



Automatic design of gene regulatory mechanisms for spatial pattern formation



Reza Mousavi¹ & Daniel Lobo^{1,2}

Gene regulatory mechanisms (GRMs) control the formation of spatial and temporal expression patterns that can serve as regulatory signals for the development of complex shapes. Synthetic developmental biology aims to engineer such genetic circuits for understanding and producing desired multicellular spatial patterns. However, designing synthetic GRMs for complex, multi-dimensional spatial patterns is a current challenge due to the nonlinear interactions and feedback loops in genetic circuits. Here we present a methodology to automatically design GRMs that can produce any given two-dimensional spatial pattern. The proposed approach uses two orthogonal morphogen gradients acting as positional information signals in a multicellular tissue area or culture, which constitutes a continuous field of engineered cells implementing the same designed GRM. To efficiently design both the circuit network and the interaction mechanisms—including the number of genes necessary for the formation of the target spatial pattern—we developed an automated algorithm based on high-performance evolutionary computation. The tolerance of the algorithm can be configured to design GRMs that are either simple to produce approximate patterns or complex to produce precise patterns. We demonstrate the approach by automatically designing GRMs that can produce a diverse set of synthetic spatial expression patterns by interpreting just two orthogonal morphogen gradients. The proposed framework offers a versatile approach to systematically design and discover complex genetic circuits producing spatial patterns.

Gene regulatory mechanisms (GRMs) comprise a network of genes and signal molecules together with their mechanistic interactions. They can govern the development of gene expression patterns in time and space, which in turn can control the formation of anatomical structures, organ locations, and complex shapes¹. Understanding the regulatory mechanisms that can produce particular spatial patterns is crucial for the prediction of phenotypes and the identification of interventions toward desired outputs². Furthermore, the design, engineering, and experimental control of complex GRMs in synthetic developmental biology^{3,4}, such as synthetic bistable behaviors⁵, will allow the manufacturing of multicellular patterns, value-added bioproducts, and synthetic smart biomaterials for medical and industrial applications^{6–9}. Towards this, automated algorithms have been proposed to aid the molecular engineering of pre-designed synthetic gene circuits^{10–12} as well as for the design of novel circuits that can implement a given behavior, such as switches, oscillators, or temporal functions^{13–18}. However, automatically discovering or designing a GRM that can produce a target multi-dimensional spatial expression pattern, is still a current major challenge^{19,20}. Although random search has been used for exploring and

engineering GRMs for relatively simple spatial patterns such as a stripe^{21,22}, heuristic optimization methods are required for automatically designing complex GRMs²³.

Different heuristic methods exist for reconstructing complex gene regulatory networks (GRNs)—a network of predicted gene-gene regulatory links lacking mechanistic information. These approaches infer links in the network as probabilistic gene-gene interactions²⁴, typically from large-scale transcriptomics data. Unsupervised machine learning can take unlabeled data from experimental or synthetic nonspatial gene expression patterns, such as microarray or RNA-Seq transcriptomics, to predict GRNs^{25–32}. In contrast, supervised machine learning uses known gene-gene interactions as labeled data for training the models and then infer GRNs from non-spatial transcriptomics data^{33–36}. In addition, automated methods have been proposed to infer GRNs from spatial gene expression patterns in in situ hybridization images, such as those during early *Drosophila* development^{37,38}. However, these GRN-inference methods are limited to inferring the topology of the gene regulatory network through link prediction and hence lack the mechanistic details in the gene interactions

¹Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD, USA. ²Greenebaum Comprehensive Cancer Center and Center for Stem Cell Biology & Regenerative Medicine, University of Maryland, Baltimore, Baltimore, MD, USA. e-mail: lobo@umbc.edu

essential for predicting spatial pattern formation dynamics. Indeed, inferring such GRMs needs computational methods based on dynamic mathematical formalisms that can mechanistically model signal- and gene-gene interactions to predict the resulting spatial phenotypes^{39–41}.

Several heuristic optimization methods have been proposed for the inference of dynamic GRMs from spatial expression patterns¹⁹. These pattern-forming GRMs are typically formalized with partial differential equations (PDEs) due to their ability to combine controlling mechanisms of gene regulation with spatial signaling and their resulting cellular behaviors in time and space^{21,22}. Evolutionary computation is a heuristic population-based method that can optimize complex solutions^{42–46}. This evolutionary approach can infer the parameters of one-dimensional PDE models of the development of natural gene expression patterns, such as in the early *Drosophila* embryo^{47–49}, as well as both the parameters and structure of GRMs for one-dimensional embryonic dynamic patterns^{50,51} and simple two-dimensional planarian head-trunk-tail body spatial patterns^{52–54}. In addition, evolutionary computation has been proposed for designing synthetic gene circuits⁵⁵ that can perform relatively basic tasks, including switches⁵⁶, logic gates⁵⁷, and oscillators⁵⁸. Evolutionary computation methods have been combined with Mixed Integer Nonlinear Programming local solvers⁵⁹ to design biological circuits with dynamic behaviors such as switches and oscillators⁶⁰ that also can be resilient to molecular noise⁶¹. These methods also have been demonstrated for the optimization of the regulatory interactions in three-gene circuits capable of forming a stripe pattern⁶². However, the automatic design of GRMs that can produce an arbitrarily complex, multi-dimensional spatial pattern is still a current challenge.

Here, we present a novel methodology for the automatic design of GRMs that can dynamically produce any given spatial gene expression pattern in response to positional information gradient signals. The method leverages the advantages of evolutionary computation and high-performance computing⁶³ to rapidly design spatiotemporal GRMs. We defined a versatile set of non-linear gene regulatory mechanisms that serve as building blocks for the optimization method to design GRMs that develop the target spatial pattern. Furthermore, the method can be tuned to produce

either complex GRMs that develop precise patterns or simple GRMs that develop approximate patterns. We evaluate the performance of the methodology by successfully inferring GRMs for a diverse set of synthetic two-dimensional spatial patterns, including geometric shapes, symbols, and characters.

Results

A system for spatial pattern formation based on orthogonal gradient signals

Patterns in developmental biology are often formed by GRMs that react to diffusible morphogen signals producing spatial gradients⁶⁴. These signals act as a positional information system for cells to react differentially depending on their location⁶⁵, a process that can be applied to synthetic spatial behaviors⁶⁶ and reinforced with synthetic bistable switches⁵. Here we employed a similar in silico approach based on continuous orthogonal morphogen gradients that serve as input signals for an automatically designed GRM to form a target spatial pattern (Fig. 1). The two input signals (labeled red and green) are produced from the top and left sides, respectively, of a two-dimensional cell culture domain and form similar but orthogonal static gradients. Each cell in the domain encodes the same GRM, which takes as input the two input signals and through a cascade of regulatory interactions expresses a non-diffusible reporter gene (blue).

Starting with the input morphogens forming the orthogonal gradients and all the other products at zero concentration, the goal of the designed GRM is to dynamically process the input morphogen signals to express the reporter gene in a stable spatial pattern similar to a target pattern. In addition to the input signals and the reporter gene, the GRM can include intermediate genes to form complex regulatory networks. The input morphogens and intermediate genes can regulate the reporter gene and other intermediate genes, but the reporter gene cannot regulate any other gene. Except the input gradient morphogens, all products are confined intracellularly. GRMs are modeled as a system of partial differential equations (PDEs) where each gene is represented by an equation defining its rate of change in product concentration due to their regulatory interactions and decay.

