Mechanistic causes of constrained phenotypic variation revealed by synthetic gene regulatory networks

Yolanda Schaerli[1,2,*], Alba Jiménez[3], José M. Duarte[1], Ljiljana Mihajlovic[1,2], Julien Renggli[4], Mark Isalan[5], James Sharpe[3,6], Andreas Wagner[1,7,8,*]

1) University of Zurich, Department of Evolutionary Biology and Environmental Studies, Winterthurerstrasse 190, 8057 Zürich, Switzerland
2) Department of Fundamental Microbiology, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland
3) European Molecular Biology Laboratory-CRG Systems Biology Program, Centre for Genomic Regulation (CRG), and Universitat Pompeu Fabra, 08003 Barcelona, Spain
4) Independent researcher, 1025 St-Sulpice, Switzerland
5) Department of Life Sciences, Imperial College London, London, SW7 2AZ, UK
6) Institucio Catalana de Recerca i Estudis Avancats, 08010 Barcelona, Spain
7) The Swiss Institute of Bioinformatics, Bioinformatics, Quartier Sorge, Batiment Genopode, 1015, Lausanne, Switzerland.
8) The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM, 87501, USA

*Correspondence: yolanda.schaerli@unil.ch; andreas.wagner@ieu.uzh.ch

Abstract

Phenotypic variation is the raw material of adaptive Darwinian evolution. The phenotypic variation brought forth by organismal development is biased towards specific phenotypes, but the molecular mechanisms behind such biases are still poorly understood. Here we study these biases in two synthetic gene regulatory networks expressed in *E. coli* that produce a gene expression stripe, which is a pivotal pattern in embryonic development. The two gene regulatory networks display the same phenotype, but create it through different regulatory mechanisms. Because these regulatory mechanisms are well understood, we can use a combination of experimental measurements, mathematical modelling and DNA sequencing to understand why mutations bring forth only some but not other novel gene expression phenotypes. Our observations emphasize the interactions between genes in a regulatory network as a mechanistic cause of biased phenotype production.

Introduction

The ability of biological systems to bring forth novel and beneficial phenotypes as a consequence of genetic mutations is essential for evolutionary adaptation and innovation. This ability is encapsulated in the concept of evolvability (Kirschner and Gerhart, 1998; Wagner, 2005b). Evolvability can be limited by evolutionary constraints, which are biases or limitations in the production of novel phenotypes (Smith et al., 1985). An example of such constraints comes from laboratory selection experiments with butterfly populations for enhanced wing eyespot colours (Allen et al., 2008). Selection was able to increase the amount of black or gold colouring in the two eyespots simultaneously, but was unable to do so for the two different colours independently in the two eyespots. Constrained variation can have multiple genetic and developmental causes that can be difficult to disentangle in a complex developing organism (Arnold, 1992; Wagner, 2011). Therefore, few experimental demonstrations of evolutionary constraints exist. What is more, thirty years after this concept rose to prominence (Smith et al., 1985), we still do not understand the mechanistic causes of evolutionary constraints.

The instructions for an organism's development are encoded in gene regulatory networks (GRNs) – networks of interacting transcription factors that control gene expression in both time and space (Davidson, 2006). Mutations in the *cis*-regulatory regions of GRNs play an important part in evolutionary adaptation and innovation (Payne and Wagner, 2014; Prud'homme et al., 2007; Wray, 2007). Examples include the evolution of the vertebrate spine (Guerreiro et al., 2013), of wing pigmentation in butterflies (Beldade and Brakefield, 2002) and of hindwing reduction in flies (Carroll et al., 2001). GRNs are thus primary candidates for systems that might lead to the production of constrained variation (Gompel and Carroll, 2003; Sorrells et al., 2015). However, no experimental work exists to find out whether GRNs might constrain novel gene expression phenotypes, and what the mechanistic causes of such constraints might be. These questions require us to study the relationship between genotypic and phenotypic changes in GRNs. Computational models of gene regulation provide one avenue to understand such genotype-phenotype maps (Ciliberti et al., 2007a, b; Cotterell and Sharpe, 2010; Ma et al., 2006; MacCarthy et al., 2003; Payne and Wagner, 2015; Wagner, 2005a). Such models predict that GRNs with different topologies – qualitatively different patterns of interaction between a GRN's genes – can achieve the same gene expression phenotypes, while they differ in their ability to bring forth novel phenotypes through DNA mutations (Ciliberti et al., 2007a, b;

Jimenez et al., 2015; MacCarthy et al., 2003; Payne and Wagner, 2015). However, experimental validation of the latter prediction is still lacking.

To help fill these gaps in experimental evidence, we here use the toolbox of synthetic biology. It allows us to create novel GRNs by assembling well-characterised parts. We are therefore no longer limited to studying GRNs *in situ*, that is, in one or few well-studied organisms where influences of genetic background or environment may be difficult to control. Instead, we can construct and modify GRNs to understand the properties and potential of GRNs to create novel phenotypes (Davies, 2017; Lim et al., 2013; Mukherji and van Oudenaarden, 2009; Wang et al., 2016). We previously built multiple 3-gene synthetic networks that display the same gene expression phenotype, but create this phenotype through different regulatory mechanisms – different regulatory dynamics and regulatory interactions among network genes (Schaerli et al., 2014). Their phenotype is a "stripe" of gene expression (low-high-low) along a spatial axis in response to a chemical concentration gradient that is analogous to a morphogen gradient in development. A GRN's ability to "interpret" such a gradient by producing such stripes is crucial in the development of many organisms and body structures, such as axial patterning of the *Drosophila* embryo and vertebrate neural tube differentiation (Lander, 2007; Rogers and Schier, 2011; Sagner and Briscoe, 2017; Wolpert, 1996). The question of which regulatory mechanisms can produce stripes is therefore itself crucial for developmental genetics (Cotterell and Sharpe, 2010). Here we go beyond this question to ask whether *different* GRNs that have the same phenotype (a "stripe" of gene expression) can produce *different* novel gene expression phenotypes in response to mutations, and if so, why.

Specifically, we use here two synthetic circuits that employ different regulatory mechanisms to produce a striped gene expression pattern. Both of these circuits are hosted by *Escherichia coli* bacteria. When these bacteria are grown as a lawn in the presence of a concentration gradient of the morphogen analogue, they display a spatially striped gene expression pattern (Fig. 1 c). We introduced random mutations into the regulatory regions of these circuits, and analysed the resulting phenotypes. The two circuits indeed produce a different spectrum of novel gene expression phenotypes. That is, the gene expression variation they produce is constrained. To identify the mechanistic causes of these constraints, we combined experimental DNA sequence and phenotypic data with a mathematical model of gene expression dynamics.

Results

Two networks with distinct regulatory mechanisms differ in their mutant phenotype distributions

Fig. 1 shows the topologies (Fig. 1a) and the molecular implementations (Fig. 1b) of our two starting networks, which we had constructed and characterized previously (Schaerli et al., 2014). Briefly, their regulatory input is the sugar arabinose, which serves as a molecular analogue of a developmental morphogen. The arabinose is sensed by the arabinose-responsive promoter pBAD that acts in a concentration dependent-manner. The observable network output is fluorescence, which is produced by superfolder green fluorescent protein (GFP) (Pedelacq et al., 2006). Positive regulatory interactions are encoded by T7 and SP6 phage RNA polymerases (RNAPs), which start transcription at T7 or SP6 promoters, respectively. Negative interactions are encoded by the transcriptional repressors LacI (lactose operon repressor protein) and TetR (tetracycline repressor). They inhibit transcription when bound to their operator sites (LacO, TetO), which are placed downstream of promoters. The two networks employ distinct mechanisms to produce a gene expression stripe pattern (Cotterell and Sharpe, 2010; Jimenez et al., 2015). We call these mechanisms the "*opposing gradients"* and the "*concurring gradients"* mechanisms. They essentially correspond to the well-studied type 2 and type 3 incoherent feedforward motifs (FFM) (Mangan and Alon, 2003) (see Box 1 for explanations). Fig. 1c schematically shows steady-state expression profiles of the three genes (color-coded as in Fig. 1a) under varying arabinose concentrations. Whereas the opposing gradients mechanism is known to be involved in *Drosophila melanogaster* anterior–posterior patterning (hunchback, knirps, krüppel) (Jaeger, 2011), to the best of our knowledge the concurring gradients mechanism has so far not been observed in a natural stripe-forming regulatory network.

Box 1. Two starting circuits producing stripes through two different mechanisms

Opposing gradients mechanism (Incoherent FFM type 2): The "red" gene (with the open reading frames (ORFs) for LacI and TetR encoded on the same transcript) is activated by the "morphogen" arabinose (vertical arrow). Its products thus form a gradient of increasing concentration with increasing arabinose concentration. The "blue" gene (LacI) and the "green" gene (GFP) are expressed from constitutive promoters. However, the "blue" gene is also repressed by the "red" gene product (TetR). Thus, the "blue" gene product forms an *opposing*

gradient with respect to the gradient of the “red” gene product. Both the “blue” (LacI) and “red” (LacI) gene products repress the “green” gene. The GFP thus reaches a high expression only at medium morphogen concentration where the repression from the “red” and “blue” genes are low.

