regulation. The activating isoform binds with lower affinity to the P1 Antp promoter, but is produced in excess, relative to the repressing one, resulting in preferential activation of transcription. (F-G') Similar to (B-C') for the short linker isoform. (H) Similar to (D) for comparison. (I) Updated qualitative model representation of Antp repression as in (E), whereby at later stages excess of Antp auto-repressor accounts for negative feedback on transcriptional regulation of the P1 promoter, resulting in partial activation of transcription, hence expression is maintained.

Figure 5: **Concentrations resulting in low variability are required for Antp homeotic function.** (A-D) Transformations of the distal antenna into a tarsus in adult flies, caused by *SynthAntp-eGFP* overexpression in antennal discs (Supplemental Figure S4A-D). Ectopic tarsi range from complete (A) to milder transformations of the arista (B and C) or ectopic leg bristles in the third antennal segment in (C, D) (black arrows). (E-F) Measurements of SynthAntp concentration and cell-to-cell variability of antennal discs (Supplemental Figure S4A-D) in the corresponding antennal discs (A-D). The three Gal4 drivers (blue font) result in partial transformations, despite being expressed at similar levels as the wild type Antp protein in the leg disc. However, their variability is higher than the endogenous variability ($CV^2 = 0.1$). In contrast, Dll-Gal4 (MD23), results in much more robust homeotic transformations (A), accompanied by the lowest variability and closest to the wild type condition. (G-J) A dynamic promoter, which drives transcription of *Antp* followed by a splicing step, leads to either the repressing ("R" in (G)) or activating ("A" in (G)) isoform of Antp. In the absence of Antp, the promoter is inactive and transcription cannot take place ("U" in (G)). This promoter configuration leads to suppression of variability and increase in Antp concentration (J).

Trajectories of individual simulations are presented in (H) and the distribution of the Antp isoforms, predicted by the model, in (I). (K-L') Model predictions (K and L) and experimental data validation (K' and L') of variability (K) and protein Fano factor (L) as a function of Antp concentration.

Figure 6: **Response of *Antp* to genetic perturbations.** (A-B) Overexpression of SynthAntp-eGFP long or short linker isoform result in tarsal transformations of the antenna (A), but normal leg development (B). These flies are fully viable and can be maintained as a stock. (C-C') Antp concentration and variability, measured by FCS, in leg discs of second and third instar larvae upon SynthAntp-eGFP long or short linker isoform expression. Despite persistent high concentration of Antp due to overexpression, variability is reduced. (D-I) Model response upon overexpression of Antp activating or repressing isoforms (similar to Fig. 5H-J). (J) Overexpression of an exogenous repressor (Scr) results in abnormal distal leg development, bearing malformations of the tarsus and femur. (K-N) Similar to (C-I) (see also Supplemental Figure S13E-I'). Antp concentration and variability, measured by FCS in the proximal leg disc of second (early) and third (late) instar larvae upon overexpression of mCherry-SynthScr.