Fig. 1 | Designing gene regulatory mechanisms for forming a target spatial gene expression pattern in response to two-dimensional orthogonal morphogen gradients. **A.** Two static input morphogens (labeled red and green) form orthogonal gradient signals across a two-dimensional tissue or cell culture. The green morphogen forms a gradient from left to right (left being high), and the red morphogen forms a gradient from top to bottom (top being high). **B.** Each cell in the domain implements the same gene regulatory mechanism (GRM) including the input morphogens, an output reporter gene for the spatial pattern (blue), and intermediate genes (a–h). **C.** The regulatory mechanisms defined in the GRM process the input morphogen gradients so that the expression pattern of the output gene forms the target spatial pattern.

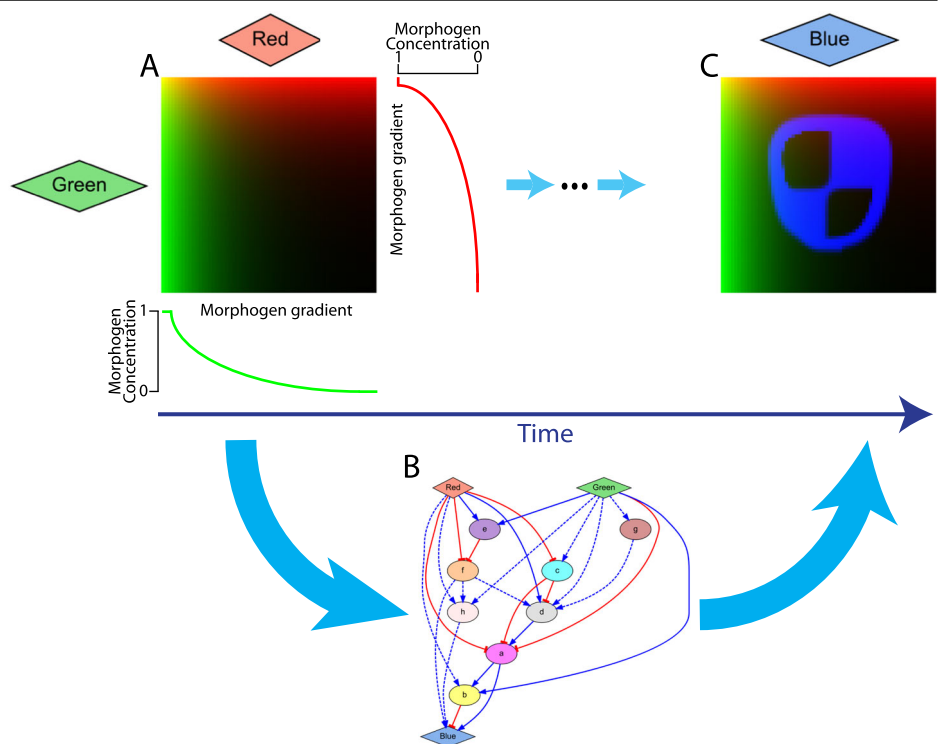
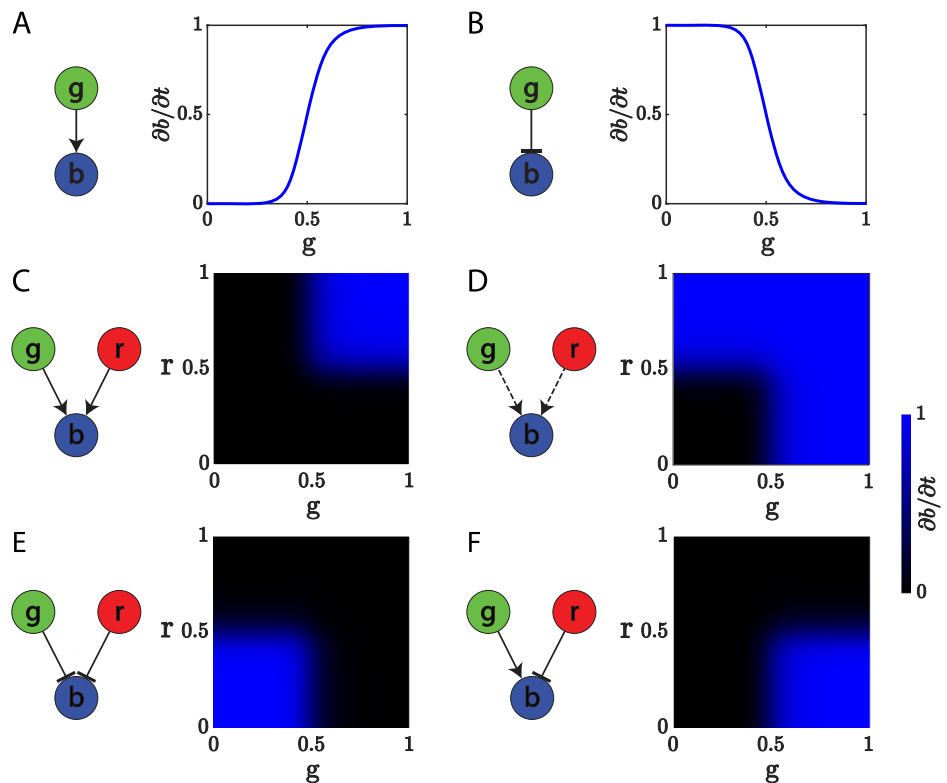


Fig. 2 | A versatile modeling framework for gene regulatory mechanisms. Using Hill equations combining different terms, continuous non-linear regulatory interactions can be modeled as positive (point arrow) or negative (blunt arrow) and can be grouped as necessary (solid line) or sufficient (dashed line) manner. Illustrative mechanisms are shown with one or two interactions between two input genes ('g' and 'r') regulating the rate of expression of a reporter gene ('b'). **A.** Single positive regulation. **B.** Single negative regulation. **C.** Double necessary positive regulation (*AND* logic). **D.** Double sufficient positive regulation (*OR* logic). **E.** Double negative regulation (*NOR* logic). **F.** Double positive and negative regulation (*NIMPLY* logic). Three or more regulatory genes can be combined in a similar fashion to produce complex regulatory interactions.



A versatile modeling framework for non-linear regulatory mechanisms

Biological regulatory interactions between signals and genes are continuous and non-linear and can act as enhancers (positive regulation) or inhibitors (negative regulation). Furthermore, multiple regulatory interactions can affect the same gene in a necessary or sufficient fashion. To allow the design and simulation of such a large variety of gene regulatory mechanisms, we have designed a versatile mathematical approach based on Hill equations combining different terms as building blocks to model the integration of any number of regulatory interactions in GRMs. Single positive and negative regulations follow a sigmoidal response modeled by a simple Hill equation. Multiple positive regulations can be grouped as necessary (similar to an AND gate) or sufficient (similar to an OR gate) by combining multiple terms in the Hill equation numerator with either multiplication or multiplication and summation operators, respectively. For simplicity, negative regulations are always combined with a multiplication operator (AND gate) in this work. Figure 2 illustrates the approach with examples including one and two regulatory interactions (see Methods for equations). The rate of expression of a product ('b') depends on the concentration levels of its regulatory products ('g' and 'r') as well as the sign (positive or negative) and grouping (necessary or sufficient) of its regulatory interactions. In this way, multiple types of gene regulations can be combined to produce a large variety of continuous regulatory mechanisms, such as the *AND*, *OR*, *NOR*, and *NIMPLY* logic illustrated. Similarly, any number of regulatory interactions can be combined to produce a versatile set of possible genetic mechanisms.