Concurring gradients mechanism (Incoherent FFM type 3): The “red” gene (with the ORFs for SP6 RNA polymerase (RNAP) and LacI encoded on the same transcript) is activated by the “morphogen” arabinose, just as in the previous circuit. Its expression thus also mimics the arabinose gradient. However, in this circuit the “red” gene product SP6 RNAP activates the “blue” gene, which thus forms a *concurring* gradient with respect to the gradient of the “red” gene product. The “green” gene is activated by the “blue” gene (T7 RNAP) and repressed by LacI of the “red” gene. Its maximum expression occurs at medium arabinose concentration where there is already activation from the “blue” gene, but not yet a high level of repression of the “red” gene.

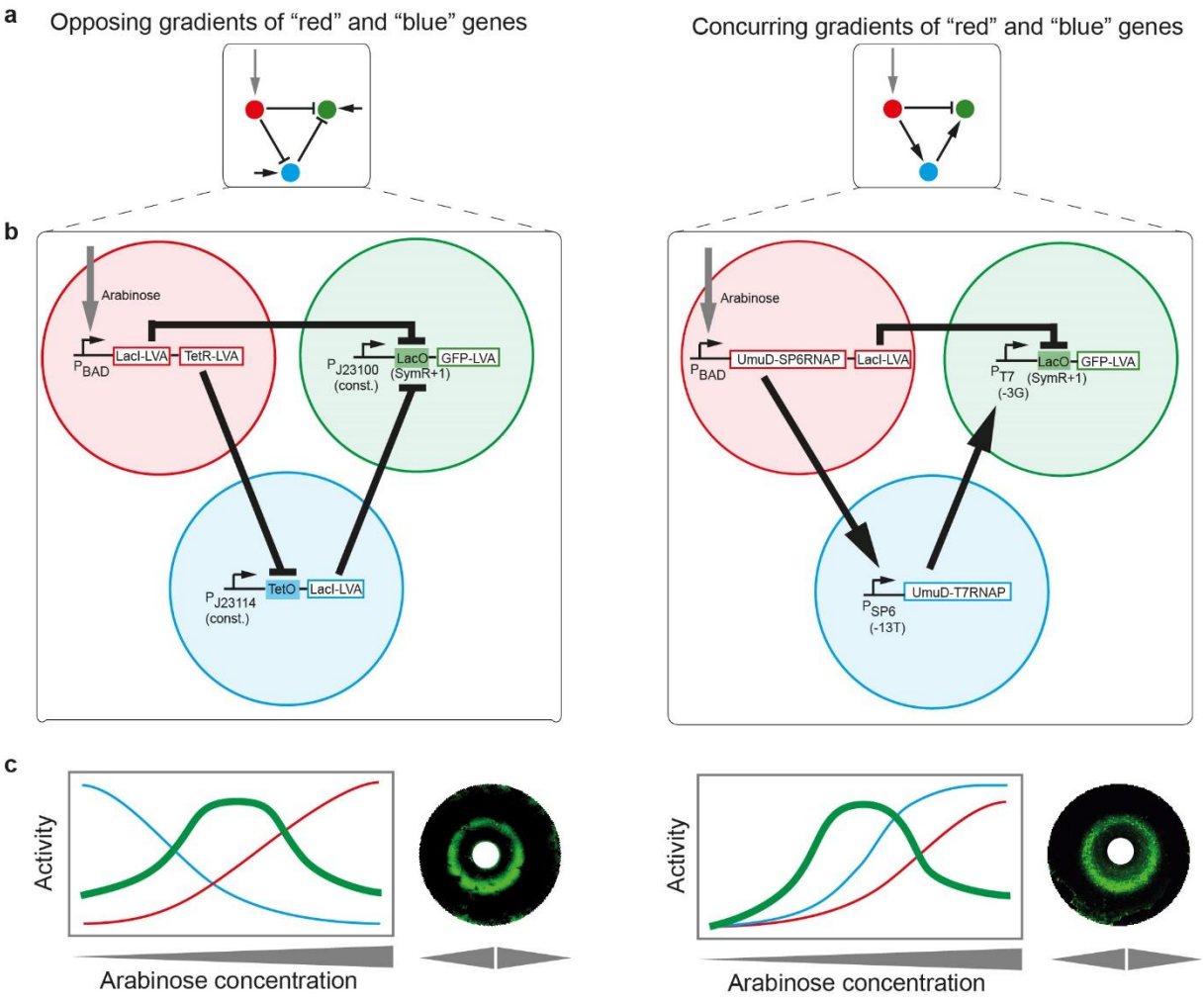

Figure 1: Topologies, synthetic implementations and expression profiles of the networks studied a Topologies of the networks using the *opposing gradients* (left) and *concurring gradients* (right) mechanisms. Arrow: activation; small horizontal arrow: constitutive promoter; bar: repression; red: morphogen input receiver gene; blue: intermediate loop gene; green: stripe output gene. **b** Synthetic implementations of the circuits (Schaerli et al., 2014). Open rectangle: open reading frame; filled rectangle: operator; bent arrow: promoter. All genes carry a degradation tag (LVA (Andersen et al., 1998) or UmuD (Gonzalez et al., 1998)). Indicated variants of T7 promoter, SP6 promoter and LacO were used (Schaerli et al., 2014). J23114 and J23100 are constitutive promoters (Anderson). **c** Rectangles: Steady-state expression profiles of the three genes (colour-coded as in **a**) shown schematically for the two networks (see Box 1). The expression level of the "green" gene is the phenotypic "output" of the network. Circles: Bacterial lawns display green fluorescent rings as a function of arabinose gradients from central paper disks (white). Figure adapted from (Schaerli et al., 2014).

We introduced mutations into the regulatory regions of these two networks by replacing the wild-type regulatory sequence with semi-randomised weighted oligonucleotides (Isalan, 2006). Resulting mutation rates ranged from 2.6 to 3.5 mutations (point mutations, insertions and deletions, SI Table 4) per regulatory region. For each of our two networks, we first generated three libraries of mutant networks in which mutations were restricted to regulatory regions of the "red", "blue" or "green" gene (Fig. 1). After plating cells from a population whose members harboured a synthetic network variant, we randomly picked colonies, grew them in liquid culture, and measured their GFP expression at low (0%), middle (0.0002%) and high (0.2%) arabinose concentrations. We classified the observed fluorescence phenotypes into six categories (Fig. 2a) (see Methods for exact definitions): "stripe", "increase", "decrease", "flat" and "broken" (all expression values below a threshold) and "other" (phenotypes that do not fall in any of the previous categories). We emphasize that phenotypes like these are not peculiarities of our synthetic system. They have also been observed in developing organisms, for example in the anterior–posterior patterning of mutant fly embryos (*Drosophila melanogaster* and *Megaselia abdita*) (see SI discussion "Biological examples of our phenotype categories in fly embryos").

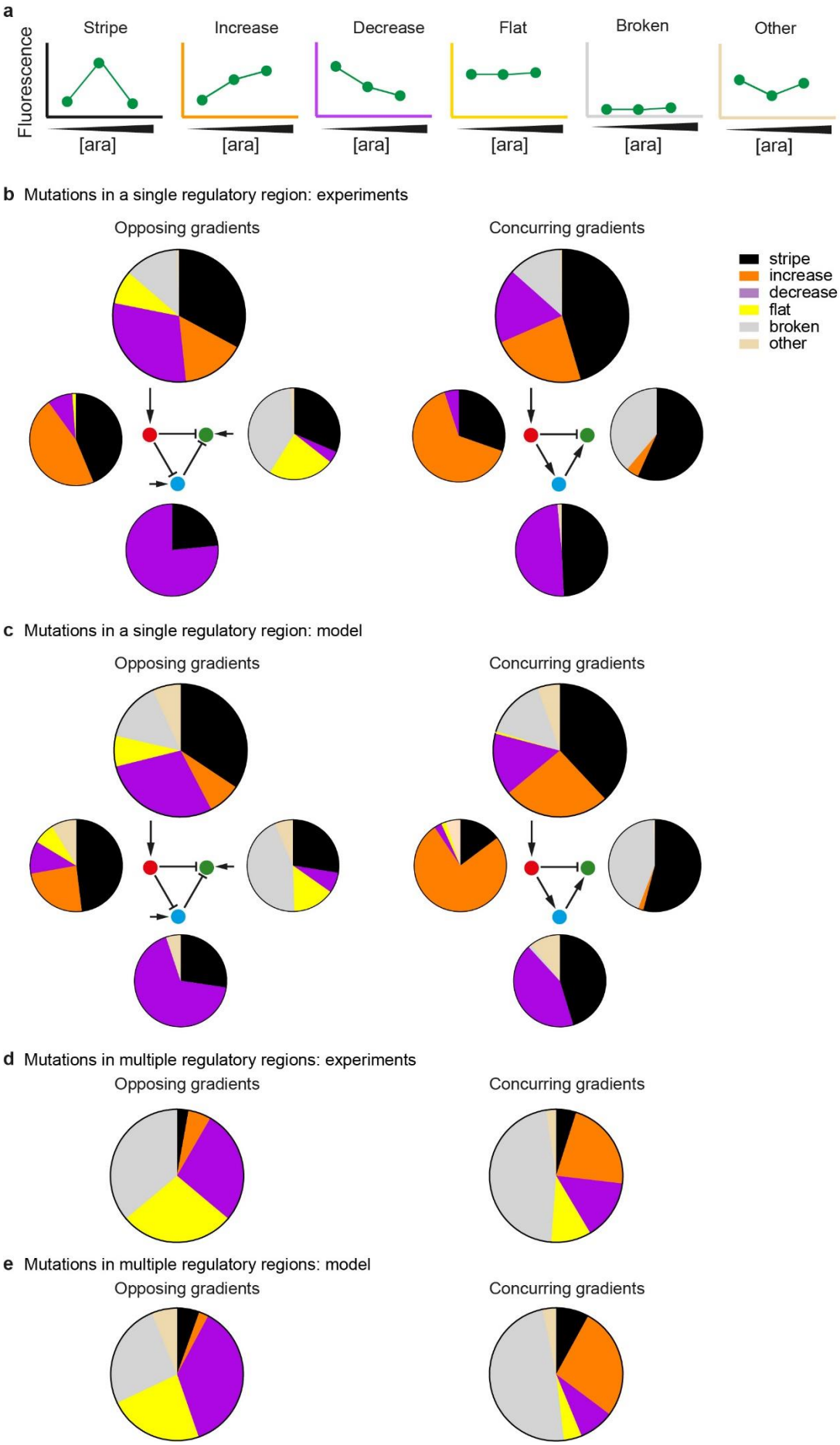
a
Stripe
Increase
Decrease
Flat
Broken
Other
Fluorescence
[ara]
b Mutations in a single regulatory region: experiments
Opposing gradients
Concurring gradients
stripe
increase
decrease
flat
broken
other
c Mutations in a single regulatory region: model
Opposing gradients
Concurring gradients
d Mutations in multiple regulatory regions: experiments
Opposing gradients
Concurring gradients
e Mutations in multiple regulatory regions: model
Opposing gradients
Concurring gradients