Automatic design of GRMs

Designing a GRM that can produce a given target function is a current challenge, especially when including spatial features. To streamline this task, we developed a machine learning methodology to automatically design GRMs able to recapitulate the formation of a given spatial pattern. The method makes use of the two-dimensional orthogonal input morphogen gradients together with the versatile regulatory modeling framework based on Hill equations to design GRMs that can form a stable spatial expression pattern. The approach is based on parallel evolutionary computation, where

a population of candidate GRMs evolve by iteratively crossing, mutating, simulating, and scoring them until a GRM that can recapitulate the target pattern is found (Fig. 3). The method takes as input a target spatial pattern (e.g., a square shape) and returns a complete GRM—including the number of intermediate genes, regulatory interactions, and parameters—that when simulated produces a reporter gene with a spatial expression pattern similar to the given target pattern.

The pseudocode for the evolutionary algorithm to design regulatory mechanisms is described in Box 1 (see Methods for details). The algorithm starts with a random population of GRMs, each including the gradient input signals and the reporter gene, a random number of intermediate genes, random interactions among all products (except the gradient signals, which have no regulators, and the reporter gene, which cannot regulate other products), and random parameters. Based on the presented framework for non-linear regulatory mechanisms, GRMs are translated into a system of partial differential equations, which can then be numerically simulated to score their capacity to produce the target expression pattern (error of the model). The GRMs that produce the most similar and stable patterns as compared to the input target pattern are kept in the population, while GRMs with higher errors are discarded. The population then produces new offspring GRMs by stochastically crossing those in the current population and adding random mutations. The new offspring GRMs are then simulated, scored, and added to the population to select the next generation. This iterative process continues until a GRM with zero error is found, representing a GRM that can produce the input target pattern.

The fitness of a GRM scores its ability to stably form the target spatial pattern and is defined with an error function (see Methods for its mathematical expression). The error of a GRM is computed at the last time step of the simulation as the sum of the average difference between each domain location (pixel) in the target and the developed pattern plus the maximum concentration change (penalizing patterns not in equilibrium). Two thresholds (parameters α and β , respectively, in the error function) are defined for both measurements to avoid overfitting and bloating. In this way, GRMs with concentration and stability scores below these thresholds have an error of zero, which avoids further complexification of the GRMs

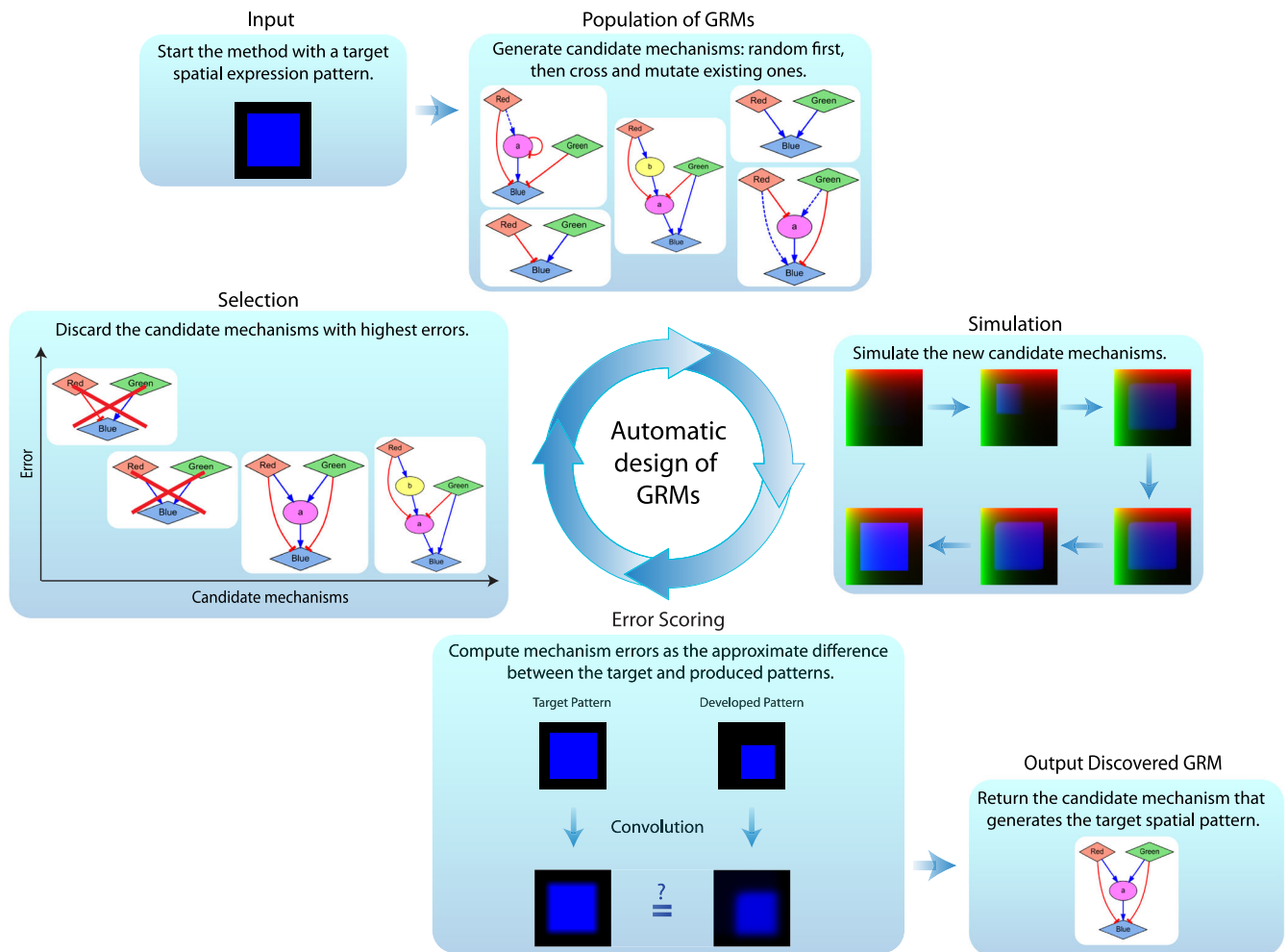


Fig. 3 | Automatic methodology based on evolutionary computation for the design of gene regulatory mechanisms producing a given spatial expression pattern. The method takes as input a target spatial pattern (a square shape in this example), iteratively generates a population of candidate mechanisms by crossing

and mutating them, simulates and computes their errors as their ability to produce the target pattern, and discards those with the highest errors. This iterative process repeats until a regulatory mechanism with zero error is discovered.

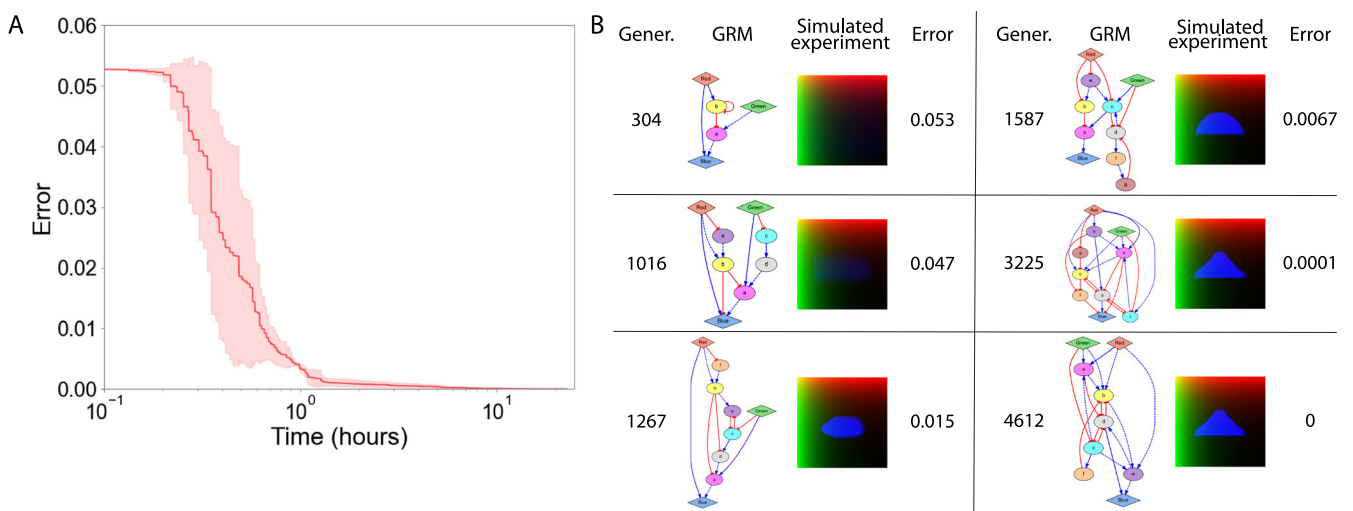
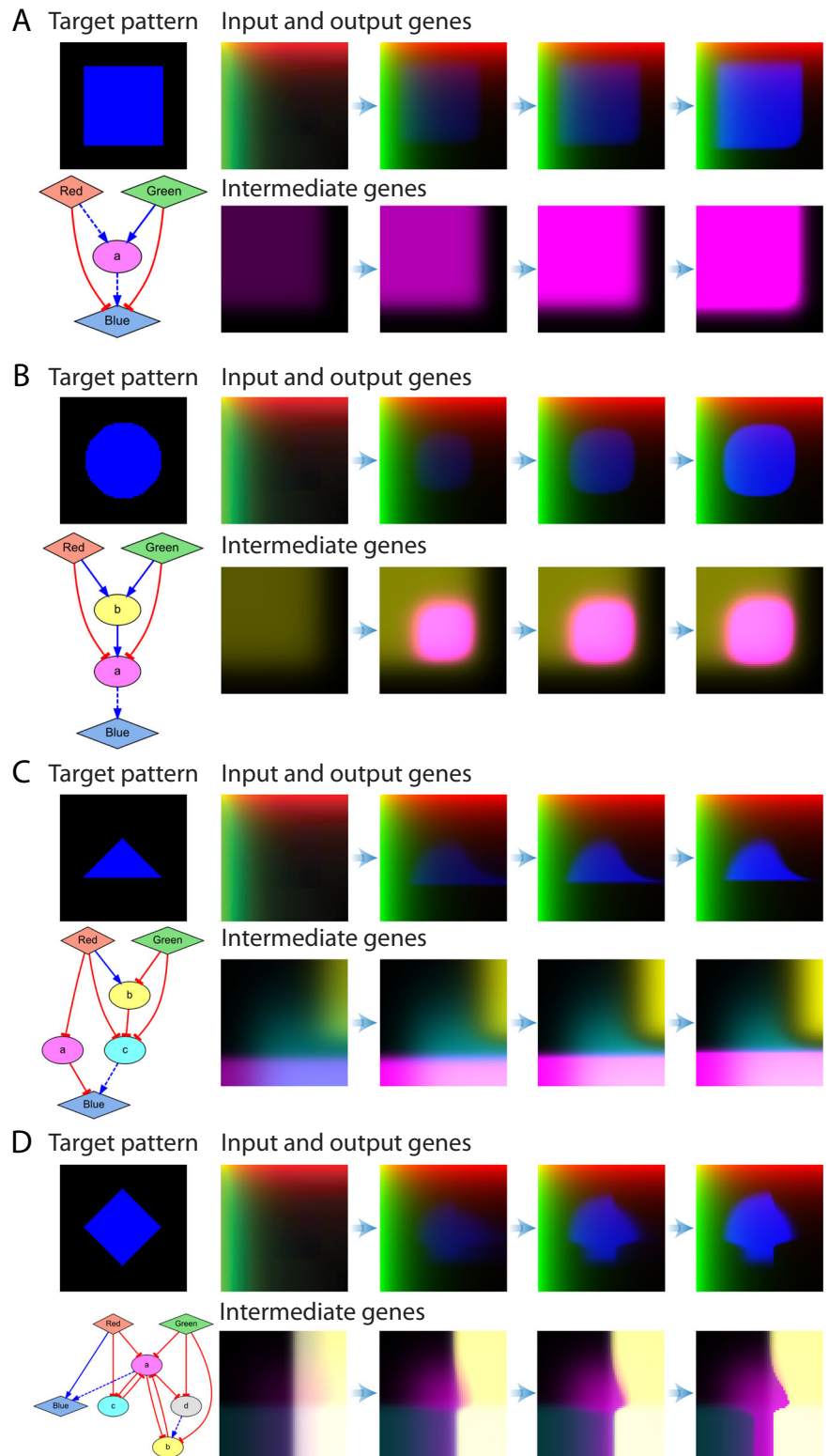


Fig. 4 | Evolutionary dynamics for the design of a GRM producing a triangle pattern. A. Average error of the best GRM across three independent runs of the algorithm, all reaching zero error after ~20 h of execution. The shaded area represents the standard deviation. Error parameters: $k = 5$, $\alpha = 0.1$, $\beta = 0.001$. B.

Representative candidate GRMs at different generations during the evolutionary search, showing the gradual increase in complexity and improvement in their capacity to produce the target triangle pattern.

Fig. 5 | Automatically designed GRMs for different geometric target shapes. The simulation of a GRM starts with all the genes at zero concentration except the input signals (green and red), which form static orthogonal gradients. All the resulting GRMs correctly produce a steady-state target pattern (blue) with zero error for each geometric shape, including a square (A), circle (B), triangle (C), and diamond (D). The dynamics of the output and intermediate genes reveal the spatial computations performed by each mechanism to produce their target patterns. Expression colors in the simulation correspond to the node colors in the GRM network diagrams. Error parameters: $k = 7$, $\alpha = 0.25$, $\beta = 0.001$.