Figure 2: Different networks create different spectra of novel phenotypes after mutation. **a** Phenotype categories used in this study. See methods for exact definitions. [ara]: arabinose concentration. The colours of the axes are used throughout the paper to colour-code the phenotypes. **b** Experimentally observed phenotype distributions when mutating one regulatory region at the time. The top (largest) pie chart summarises the spectrum of all mutant phenotypes observed in a network. The smaller pie charts show the spectrum of phenotypes obtained after mutagenising individual genes. The data is based on 234 and 215 mutants of the opposing and concurring gradients networks, respectively. **c** Phenotype distributions produced by the model when mutating one regulatory region at a time (see text for details). **d** Experimentally observed phenotypes of networks with at least two mutated regulatory regions. The data is based on 36 and 41 mutant networks for the opposing and concurring gradients mechanisms, respectively. **e** Phenotype distributions produced by the model when simultaneously mutating multiple regulatory regions.

Fig. 2b summarises the spectrum of phenotypes we observed after mutagenesis. We first note that both networks are to some extent robust to mutations; that is, a considerable fraction of mutations do not change the "stripe" phenotype (black sectors in Fig. 2b). What is more, the two types of networks we study differ in their robustness. Averaged across the three genes, 45.5 % of analysed mutants preserve the stripe phenotype in the concurring gradients network, whereas only 32.9% do so in the opposing gradients network. The concurring gradients network is thus significantly more robust to mutations [Chi-square goodness of fit test, $X2$ (1, N = 215) = 13.67, p =0.0002]. Next, we note that within any one of the two networks the *novel* phenotypes do not occur at the same frequency, providing evidence for the biased production of novel phenotypes, where certain types of phenotypes are more common than others. We also observed differences in the types of novel phenotypes between the two networks. For example, 8.2% of mutants of the opposing gradients networks show a "flat" GFP expression phenotype, where the GFP expression is invariant to arabinose concentrations (yellow sector in Fig. 2b). In contrast, mutations in the concurring gradients network did not produce a single such phenotype. In addition, mutations in the opposing gradients network are more likely to create a "decrease" phenotype (purple, 29.8% of all novel phenotypes) rather than an "increase" phenotype (orange, 15.4%). For the concurring gradients network, the opposite is true: Mutations are more likely to create "increase" (23.0%) rather than "decrease" (18.1%) phenotypes.

Subsequently, we analysed the differences in novel phenotypes created by mutations in specific regulatory regions (i.e. of the "red", "blue" or "green" gene). Within any one of the two network types, regulatory mutations in the "red" gene most often create "increase" phenotypes (Fig. 2b, small pie charts left to the "red" genes), whereas those in the "blue" gene most often create "decrease" phenotypes (Fig. 2b, small pie charts at the bottom of the "blue" genes), and those in the "green" gene preferably create "broken" phenotypes (Fig. 2b, small pie charts to the right of the "green" genes). As a consequence, not all phenotypes can be reached by introducing mutations in the regulatory region of any of the three genes. For example, in the opposing gradient network the "increase" phenotype is only reachable by introducing mutations in the "red" gene, but not in the "blue" and "green" genes.

The two networks differ in the spectrum of novel phenotypes that mutations in individual genes create, which is especially obvious for mutations in the "green" gene: Unless regulatory mutations in this gene lead to a complete loss of expression ("broken"), the opposing gradients network is >5 times more likely to create a "flat" phenotype (23.2 %) than a "decrease" phenotype (4.1 %). In contrast, the concurrent gradients network does not produce any "flat" phenotype at all, but readily produces "increase" phenotypes (4.5 %)**.** In sum, mutations in networks which start with the same phenotype (single stripe formation), but which have alternative topologies and regulatory mechanisms, create different kinds of novel phenotypes. Phenotypic variation is subject to constraints, and these constraints differ between regulatory regions and networks.

Differences in constrained variation can be explained by differences in the regulatory mechanisms behind stripe formation

We next asked whether the regulatory mechanisms contributing to stripe formation can help explain these phenotypic constraints. In doing so, we focused on novel phenotypes produced by regulatory mutations in the "green" gene, because such mutations produced the most distinct spectrum of novel phenotypes (Fig. 2b). Also, the regulation of this gene is most complex, because it receives two regulatory inputs instead of just one for the other genes (Fig. 1). (Similar analyses for the "red" and "blue" genes can be found in SI Figures 2-4.)

To address this question, we first used a mathematical model that we had developed previously and validated experimentally to describe the regulatory dynamics of our networks (Schaerli et al., 2014). Briefly, the model uses Hill-like functions to represent gene regulation changes based on equilibrium binding of transcription factors to their DNA binding sites (Bintu et al., 2005)

(see SI Tables 1 and 2 for details). The unmutated ("wild-type") model for each circuit used parameter values determined in our previous study (Schaerli et al., 2014). Into these models we now introduced quantitative changes in the parameters relating to the promoter activity (binding constants of activators and transcription rates) and to the operator activity (binding constants of repressors) in order to predict phenotypes that are accessible by mutations (see Methods for details). We represent the unmutated network as a point in parameter space, and study regions near this point that are accessible by mutations, and the novel phenotypes they contain (Dichtel-Danjoy and Felix, 2004). We visualise the results in phenotype diagrams (Fig. 3a and SI Fig. 2), which are projections of the higher-dimensional parameter space onto two dimensions. Coloured regions indicate the phenotype (see legend, Fig. 3a) that a circuit produces if mutations change parameters from their wild-type values to other values in this region, expressed as a percentage of the wild-type value. For example, the black region in Fig. 3 corresponds to mutant parameter combinations that maintain the "stripe" phenotype. Its area is therefore a measure for a network's robustness to parameter changes.

For each of the novel phenotypes observed experimentally when mutating the regulatory region of the "green" gene (Fig. 2b), some "mutated" model parameter values exist which reproduce the phenotype (Fig. 3a). However, the phenotype diagrams of the two networks (Fig. 3a) are visually very distinct, indicating that the two networks have different potentials to access specific phenotypes. Specifically, for the opposing gradients network we find regions corresponding to the "broken" (grey), "decrease" (purple), "flat" (yellow) and "other" (beige) phenotypes, whereas for the concurring gradients network we find regions for the "broken" and "increase" (orange) phenotypes. Especially instructive are mutants with strongly decreased repressor binding (i.e. reduced operator activity, arrows in Fig. 3a). Such mutants produce a "flat" phenotype (yellow region) in the opposing gradient network, but an "increase" phenotype (orange region) in the concurring gradient network (Fig. 3a). In other words, even though both networks contain the same operator (LacO) in the "green" gene, the model predicts that identical operator mutations can lead to different novel phenotypes. Fig. 3b illustrates how this is possible: Through an operator mutation that removes the incoming negative interaction of the "green" gene in the opposing gradient network, the constitutive promoter becomes the sole driver of "green" gene expression. Consequently, GFP expression becomes independent of arabinose concentrations, which results in a "flat" phenotype. In contrast, after removing the repression of the "green" gene in the concurring gradient network, the "green" gene is still regulated by the activating "blue" gene (T7 RNAP) in an arabinose dependent-manner. Hence, in this mutant circuit, GFP expression increases with increasing arabinose concentrations. In

sum, different biases in the production of novel phenotypes can be explained by differences in the regulatory mechanisms of the networks.