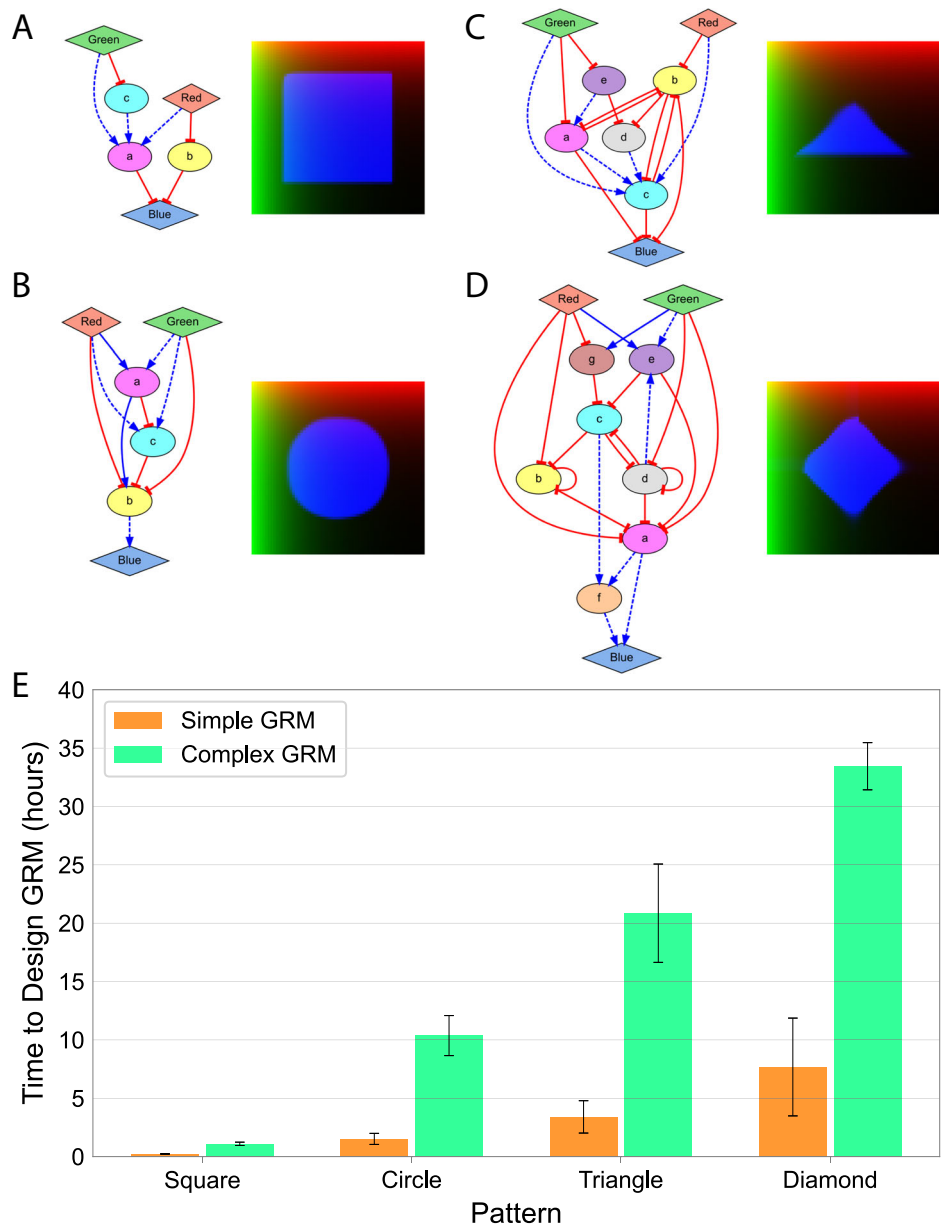
without meaningful improvements in fitness. Importantly, the spatial pattern fidelity required for the designed GRMs can be adjusted in the method, since sharp and complex pattern boundaries may require excessively complex GRMs for a given application. Before calculating the fitness, both the target and developed patterns are processed with a box blur kernel convolution to eliminate sharp spatial features. Similar to the concentration and equilibrium thresholds, the strength of the convolution function can be adjusted with a parameter defining the kernel size (parameter k ; higher

values representing more approximate patterns). Thus, the fidelity of the GRM to produce the target pattern can be adjusted with error parameters at the concentration, equilibrium, and spatial levels.

Inferred GRMs for geometric patterns

We tested the proposed methodology for the design of GRMs that can produce different geometric target spatial patterns. Figure 4 shows an illustrative example of the evolutionary dynamics of the algorithm across

Fig. 6 | The method includes tolerance parameters to control the complexity of the designed GRMs. The kernel size (k) and error concentration threshold (α) parameters determine the complexity of the designed GRMs and their precision in producing the target patterns. **A–D.** Lowering the tolerance parameters results in more complex GRMs but with more precise patterns as compared to Fig. 5. Error parameters: $k = 5$, $\alpha = 0.1$, $\beta = 0.001$. **E.** Time needed by the algorithm to design simple (Fig. 5) or complex (A–D) GRMs with zero error. Each condition was run three times. Error bars denote standard deviation.



three independent runs for the design of a GRM to develop a spatial pattern with a triangle shape. The initial models are random and produce patterns far from the target, hence with high error scores. After many generations of crossover and mutations, including adding de novo intermediate nodes, the models increase their complexity in terms of number of genes and links and their ability to produce the target spatial pattern. After ~20 hours of computation time, the evolutionary process finds GRMs that can produce the target triangle pattern with zero error (for the given error thresholds).

The regulatory links and intermediate genes automatically added by the evolutionary algorithm perform the necessary spatial computations to produce the target spatial patterns. Figure 5 shows examples of the resultant GRMs discovered by the automated methodology for four different target geometric shapes, all found in less than 35 hours of computational time and 23,500 generations (Supplementary Fig. 1 and Supplementary Movies). Target patterns with edges parallel to the orthogonal gradients (Fig. 5A; square pattern) require less complex GRMs than patterns with oblique lines (Fig. 5C-D; triangle and diamond). To produce curved edges (Fig. 5B, circle), the algorithm takes advantage of regulatory interactions with gradual

slopes, which produce softer responses at the edges. Possible synthetic realizations of the discovered GRMs are illustrated using SBOL notation⁶⁷ in Supplementary Fig. 2.

The simulation dynamics show how the regulatory interactions designed in the GRMs translates into spatial computations to produce the target patterns. For the square pattern (Fig. 5A), the negative regulations between the input gradient signals (red and green) and the output reporter gene (blue) prevent the latter from being expressed at high input concentrations (top and left sides of the domain). In addition, the input signals positively regulate with a low threshold an intermediate gene ('a'), resulting in no expression in the bottom and right areas. Then, a positive regulation between the intermediate and output gene defines the bottom and right edges of the square pattern. The GRM for the circle pattern (Fig. 5B) extends this design with an additional gene ('b') at the end of the pathway that together with more gradual regulatory interactions produce the curved edges needed for the circle pattern. The discovered GRMs for the triangle and diamond patterns (Fig. 5C-D) include three and five intermediate genes, respectively, that define intermediate expression patterns at different domain locations to produce the target geometric shapes.