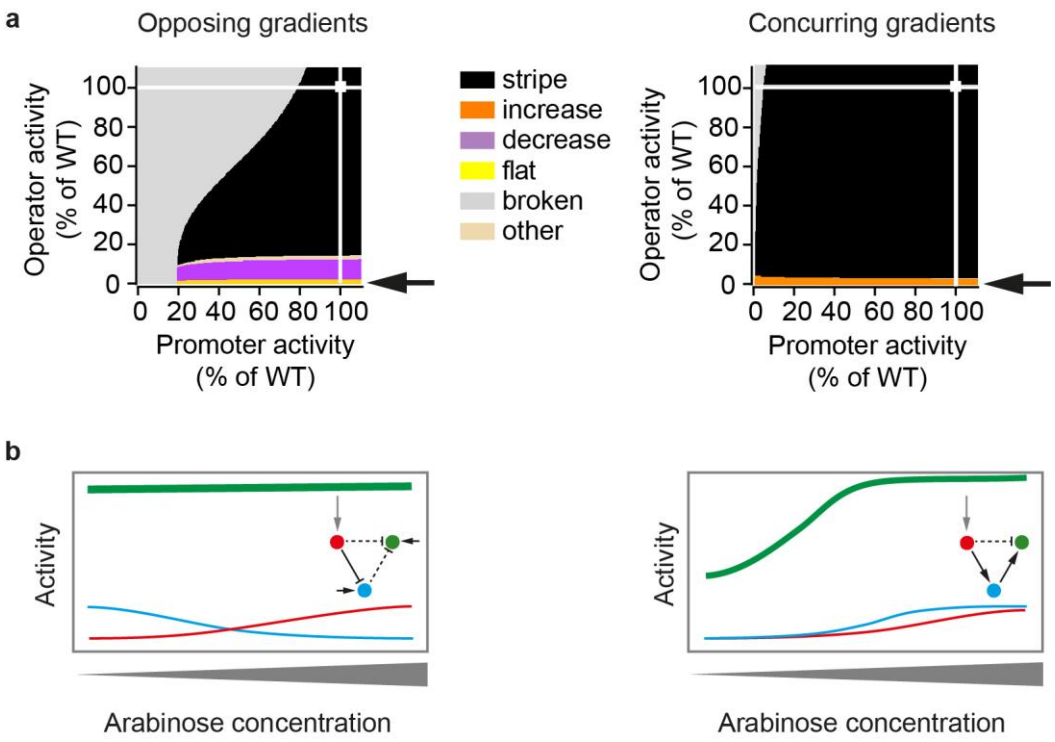

Figure 3: The same parameter change leads to different phenotypes in the two network types. a Phenotype diagrams for parameters that describe the activity of the "green" gene. Horizontal and vertical axes indicate promoter and operator activities of the "green" gene relative to the wild-type value (WT, 100%). Colours indicate phenotypes predicted by the model over the whole range of promoter and operator activity values. White squares indicate the parameter combination of the unmutated circuit, which produces the "stripe" phenotype, and white lines are visual guides that project these values onto the two parameter axes. Arrows point to the phenotype observed when operator activity decreases to a value near zero percent. **b** Schematic drawing of how a strong reduction in operator activity of the "green" gene affects gene expression patterns differently in the two types of networks. Insets: Topologies of the networks, with dashed lines indicating interactions affected by mutations in the operator.

Sequence analysis confirms model predictions of constrained phenotypic variation

We next validated our model predictions with DNA sequence analysis. We sequenced the regulatory regions of the mutated "green" genes in the 73 opposing and 67 concurrent gradient networks, and studied their phenotypes. Because many circuits have multiple regulatory mutations, we first categorised circuits according to the number of mutations that they

contained, counted the number of circuits in each category, and classified their phenotypes (Figure 4a, big diagrams: All mutations). We followed the same procedure for the subset of circuits that have mutations only in the promoter sequence or only in the operator sequence (Fig. 4a, smaller diagrams). This categorisation reveals that mutations in the operator produce a "flat" phenotype in the opposing gradient network, but an "increase" phenotype in the concurring gradient network (Fig. 4a, smaller diagrams), thus validating model predictions (Fig. 3). In addition, the model predicts that mutations in the operator of the opposing gradient network are able to produce a "decrease" phenotype (Fig. 3a). Even though we did not find a circuit with a "decrease" phenotype that has *only* operator mutations, all circuits with this phenotype carry at least one mutation in the operator (and additional mutations in the promoter, SI sequences).

Subsequently, we analysed the locations of mutations in networks with a given phenotype in greater detail (Fig. 4b for mutations in the "green" gene, SI Fig 4 for mutations in the other genes). Many networks carry mutations in both the promoter and the operator. Nevertheless, for the "flat" phenotype of the concurring gradients network, operator mutations are significantly enriched [Chi-square goodness of fit test, $X2$ (1, N = 59) = 7.64, p =0.006], as predicted by the model (Fig. 3). For the "decrease" phenotype in the opposing gradients network and the "increase" phenotype in the concurring gradients network our dataset is too small to detect the predicted enrichment of operator mutations.

Especially informative are those mutants with a novel phenotype that carry only a single point mutation (red arrows in Fig. 4b). Among them are two different mutants of the opposing gradient network with a "flat" phenotype. The mutations in them affect the two central nucleotides of the Lac operator, which are known to be critical for operator function (Falcon and Matthews, 2000; Kalodimos et al., 2001; Lehming et al., 1987; Zhang and Gottlieb, 1995). Mutations in these positions reduce the operator's binding affinity for the LacI repressor dramatically (Lehming et al., 1987), and the observed "flat" phenotype supports our model predictions.

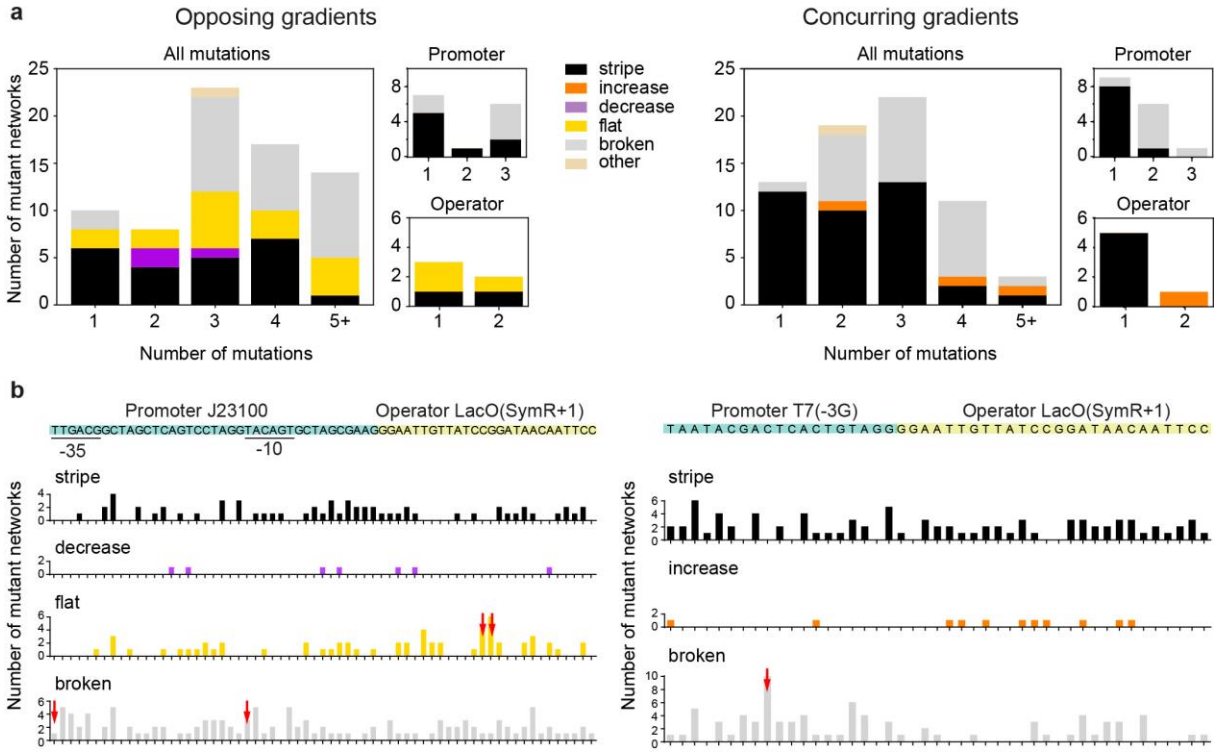

Figure 4: Sequence analysis of the regulatory regions. a Distribution of observed phenotypes of mutant circuits with mutations in the regulatory region of the "green" gene. Phenotypes are colour-coded (legend). Smaller graphs to the right indicate the subset of networks that have mutations only in the promoter or only in the operator. **b** Wild-type sequences of regulatory regions (top of each panel, important elements labelled and coloured) together with the number of mutations at each site of a regulatory region that produce phenotypes of a given kind (bar-charts below sequence, phenotypes labelled and colour-coded). The height of each bar corresponds to the number of mutant networks with a mutation at a given position, where these mutations produced the indicated phenotype. Only phenotypes produced by at least three mutant circuits are shown. Red arrows indicate genotypes that can produce a novel phenotype with a single mutation at the indicated position.

Encouraged by this agreement between model and mutational data, we also aimed to see whether a model of mutational effects can correctly fit the *frequencies* instead of just the *kinds* of phenotypes caused by mutations. To find out, we simulated mutations by changing specific parameters of the model. If a mutation affected the promoter (or operator), we changed all the parameters determining promoter (or operator) activity. We did so by replacing the wildtype promoter (or operator) activity with an activity drawn from a uniform distribution, and aimed to identify upper and lower bounds for this distribution that reproduce the observed frequencies

of novel phenotypes (SI Table 3, see Methods). Such bounds do indeed exist, and enable our model (Fig. 2c) to reproduce experimental phenotype distributions (Fig. 2b) very well.

Non-additive interactions of mutations in multiple regulatory regions are explained by the regulatory mechanisms of the networks

Because mutations rarely occur in isolation in a single gene, we next asked whether mutations in different regulatory regions independently affect gene expression phenotypes. To this end, we pooled networks with mutations in single regulatory regions to obtain networks with mutations in the regulatory regions of two or three genes. We then measured the gene expression phenotypes of these multiple-gene mutants and sequenced their regulatory regions (SI sequences). Fig. 2d shows the resulting distribution of phenotypes. Similar to the one-gene mutants (Fig. 2b), some phenotypes occur more frequently than others, and the two networks produce any one phenotype at different frequencies.

Regardless of the regulatory mechanism, the frequencies of novel phenotypes differed significantly between networks with mutations in multiple versus single genes (Fig. 2d and Fig. 2b, top (largest) pie charts; (opposing: [Chi-square, $X2$ (4, N = 36) = 41.7, p <0.0001]; concurring: [Chi-square, $X2$ (4, N = 41) = 167.0, p <0.0001]). For example, both networks produce the "flat" phenotype in response to multiple mutations, but the concurring gradient networks did not produce this phenotype in response to single-gene mutations (compare yellow sectors in Fig. 2d and 2b, large circles).

Because the phenotypes we observed in the multiple-gene mutants are not just additive superpositions or "sums" of phenotypes observed when the mutations occur separately, the mutations in the different genes must interact non-additively (epistatically) to produce novel phenotypes, such that a mutation's phenotypic effect depends on the genetic background in which it occurs (Lehner, 2011; Mackay, 2014).

In SI Fig. 5 we show an experimental example of how mutations in the "green" and "blue" genes can interact to produce a "flat" phenotype in the opposing gradient network: The network with the mutated "green" gene maintains the "stripe" phenotype (SI Fig. 5a) and the network with the mutated "blue" gene leads to a "decrease" phenotype (SI Fig. 5b). When these two mutations are combined, the resulting phenotype is "flat" (SI Fig. 5c). Importantly, this new phenotype cannot just be explained as an additive superposition of the two individual phenotypes.

To understand the phenotype distributions of the multiple-gene mutants and in particular, the non-additive interactions, we turned again to our model. Analogous to the experiments, we now changed parameters of multiple genes simultaneously within the same ranges as used to model the single-gene mutants. The resulting phenotype distributions (Fig. 2e) predict the experimentally observed distributions very well (Fig. 2d). This implies that both constrained variation and non-additive interactions of mutational effects are a direct consequence of how the individual network genes interact with each other.

Discussion

Mutations in regulatory regions of GRNs play a crucial role in evolutionary adaptation and innovation (Payne and Wagner, 2014; Prud'homme et al., 2007; Wray, 2007). Here, we first introduced random mutations in the regulatory regions of two synthetic stripe-forming GRNs (Fig. 1) and analysed the resulting distributions of novel gene expression phenotypes (Fig. 2). Both networks produced a non-uniform distribution of novel phenotypes, but the different networks displayed different constraints in the production of novel phenotypes. The identity of the mutated regulatory region, and non-additive interactions among mutations in multiple regions also influenced these constraints.

A mathematical model describing the regulatory mechanisms of the two networks allowed us to understand the differences between available phenotypes for the two networks (Fig. 2 and 3) and this understanding was also supported by DNA sequencing data (Fig. 4). We thus provide for the first time empirical evidence that GRNs with different regulatory mechanisms can cause different constrained variation, as was recently proposed (Jimenez et al., 2015). Additional experiments show that the mechanism by which a network produces a stripe constrains the origin of novel expression phenotypes more than quantitative parameters driving gene expression dynamics (SI Fig. 6).

Comparisons of GRNs in related species indicate that they indeed use many distinct molecular solutions to produce a desired phenotypic outcome and that these solutions diverge substantially through time even when the ultimate phenotype stays qualitatively constant (True and Haag, 2001; Weiss and Fullerton, 2000). Examples include the GRN that regulates mating in yeast: Even though both *Saccharomyces cerevisiae* and *Candida albicans* produce two mating types (a-cells and α-cells), the circuit responsible for determining these mating type has changed substantially during evolution (Sorrells et al., 2015; Tsong et al., 2006). Our observations show

that such seemingly equivalent solutions to solve a biological problem can indeed be distinct in how they constrain the outcome of further evolution.

While evolutionary constraints are an important concept in evolutionary theory (Arnold, 1992; Smith et al., 1985), few experimental studies have aimed to detect or quantify them (Allen et al., 2008; Beldade et al., 2002; Bolstad et al., 2015; Frankino et al., 2005; Teotonio and Rose, 2000; Zalts and Yanai, 2017). These studies emphasize the importance of natural selection in determining the outcome of adaptive evolution (Beldade et al., 2002; Frankino et al., 2005), but they also show that evolution of development can be biased by the constrained production of phenotypes (Allen et al., 2008; Bolstad et al., 2015; Kiontke et al., 2007). Laboratory selection experiments have proven to be a powerful tool for detecting evolutionary constraints (Allen et al., 2008; Bolstad et al., 2015; Kiontke et al., 2007), for example in the colour pattern of butterfly wing spots (Allen et al., 2008). However, they have been less successful in disentangling their genetic and developmental causes (Arnold, 1992; Wagner, 2011). This is not surprising, because studying the evolution of developmental GRNs in a multicellular organism is extremely difficult: GRNs are complex, highly interconnected, often incompletely understood, and their genes may be highly pleiotropic, serving multiple functions outside any one GRN (Stearns, 2010). In addition, essential molecular tools for manipulating GRNs are often unavailable, especially in non-model organisms.

As a complementary approach, experimental studies on proteins (Harms and Thornton, 2013; Miller et al., 2006) with simple phenotypes have demonstrated how structure-function relationships of macromolecules can constrain evolution. Here, we apply this approach for the first time to GRNs by studying synthetic circuits in *E. coli* (Fig. 1). We performed experiments with easily modifiable, well characterised synthetic circuits that are not essential for the survival of their host organism. However, the chosen phenotype – stripe formation in a chemical gradient – is crucial for embryonic development of many organisms and body structures (Lander, 2007; Rogers and Schier, 2011; Sagner and Briscoe, 2017; Wolpert, 1996). For example, an opposing gradients network is part of the gap network responsible for axial patterning in the *Drosophila* embryo (Jaeger, 2011). And while a lawn of *E. coli* cells carrying a synthetic GRN does not capture the complexity of a developing animal, this reduced complexity also allowed us to study the potential of GRNs to bias phenotype production without cofounding effects.

Recent developmental and quantitative genetics studies demonstrated that epistatic interactions among genes are of paramount importance in determining phenotypes (Haag, 2007; Lehner,

2011; Mackay, 2014). For example, experiments in *D. melanogaster* have shown that the phenotypic effects of variation in single genes depend heavily on the genetic background, and they do so for phenotypes as varied as bristle shape and number, haltere and eye size, and wing morphology (reviewed in (Lehner, 2011; Mackay, 2014)). Non-linear regulatory mechanisms of gene regulation such as those we study are one possible cause of epistatic interactions (Lehner, 2011). To our knowledge, our study shows for the first time that non-additive interactions in a nonlinear gene regulatory network help produce constrained phenotypic variation. These interactions enabled the origin of novel phenotypes that were not observed when we mutated a single gene (e.g., the "flat" phenotype in concurring gradients network, SI Fig. 5). Our results suggest that these epistatic interactions can also be predicted, if the corresponding GRN, its regulatory mechanism, and the effect of mutations in single regulatory regions are known. Ultimately, understanding the nonlinearities inherent in complex biological systems will be essential to understand how such systems constrain the production of phenotypes.

Since the 19th century, Darwinian evolutionary biology has focused on natural selection and its power to shape populations and species. Natural selection, however, requires phenotypic variation, and the molecular mechanisms by which DNA mutations produce novel phenotypes have only become understood in recent years. While orthodox evolutionary theory assumed, often tacitly, that DNA mutations may produce any kind of variation (Mayr, 1963), the discovery of constrained phenotypic variation challenged this view (Arnold, 1992; Smith et al., 1985). As we show here, constrained variation in simple yet important spatial gene expression patterns can be explained by the simple fact that genes are embedded in regulatory networks. What is more, the regulatory mechanisms of these GRNs can help explain why specific gene expression patterns originate preferentially. Given the pervasive nonlinearity of gene regulatory networks (Davidson, 2006), we surmise that constraints like those we observe are pervasive in biological pattern forming systems. Future work will show whether they can also influence the trajectories of adaptive evolution.

Materials and Methods

Media

Cloning experiments used Luria-Bertani medium (LB:10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl per 1 l) supplemented with appropriate antibiotic (100 μg/ml ampicillin, 30 μg/ml kanamycin or 50 μg/ml spectinomycin). Experiments with the complete synthetic circuits used 'Stripe Medium' (SM: LB plus 0.4% (w/v) glucose, 50 μg/ml ampicillin, 15 μg/ml kanamycin and 25 μg/ml spectinomycin). For the opposing gradients network SM was supplemented with 5 μM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Molecular cloning reagents

Restriction enzymes and T4 DNA ligase were purchased from New England BioLabs (NEB). Oligonucleotides were obtained from Microsynth and chemicals were obtained from Sigma-Aldrich. Polymerase Chain Reactions (PCRs) were carried out with KOD Hot Start polymerase (MERCK MILLIPORE). Plasmids were purified using the QIAprep Spin Miniprep Kit (QIAGEN).