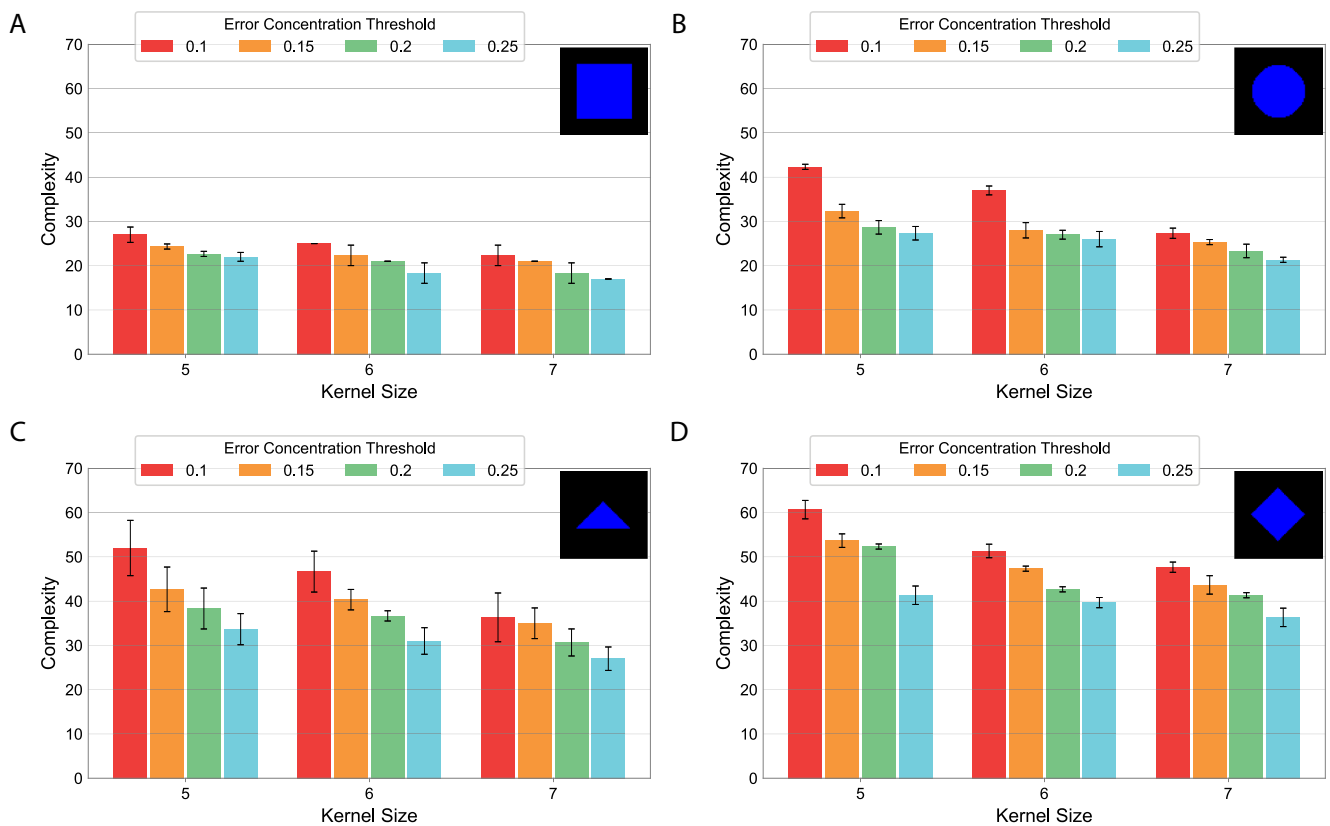


Fig. 7 | Complexity of the designed GRMs discovered by the automated search method for the geometric shapes with different tolerance parameters. The complexity of a GRM was measured as the number of links plus three times the number of genes. Three runs were performed for each target pattern, including the

square (A), circle (B), triangle (C), and diamond (D), and parameter values for kernel size (k) and error concentration threshold (α). Error bars denote standard deviation.

Adjusting the complexity and pattern precision of the designed GRMs

The tolerance of the proposed method can be adjusted for different design needs, from complex regulatory mechanisms that produce exact patterns to simple mechanisms that produce approximate patterns. For this, different values can be set for the kernel size (k) and concentration threshold (α) parameters used in the error fitness function. Figure 6 shows the results of lowering these parameters to design more complex GRMs for the same geometric target shapes as for the previous simpler networks, all with zero error. While the simpler GRMs produced approximate patterns with diffuse edges (Fig. 5), especially those that are not parallel to the gradient signals, the complex GRMs produced precise patterns with sharp edges, even at different angles to the gradient signals or forming curves (see also Supplementary Movies). Conversely, the time needed by the algorithm to design complex networks that can produce precise patterns was significantly longer (about 5x) than for designing simpler networks to produce approximate patterns (Fig. 6E).

The complexity of the GRMs designed consistently depends on the target pattern and the tolerance parameters used. Figure 7 shows the average complexity of GRMs for the square, circle, triangle, and diamond target patterns obtained with different values of kernel size and error concentration thresholds. The results illustrate how the complexity of the discovered GRMs for each pattern decreases as the kernel size and the error concentration threshold parameters increase. The effect of these parameters is more acute in the case of complex patterns (Fig. 7D) as compared to simpler ones (Fig. 7A). Moreover, increasing the error concentration threshold results in more diffuse edges in the developed patterns, while increasing the kernel size results in less precise shapes (Supplementary Figs. 3–6).

To test the ability of the method to design constrained GRMs in terms of interaction types and genes, we performed runs limiting the Hill

coefficients (which define the slope of the interactions) or the maximum number of genes. The results demonstrated that with a limited set of Hill coefficients (1, 2, 4, 8, or 10) the method is still able to find both simple and complex GRMs with zero error for all the target geometric shapes (Supplementary Fig. 7). Limiting the maximum number of genes below the minimum required to form a particular pattern results in GRMs with error values proportional to the number of genes missing (Supplementary Fig. 8). Crucially, setting the limit higher than the minimum number of genes required still results in GRMs containing the minimum number of genes.

Designing GRMs for arbitrary shapes and biological patterns

To test the ability of the method to design GRMs for arbitrary shapes, we applied it to discover models that can produce gradients, periodic shapes, symbols, and characters. Figure 8 shows the developed expression patterns produced by the designed GRMs, all reaching zero error (see Supplementary Figs. 9,10 and Supplementary Movies for the target patterns and the expression of intermediate genes). The complexity of the discovered networks varies from 29 to 91 (as the number of edges plus three times the number of genes). The results show how the method can design GRMs for complex shapes with fine details such as gradients, curved lines, and pointed ends—all produced by the interpretation of two orthogonal morphogen gradients (red and green).

To assess the capability of the automated methodology to design GRMs for biological patterns—including multiple output reporter genes—we sought to reverse engineer the gap gene expression pattern observed during early *Drosophila* development. The input and target gene expression patterns used in the method are wild-type concentrations of the protein products at the late syncytial blastoderm stage of *Drosophila melanogaster* (cleavage cycle 14 A, $t = 62$ min), as reported in⁶⁸. The products include the two input signals Bicoid (Bcd) and Caudal (Cad) (Fig. 9A, red and green),