Generation of libraries

Two stripe-forming synthetic circuits of our previous work (Schaerli et al., 2014) which implemented the opposing gradients mechanism (GenBank accession codes of plasmids: KM229377, KM229382, KM229387) and the concurring gradients mechanism (GenBank accession codes of plasmids: KM229378, KM229383, KM229388) were used as starting points ("wild-types", WT) of our experiments.

Mutating one regulatory region of a circuit at a time

Oligonucleotides covering the regulatory regions of these circuits were synthesised with weighted base mixtures (Isalan, 2006). That is, four custom-weighted phosphoramidite mixtures were prepared, with the WT base pair constituting 95% and each of the other bases constituting 1.67% of any one mixture. These mixtures were used to randomise the regulatory regions (Table 1) during oligonucleotide synthesis (Microsynth). These semi-randomised weighted single-stranded oligonucleotides (2 μM) were annealed to a reverse primer (Table 2, 2.4 μM) to render them double-stranded by primer extension (2 min 95º C, cooling down to 72º C in 7 min, 10 min 72º C). The resulting library of double-stranded oligonucleotides was then purified with the QIAquick nucleotide removal kit (QIAGEN). Next, these double-stranded

oligonucleotides encoding the mutated regulatory regions were cloned into the plasmid encoding the gene whose regulatory region was to be mutated. For the "blue" and "green" regulatory regions this was done by restriction enzyme digest and ligation. For the "red" regulatory region Gibson assembly (Gibson et al., 2009) was performed instead.

The cloning of the "blue" and "green" regulatory regions by restriction enzyme digest and ligation was performed as follows: double-stranded oligonucleotides encoding one of the regions were digested with EcoRI and SacI. The plasmid into which the region was to be inserted was also digested with EcoRI and SacI, dephosphorylated with CIP, and gel purified (QIAquick gel extraction kit, QIAGEN). The double-stranded digested oligonucleotide library (25 ng) was then ligated into the cut plasmid (70 ng).

The cloning of the "red" regulatory regions by Gibson assembly was performed as follows: The plasmid into which the region was to be inserted was first amplified using the primers pBAD_Gibson_f and pBAD_Gibson_r (Table 2), and was then assembled with the double-stranded oligonucleotide library using the Gibson assembly master mix (NEB). Due to a mistake in primer design, Gibson assembly produced a two-nucleotide deletion (CT) downstream of the pBAD promoter. Therefore, the library contains sequences with and without this deletion. We confirmed that this change did not affect the initial "stripe" phenotype.

Ligation ("blue" and "green" regulatory regions) and Gibson assembly reaction products ("red" regulatory region) were transformed into electrocompetent MK01 cells (Kogenaru and Tans, 2014) that carried already the two other plasmids necessary to complete the synthetic circuit. Transformants were plated out on SM-agar plates.

Mutating multiple regulatory regions of a circuit

For experiments where multiple regulatory regions of a circuit were to be mutated at the same time, cloning was performed as described above for mutating one regulatory region at a time. However, instead of transforming plasmids with mutagenised regulatory regions directly into MK01 cells, ligation products and Gibson assembly reaction products were first transformed into electrocompetent NEBα cells, and plated out on LB-agar plates containing the appropriate antibiotic (100 μg/ml ampicillin, 30 μg/ml kanamycin or 50 μg/ml spectinomycin). All colonies were resuspended in LB, diluted 100-fold into LB containing the appropriate antibiotic, and grown overnight. Plasmids libraries were extracted from the resulting culture (QIAprep Spin

Miniprep Kit, QIAGEN). The extracted plasmid libraries were mixed with plasmids containing no mutations in a ratio of 70:30 to generate mutant circuits that have mutations in two or three genes. The resulting plasmid mix was transformed into electrocompetent MK01 cells (Kogenaru and Tans, 2014). Circuits that had no mutations or only mutations in one gene were not considered in the analysis.

Network + gene	**Weighted olignonucleotide for randomisation (5' -> 3')**
OG[1] red	CGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGTTCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCGAATTCATTTCACC
OG blue	ATTGGAATTCTTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGCGAAGGGTCCCTATCAGTGATAGAGAGAGCTCGTTGAGTTACCTGC
OG green	ATTGGAATTCTTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCGAAGGGAATTGTTATCCGGATAACAATTCCGAGCTCGTTGAGTTACCTGC
CG[2] red	CGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGTTCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCGAATTCATTTCACC
CG blue	ATTGGAATTCATTTTGGTGACACTATAGAAGGGGCCAAGCAGGGGGCCAAGCAGGGGGGCCAAGGAGCTCGTTGAGTTACCTGC
CG green	ATTGGAATTCTAATACGACTCACTGTAGGGGAATTGTTATCCGGATAACAATTCCGAGCTCGTTGAGTTACCTGC
CG green mutant A	ATTGGAATTCTAATACGACTCACTGTAGGGGAATTGTTAACCGGATAACAATTCCGAGCTCGTTGAGTTACCTGC
CG green mutant B	ATTGGAATTCTTAATACGACTCACTGTAGGTGAATTGTTATCCGGATAACATTCCGAGCTCGTTGAGTTACCTGC

Table 1: Sequences of oligonucleotides synthesized with weighted base mixtures. The underlined region was randomised, the remaining sequence was constant. [1]OG: opposing gradient network; [2]CG: concurring gradient network

name	**sequence (5' -> 3')**
rev_green/blue	GCAGGTAACTCAACGAGCTC
rev_red	GGTGAAATGAATTCGGTATGGA
pBAD_Gibson_f	TCCATACCGAATTCATTTCACC
pBAD_Gibson_r	AGTGTGACGCCGTGCAAATAATC

Table 2: Sequences of primers used for cloning

name	sequence (5' -> 3')	used to sequence
pET_Seq	CCGAAAAGTGCCACCTGAC	OG green
pET_Seq_Amp	GACACGGAAATGTTGAATACTCATAC	CG green, OG green
pBAD_f	GCCGTCACTGCGTCTTTTAC	OG red, CG red
pCDF_Seq_ori	GAGTTCGCAGAGGATTTGTTTAGC	OG blue, CG blue
pCDF_I2_rev	TCTACTGAACCGCTCTAG	OG blue

Table 3: Sequences of primers used for sequencing

Fluorescence measurements of mutagenised circuits

Colonies were picked from agar plates, inoculated into SM medium in a single well of a 96-well plate and grown overnight. Each plate also contained three clones of the WT circuit and a "blank" (SM medium only). A glycerol stock plate was prepared from the overnight cultures. This plate was used to inoculate three further 96-well overnight pre-culture plates with SM medium. Five µl of each well from the 96-well plate were transferred to four wells of a 384-well plate containing 55 µl of SM medium, arabinose, and IPTG. Specifically, the four wells contained the following amounts of arabinose and IPTG:

1) 0% arabinose ("low")
2) 0.0002% arabinose ("medium")
3) 0.2% arabinose ("high")
4) 0.2% arabinose with 700 µM IPTG for the opposing gradients network and 0.2% arabinose with 100 µM IPTG for the concurring gradients network ("metabolic load control", see section "*Phenotype classification, metabolic load"* for details).

The pipetting steps for this part of the experiment were carried out with a manual pipetting system (Rainin Liquidator 96, METTLER TOLEDO).

The 384-well plate was incubated at 37 °C in a Tecan plate reader (Infinite F200 Pro or SPARK 10M) until the *E. coli* cells had reached stationary phase (approximately 5 h). During this incubation absorbance at 600 nm and green fluorescence (excitation: 485 nm, emission: 520 nm) were measured every 10 min. Between readings, plates were shaken continually (orbital, 2 mm). Plates were incubated and read with their lids in place to reduce evaporation. Absorbance and green fluorescence was measured for each colony in three independent experiments, each started from a separate pre-culture plate.

Analysis

The time-point when the fluorescence of the WT network at the medium arabinose concentration (0.0002%) peaked was chosen for further analysis of all fluorescence measurements. The background fluorescence of the SM medium was subtracted from each culture's fluorescence. Likewise, the background absorbance was subtracted from each culture's absorbance. Background-corrected fluorescence was then normalised for the number of cells by dividing it by the background-corrected absorbance. This background corrected normalised fluorescence (nF) was used for all further analyses.

Exclusions

A circuit was excluded from further analysis if

(i) any of its nF values (except the one at 0% arabinose) was smaller than zero

or

(ii) the absorbance of the circuit differed by more than 0.1 from the absorbance of the WT controls in any of the four conditions (which indicates a substantially different growth rate of the circuit)

or

(iii) it suffered from metabolic load (see below)

Metabolic load: The 4th condition (highest arabinose concentration with IPTG) served as a metabolic load control. We previously noted that expressing the genes of synthetic networks at high levels can induce a strong bacterial stress response that affects the expression of genes and the growth rate of the cell (Schaerli et al., 2014). This can lead to a spurious "stripe" phenotype caused by a stress-induced GFP expression that is shut down at the highest arabinose concentration (Schaerli et al., 2014). We therefore checked for each circuit whether it suffers from this metabolic load problem. To this end, we removed LacI repression at the highest of our three arabinose conditions through addition of IPTG. Without repression, we are no longer expecting to observe a "stripe" phenotype. If the observed phenotype is nevertheless a "stripe", this is a strong indication that the network suffers from metabolic load. Specifically, a circuit was excluded due to high metabolic load if its nF value in the 4th condition was lower than 90% of its nF value at the medium arabinose concentration, and if the nF value at the medium arabinose concentration was higher than the corresponding nF of the WT controls (we know that expression levels as high as that of the WT do not induce metabolic load (Schaerli et al., 2014)).

Phenotypic categories

All mutant circuits that remained after the filtering procedure just described were classified into the following phenotypic categories (in this order):

Broken: A threshold value of nF below which a phenotype was considered "broken" was defined as follows: For the opposing gradients networks, nF needed to lie below the nF of the WT controls at the highest of our three arabinose concentration. For the concurring gradients networks, nF needed to lie below 1/3 of the nF of the WT controls at this highest arabinose concentration. Any circuit whose nF was below this threshold in all four conditions was assigned the "broken" phenotype. The reason for the different definitions of the threshold for the two networks is that the WT concurring gradient network circuit has a much higher level of basal fluorescence.

Flat: The average nF at the lowest, medium and highest arabinose concentration was calculated. If all three nFs differed by less than 10% from this average, the circuit was assigned the "flat" phenotype.

Decrease: If the following three statements were true, the mutant circuit was assigned the "decrease" phenotype:

1) The nF at the lowest arabinose concentration was higher than 90% of the nF at the medium arabinose concentration.
2) The nF at the medium arabinose concentration was higher than 90% of the nF at the highest arabinose concentration.
3) The nF at the lowest arabinose concentration was higher than 120% of the nF at the highest arabinose concentration.

Increase: If the following three statements were true, the mutant was assigned the "increase" phenotype:

1) The nF at the highest arabinose concentration was higher than 90% of the nF at the medium arabinose concentration.
2) The nF at the medium arabinose concentration was higher than 90% of the nF at the lowest arabinose concentration.
3) The nF at the highest arabinose concentration was higher than 120% of the nF at the lowest arabinose concentration.

Stripe: If the nF at medium arabinose concentration was higher than 120% of the nF at the lowest and highest arabinose concentrations, the mutant was assigned the "stripe" phenotype.

Other: Any phenotype that did not fall into one of the previous categories.

Sequencing

The mutagenised region(s) of all circuits whose phenotype fell into one of our six main categories, and did so consistently in three independent measurements were sent for Sanger sequencing (High-throughput service, Microsynth, see Table 3 for primers).

Mutagenised circuits with WT regulatory sequences despite mutagenesis, with polymorphic nucleotides in the sequences (mainly due to transformation of multiple plasmid variants into the same cell) or with cloning artefacts (shortened or multiple regulatory regions) were discarded. The remaining mutants were used for the phenotypic statistics we report, and their sequences were analysed further. Pairwise alignment to the WT sequence was performed with the Biopython (Cock et al., 2009) Bio.pairwise2 module. A custom-made Python script was used to categorise mutations as point mutations, insertions, or deletions, and to identify their positions.

All sequences are listed in the SI.

Experimental confirmation of rare observed phenotypes

Phenotypes observed fewer than three times in a library (except "others") were experimentally confirmed. The fluorescence output of these circuits was measured in a 96-well plate assay as described previously (Schaerli et al., 2014) at the arabinose concentrations also used in the 386-well plate assay. If the phenotypes of the two assays did not agree, the circuits were excluded from the data set.

Experimental confirmation of epistasis (SI Fig. 5)

The plasmids from mutant 4_4_f (see SI sequences) were isolated. For SI Fig. 5a the mutated plasmid coding for the "green" gene was transformed together with the WT plasmids for the "blue" and "red" genes into electrocompetent MK01 cells (Kogenaru and Tans, 2014). For SI Fig. 5b the mutated plasmid coding for the "blue" gene was transformed together with the WT plasmids for the "green" and "red" genes into electrocompetent MK01 cells (Kogenaru and

Tans, 2014). For SI Fig. 5c the initial 4_4_f mutant was assayed (the "red" gene is not mutated). The fluorescent phenotypes were measured in a 96-well plate assay as described (Schaerli et al., 2014) at the following arabinose concentrations (w/v): 0.2%, 0.02%, 0.002%, 0.0002%, 0.00002%, 0%.

Statistical tests

Chi-square goodness of fit tests (Snedecor and Cochran, 1989) were used to compare observed and expected frequencies. An online tool (https://graphpad.com/quickcalcs/chisquared1.cfm) was used to perform the calculations.

Phenotype diagrams

A previously developed and experimentally validated model was used to describe the regulatory dynamics of our networks (see SI Tables 1 and 2 for details) (Schaerli et al., 2014). To generate phenotype diagrams a custom-made Python script was written that systematically varies the combinations of two or more parameters between 0-110% (for the "blue" and "green" genes) or 0-200% (for the "red" genes) of the wild-type parameter value in 1000 steps. Analogously to our experiments for each parameter combination the model's phenotype was evaluated at 3 arabinose concentrations (0%, 0.000195%, 0.2% for opposing gradients network and 0%, 0.000195%, 0.1% for concurring gradients network). In order to allocate the obtained GFP expression pattern to a phenotype category the same rules as described above for the experimental data ("Phenotypic categories") were used. An R script (R Development Core Team, 2016) was applied to create a digital image where every pixel corresponds to a combination of parameter values and has a colour corresponding to the model's phenotype. For SI Fig. 2, combinations of two parameters as indicated on the axes were varied. For Fig. 3a multiple parameters affecting the promoter or operator were varied jointly and to the same extent. For the opposing gradients network, these were parameters *a* and *b* for promoter activity, and parameter *c* for operator activity. For the concurring gradients network these were parameters *b*, *c* and *e* for promoter activity and parameter *d* for operator activity.

Distributions of novel phenotypes

In order to fit quantitatively the distributions of novel phenotypes for the single-gene mutants and predict the distributions of novel phenotypes for the multiple-gene mutants a custom-made Python scripts was used. We first discuss the single-gene mutants:

Multiple iterations of a procedure were performed that consisted of the following three steps: First, a series of simulated mutants was created. For each gene, a random binary vector whose length corresponds to the number of nucleotides in the regulatory sequence (SI Table 4) was generated. In this binary vector 0's and 1's indicate whether a nucleotide is not mutated (0) or mutated (1), and the probability to obtain either 0 or 1 is given by the average mutation rate extracted from the experimental sequencing data (SI Table 4). For every network, it was then assessed which genes were mutated, according to the entries of this vector. For the single-gene mutants, mutants that had only one gene mutated were selected. This process was repeated until 1000 single-gene mutants had been obtained. For a particular mutant, all parameters related to the mutated sequence were varied (see SI Tables 1-3). A given single-gene mutant had either only its promoter mutated, only its operator mutated, or both. If a mutation affected the promoter (operator), all parameters determining promoter (operator) activity were changed (SI Table 3). A subset of the changed parameters was varied jointly and to the same extent (i.e. all of them were changed to same percentage of their wild-type parameter value), because a mutation is likely to affect these parameters in a similar way (SI Table 3). New (mutant) parameters were chosen according to a standard uniform distribution between upper and lower ranges which were kept constant for a given model iteration.

Second, for each mutant the phenotype of the model (SI Tables 1 and 2) was evaluated at 3 arabinose concentrations (0%, 0.000195%, 0.2% for opposing gradients network and 0%, 0.000195%, 0.1% for concurring gradients network). In order to allocate the resulting GFP expression pattern to a phenotype category the same rules as described above for the experimental data ("Phenotypic categories") were used.

Third, the obtained phenotype distribution of the 1000 assessed mutants was compared to the experimentally observed phenotype distribution.

After each iteration of these three steps, the upper and lower ranges of each parameter were manually adjusted to best fit the results of the model to the phenotype distributions observed in the experimental data (see SI "Discussion of intervals"). Finally, the best upper and lower ranges for each mutant parameter distribution were kept (SI Table 3) and used to produce Fig. 2c.

To predict the distributions of novel phenotypes for the multiple-gene mutants (Fig. 2e) the same procedure was used, with the following modifications: 1) only mutants containing more than one mutated gene were kept for analysis. 2) The upper and lower ranges of the parameter distributions were not adjusted, but the intervals derived from the single-gene mutants were used (SI Table 3). 3) Only one iteration of the three steps above was performed.

Schematic drawings

Fig. 3b contains schematic depictions, based on the mathematical model. The model uses the parameters as in SI tables 1-2 with following changes:

Fig. 3b: opposing gradient (left): Parameter *c* of the "green" gene was changed from 0.103 to 0; concurring gradient (right): Parameter *c* of the "green" gene was changed from 16.5 to 0.

Acknowledgments

We thank Elke Karaus Wyer for carrying out preliminary experiments, Karl Wotton for advice on fly embryo phenotypes and Joshua L. Payne for critical reading. We thank the Sanger Sequencing team from Microsynth for their competent support.