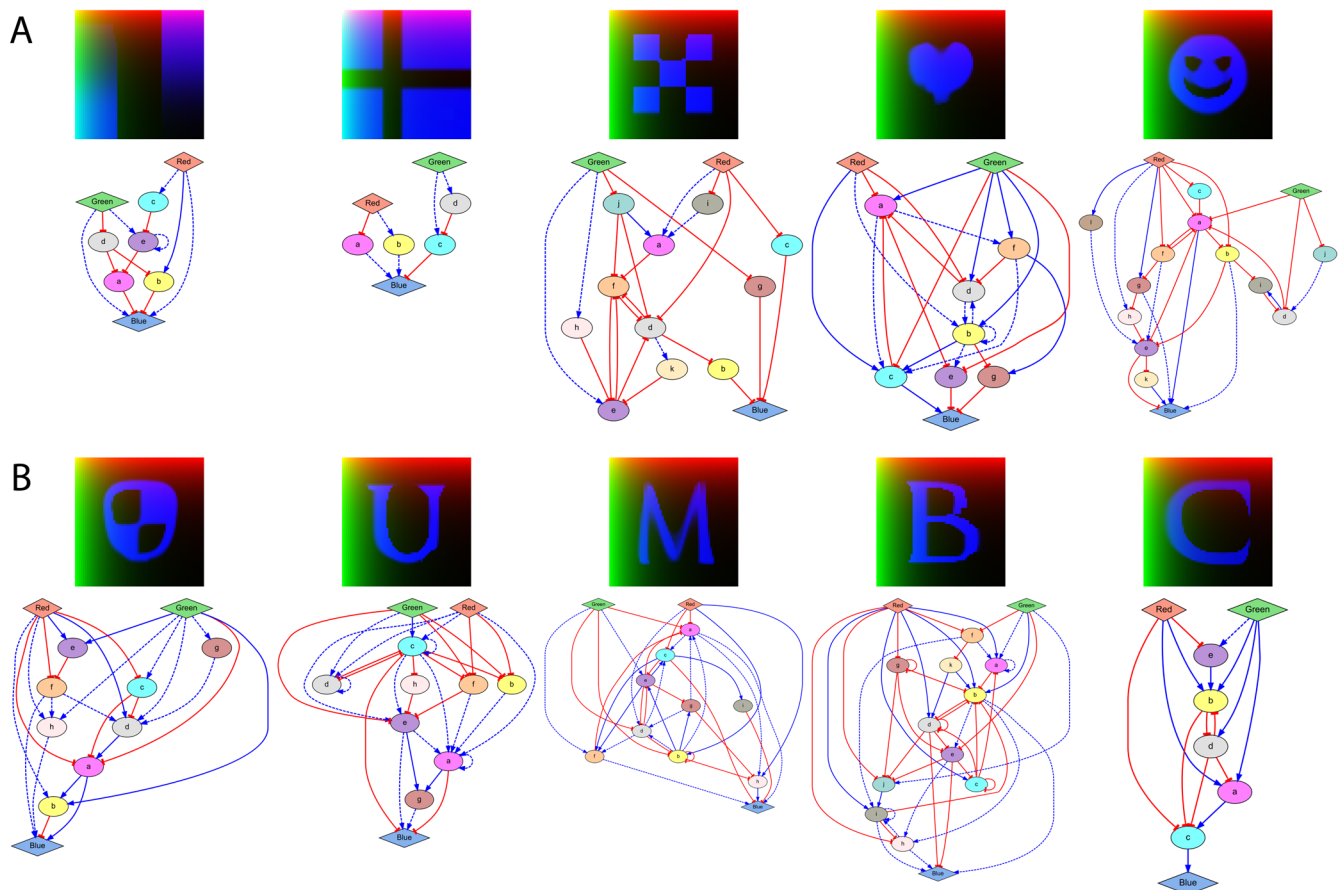


Fig. 8 | Patterns produced by GRMs discovered by the automated methodology applied to arbitrary shapes. The patterns tested for arbitrary shapes include: (A) gradients, flag, checkerboard, and symbols; (B) logo and characters. Parameters: $k = 5$, $\alpha = 0.1$, $\beta = 0.001$.

which form gradients from the anterior (left) and posterior (right) sides, respectively. These two signal gradients provide positional information to develop the target gap gene patterns, including the expression levels of Giant (Gt), Hunchback (Hb), Knirps (Kni), and Kruppel (Kr) (Fig. 9A, cyan, blue, yellow, and magenta, respectively). Candidate GRMs include the input genes, the gap genes, and any number and type of regulatory interactions from the input to the gap genes and between gap genes. Figure 9C shows the GRM discovered by the method, which when simulated can successfully produce the target gap gene expression pattern with zero error (Fig. 9D and Supplementary Movies). Furthermore, the designed GRM is very similar to recently published models^{69–71}, and includes the characteristic double-negative feedback loops between Hb/Kni and Kr/Gt.

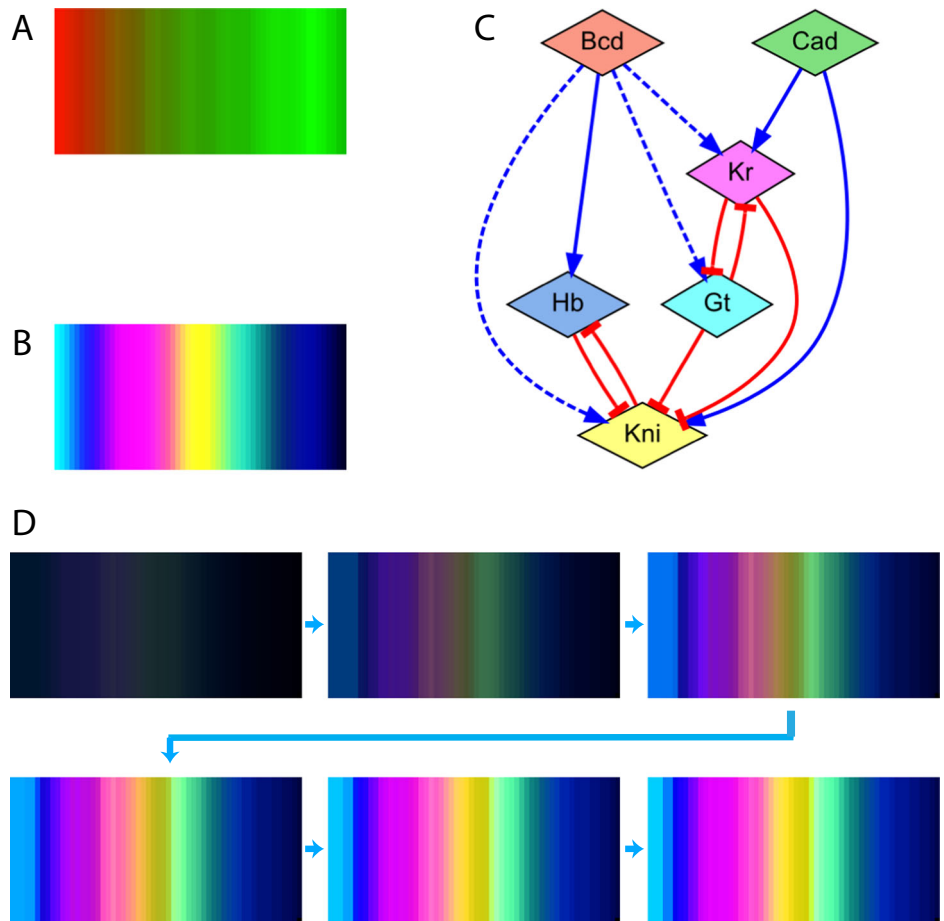
Discussions

To streamline the design and understanding of gene regulatory mechanisms (GRMs) capable of producing spatial patterns in response to morphogen gradients, here we proposed a novel methodology integrating a framework to model arbitrary genetic circuits with evolutionary computation and high-performance computing. The method can design synthetic GRMs—including all the necessary genes, regulatory links, and parameters—that when simulated develop a given target spatial pattern in response to orthogonal morphogen gradient signals. Crucially, a GRM is formalized as a PDE system, which allows the simulation of the dynamics of spatially distributed systems such as those forming spatial patterns. We demonstrated the capacity of the method to design GRMs able to interpret the positional information provided by the signal gradients and produce a variety of target patterns, including geometric shapes, symbols, and characters. The method is generic and could be adapted to particular applications spanning diverse size and time scales. For example, the discovered synthetic networks could be engineered in bacteria to form in 24 h such spatial patterns with