Funding

YS, JMD and LM acknowledge support by Swiss National Science Foundation Ambizione program (PZ00P3-148235 to YS); AW acknowledges support by Swiss National Science Foundation grant 31003A_146137, by ERC Advanced Grant 739874, by an EpiphysX RTD grant from SystemsX.ch, as well as by the University Priority Research Program in Evolutionary Biology at the University of Zurich. MI is funded by a Wellcome Trust UK New Investigator Award (WT102944) and by the Volkswagen Foundation. AJ and JS acknowledge the Spanish Ministry of Economy, Industry and Competitiveness (MINECO), BFU2010-16428, the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 670555, the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement 601062; the Spanish Ministry of Economy, Industry and Competitiveness "Centro de Excelencia Severo Ochoa 2013–2017", SEV-2012-0208 and the Cerca Programme/Generalitat de Catalunya.

Competing interests

The authors declare not competing interests.

Author contributions

YS, AW designed the project, YS, JMD, LM performed experiments, YS, AJ, AW analysed data, AJ, YS, JR did the mathematical modelling, JS supervised modelling, YS, AJ, MI, JS, AW wrote the paper.

References

Allen, C.E., Beldade, P., Zwaan, B.J., and Brakefield, P.M. (2008). Differences in the selection response of serially repeated color pattern characters: standing variation, development, and evolution. BMC Evol Biol *8*, 94.

Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M., and Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. Appl Environ Microbiol *64*, 2240-2246.

Anderson, J.C. Anderson promoter collection.

Arnold, S.J. (1992). Constraints on phenotypic evolution. The American naturalist *140 Suppl 1*, S85-107.

Beldade, P., and Brakefield, P.M. (2002). The genetics and evo-devo of butterfly wing patterns. Nat Rev Genet *3*, 442-452.

Beldade, P., Koops, K., and Brakefield, P.M. (2002). Developmental constraints versus flexibility in morphological evolution. Nature *416*, 844-847.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005). Transcriptional regulation by the numbers: applications. Curr Opin Genet Dev *15*, 125-135.

Bolstad, G.H., Cassara, J.A., Marquez, E., Hansen, T.F., van der Linde, K., Houle, D., and Pelabon, C. (2015). Complex constraints on allometry revealed by artificial selection on the wing of Drosophila melanogaster. Proc Natl Acad Sci U S A *112*, 13284-13289.

Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2001). From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design (Malden, MA: Blackwell Science).

Ciliberti, S., Martin, O.C., and Wagner, A. (2007a). Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci U S A *104*, 13591-13596.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007b). Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput Biol *3*, e15.

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422-1423.

Cotterell, J., and Sharpe, J. (2010). An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol Syst Biol *6*, 425.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Academic).

Davies, J. (2017). Using synthetic biology to explore principles of development. Development *144*, 1146-1158.

Dichtel-Danjoy, M.L., and Felix, M.A. (2004). Phenotypic neighborhood and micro-evolvability. Trends Genet *20*, 268-276.

Falcon, C.M., and Matthews, K.S. (2000). Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. Biochemistry *39*, 11074-11083.

Frankino, W.A., Zwaan, B.J., Stern, D.L., and Brakefield, P.M. (2005). Natural selection and developmental constraints in the evolution of allometries. Science *307*, 718-720.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Gompel, N., and Carroll, S.B. (2003). Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. Nature *424*, 931-935.

Gonzalez, M., Frank, E.G., Levine, A.S., and Woodgate, R. (1998). Lon-mediated proteolysis of the Escherichia coli UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. Genes Dev *12*, 3889-3899.

Guerreiro, I., Nunes, A., Woltering, J.M., Casaca, A., Novoa, A., Vinagre, T., Hunter, M.E., Duboule, D., and Mallo, M. (2013). Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. Proc Natl Acad Sci U S A *110*, 10682-10686.

Haag, E.S. (2007). Compensatory vs. pseudocompensatory evolution in molecular and developmental interactions. Genetica *129*, 45-55.

Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat Rev Genet *14*, 559-571.

Isalan, M. (2006). Construction of semi-randomized gene libraries with weighted oligonucleotide synthesis and PCR. Nat Protoc *1*, 468-475.

Jaeger, J. (2011). The gap gene network. Cell Mol Life Sci *68*, 243-274.
Jimenez, A., Munteanu, A., and Sharpe, J. (2015). Dynamics of gene circuits shapes evolvability. Proc Natl Acad Sci U S A *112*, 2103-2108.
Kalodimos, C.G., Folkers, G.E., Boelens, R., and Kaptein, R. (2001). Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. Proc Natl Acad Sci U S A *98*, 6039-6044.
Kiontke, K., Barriere, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D.H., and Felix, M.A. (2007). Trends, stasis, and drift in the evolution of nematode vulva development. Curr Biol *17*, 1925-1937.
Kirschner, M., and Gerhart, J. (1998). Evolvability. Proceedings of the National Academy of Sciences *95*, 8420-8427.
Kogenaru, M., and Tans, S.J. (2014). An improved Escherichia coli strain to host gene regulatory networks involving both the AraC and LacI inducible transcription factors. J Biol Eng *8*, 2.
Lander, A.D. (2007). Morpheus unbound: reimagining the morphogen gradient. Cell *128*, 245-256.
Lehming, N., Sartorius, J., Niemoller, M., Genenger, G., v Wilcken-Bergmann, B., and Muller-Hill, B. (1987). The interaction of the recognition helix of lac repressor with lac operator. EMBO J *6*, 3145-3153.
Lehner, B. (2011). Molecular mechanisms of epistasis within and between genes. Trends Genet *27*, 323-331.
Lim, W.A., Lee, C.M., and Tang, C. (2013). Design principles of regulatory networks: searching for the molecular algorithms of the cell. Mol Cell *49*, 202-212.
Ma, W., Lai, L., Ouyang, Q., and Tang, C. (2006). Robustness and modular design of the Drosophila segment polarity network. Mol Syst Biol *2*, 70.
MacCarthy, T., Seymour, R., and Pomiankowski, A. (2003). The evolutionary potential of the Drosophila sex determination gene network. J Theor Biol *225*, 461-468.
Mackay, T.F. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet *15*, 22-33.
Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A *100*, 11980-11985.
Mayr, E. (1963). Animal Species and Evolution (Cambridge Belknap/Harvard Univ. Press).
Miller, S.P., Lunzer, M., and Dean, A.M. (2006). Direct demonstration of an adaptive constraint. Science *314*, 458-461.

Mukherji, S., and van Oudenaarden, A. (2009). Synthetic biology: understanding biological design from synthetic circuits. Nat Rev Genet *10*, 859-871.

Payne, J.L., and Wagner, A. (2014). The robustness and evolvability of transcription factor binding sites. Science *343*, 875-877.

Payne, J.L., and Wagner, A. (2015). Function does not follow form in gene regulatory circuits. Sci Rep *5*, 13015.

Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol *24*, 79-88.

Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A *104 Suppl 1*, 8605-8612.

R Development Core Team (2016). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Rogers, K.W., and Schier, A.F. (2011). Morphogen gradients: from generation to interpretation. Annu Rev Cell Dev Biol *27*, 377-407.

Sagner, A., and Briscoe, J. (2017). Morphogen interpretation: concentration, time, competence, and signaling dynamics. Wiley Interdiscip Rev Dev Biol.

Schaerli, Y., Munteanu, A., Gili, M., Cotterell, J., Sharpe, J., and Isalan, M. (2014). A unified design space of synthetic stripe-forming networks. Nat Commun *5*, 4905.

Smith, J.M., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. The Quarterly Review of Biology *60*, 265-287.

Snedecor, G.W., and Cochran, W.G. (1989). Statistical Methods, Eighth Edition edn (Iowa State University Press).

Sorrells, T.R., Booth, L.N., Tuch, B.B., and Johnson, A.D. (2015). Intersecting transcription networks constrain gene regulatory evolution. Nature *523*, 361-365.

Stearns, F.W. (2010). One hundred years of pleiotropy: a retrospective. Genetics *186*, 767-773.

Teotonio, H., and Rose, M.R. (2000). Variation in the reversibility of evolution. Nature *408*, 463-466.

True, J.R., and Haag, E.S. (2001). Developmental system drift and flexibility in evolutionary trajectories. Evolution & development *3*, 109-119.

Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415-420.

Wagner, A. (2005a). Circuit topology and the evolution of robustness in two-gene circadian oscillators. Proc Natl Acad Sci U S A *102*, 11775-11780.

Wagner, A. (2005b). Robustness and evolvability in living systems (Princeton (New Jersey): Princeton University Press).

Wagner, A. (2011). Genotype networks shed light on evolutionary constraints. Trends in ecology & evolution *26*, 577-584.

Wang, L.Z., Wu, F., Flores, K., Lai, Y.C., and Wang, X. (2016). Build to understand: synthetic approaches to biology. Integr Biol (Camb) *8*, 394-408.

Weiss, K.M., and Fullerton, S.M. (2000). Phenogenetic drift and the evolution of genotype-phenotype relationships. Theor Popul Biol *57*, 187-195.

Wolpert, L. (1996). One hundred years of positional information. Trends Genet *12*, 359-364.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat Rev Genet *8*, 206-216.

Zalts, H., and Yanai, I. (2017). Developmental constraints shape the evolution of the nematode mid-developmental transition. Nature Ecology & Evolution *1*, 0113.

Zhang, X., and Gottlieb, P.A. (1995). Modified nucleotides reveal the indirect role of the central base pairs in stabilizing the lac repressor-operator complex. Nucleic Acids Res *23*, 1502-1511.