approximate dimensions of $1 \times 1 \text{ cm}^{5,66}$, while the patterns in *Drosophila* have approximate dimensions of $0.5 \times 0.2 \text{ mm}$ and develop in 1 h⁶⁸. Crucially, the precision of the designed GRMs to produce a given pattern can be adjusted with a convolution-based fitness function that evaluates their ability to recapitulate the target pattern. The results showed how these parameters can modulate the trade-offs between the precision of the produced pattern, the complexity of the designed mechanism, and the speed of the machine learning method. Hence, while the presented methodology can design very complex GRMs, it is also flexible enough to automatically limit the complexity of such networks for a variety of applications and studies.

The proposed methodology is highly versatile but could be extended with other functionality in future work. The products of intermediate genes are restricted to intercellular pathways, but the methodology could include intermediate genes that produce diffusible signals, such as ligands acting intracellularly. This would allow more complex spatial regulation involving dynamic morphogens, as found in developmental processes⁷² and novel synthetic biology applications⁷³. Indeed, this study is focused on morphogen gradients as the input signals to produce spatial patterns, but such diffusible morphogens could pave the way to automatically designing self-regulated mechanisms for pattern formation, such as Turing systems for the engineering of synthetic spots, stripes, and other periodic patterns^{74,75}. The framework is currently limited to two-dimensional domains and target spatial patterns, but it could be easily expanded to three-dimensional pattern systems⁷⁶. The machine learning method can explore and use a large range of in silico regulatory interactions and degradation rates. However, the method could be constrained with a predetermined set of standard biological parts to facilitate the engineering of the automatically designed GRMs into synthetic biological circuits^{11,77,78}. The presented method starts with

Fig. 9 | GRM discovered by the automated methodology for the *Drosophila* gap gene system. A. Two input morphogens Bcd and Cad (red and green) form anti-parallel gradient signals across the anterior-posterior axis in a *Drosophila* embryo (anterior to the left). **B.** The target pattern includes the expression levels of four gap genes Gt, Hb, Kni, and Kr (cyan, blue, yellow, and fuchsia, respectively) obtained experimentally. **C.** Resultant GRMs discovered by the automated methodology. **D.** Simulation of the designed GRM results in the formation of the correct target biological pattern with zero error. Regulatory interactions can act as positive (point arrow) or negative (blunt arrow) and can affect the same gene as necessary (solid line) or sufficient (dashed line). Expression colors in the simulation correspond to the node colors in the GRM network diagram. Domain size: 58×29. Error Parameters: $k = 5$, $\alpha = 0.1$, $\beta = 0.001$.



random initial mechanisms. However, to improve its efficiency, the initial population could contain models including genes and interactions of known necessary pathways, or specific GRMs known to produce related patterns. Finally, the algorithm is limited to return a single GRM for each run. However, multiple GRMs could lead to the formation of the same pattern (see Supplementary Fig. 8). Indeed, a major challenge for complex phenotypes is to discover a comprehensive set of GRNs that can develop a given gene expression pattern, i.e., an atlas of regulatory designs^{21,79}. Future work will extend the presented methodology with evolutionary multi-objective and diversity-preserving algorithms⁸⁰ for the discovery of a comprehensive set of GRMs that can produce a given spatial pattern.

In conclusion, the capacity of the presented method to automatically design GRMs for spatial patterns could be essential to transition from our current ability to understand and implement synthetic small modules to be able to identify and assemble larger scale systems⁸¹. Recent advancements have made it possible to streamline the engineering of synthetic GRMs for a given arbitrary circuit⁸². The presented methodology expands these approaches by automating also the design of such circuits for producing complex spatial patterns, which could potentially be applied to current bioengineering problems—from the synthesis of complex bioproducts in industrial, pharmaceutical, and biomaterial applications⁸³ to the construction of multicellular synthetic systems⁸⁴. In addition, these advancements could streamline the systematic study of natural genetic circuits towards the discovery of complex mechanisms controlling tissue spatial behaviors⁸⁵ and whole-body patterns⁸⁶ from morphogen gradients. Indeed, it is a current challenge to produce mechanistic hypotheses in terms of GRMs that can recapitulate observed spatial phenotypes. The methodology presented here could be employed to automatically infer such hypotheses directly from datasets of curated experimental gene expression patterns^{87,88}. Overall, the

presented method to aid in the design of GRMs able to produce arbitrary gene expression patterns has the potential to both enable the understanding of complex developmental processes as well as the design of complex dynamic synthetic systems.

Methods

Simulation of GRMs

We developed a simulator of GRMs for spatial pattern formation based on a system of nonlinear partial differential equations (PDEs). Gene regulations are based on Hill equations and a GRM consists of two input morphogens, intermediate genes, and a reporter gene. The two input morphogens (red R and green G) have a constant gradient distribution given by $M_R = d^j$, $M_G = d^i$, where $d = 0.93$ and (i, j) is the cell position in the domain. Intermediate genes can be activated or repressed by other genes except for the output gene, which can be regulated but cannot regulate other genes—since it represents a reporter signal that defines the output pattern produced by the GRM. Gene products are confined intracellularly and decay over time.

Each gene in a GRM includes four parameters: production, decay, diffusion, and basal expression. Each regulatory interaction is modeled as positive (activating) or negative (inhibiting) with a Hill equation including two parameters: Hill coefficient (modulating the link sensitivity) and binding constant (modulating the link strength). Genes can be regulated by several other genes simultaneously, and these regulations can be grouped as necessary or sufficient. Necessary positive regulations are combined with a multiplication operator (AND logic), while sufficient positive regulations are combined with both a multiplication and summation operation (OR logic). Negative regulations are combined with a multiplication operator (AND logic). This methodology guarantees that the strength of any regulation lies within the range $[0, 1]$. The illustrative examples of single and

double regulations shown in Fig. 2 are modeled by the following equations:

$$\text{Single positive regulation (Fig. 2A)} : \frac{\partial b}{\partial t} = \frac{(\alpha_1 g)^{\eta_1}}{1 + (\alpha_1 g)^{\eta_1}} \quad (1)$$

$$\text{Single negative regulation (Fig. 2B)} : \frac{\partial b}{\partial t} = \frac{1}{1 + (\alpha_1 g)^{\eta_1}} \quad (2)$$

$$\text{Double necessary positive regulation (Fig. 2C)} : \frac{\partial b}{\partial t} = \frac{(\alpha_1 g)^{\eta_1} (\alpha_2 r)^{\eta_2}}{[1 + (\alpha_1 g)^{\eta_1}][1 + (\alpha_2 r)^{\eta_2}]} \quad (3)$$

$$\text{Double sufficient positive regulation (Fig. 2D)} : \frac{\partial a}{\partial t} = \frac{(\alpha_1 g)^{\eta_1} (\alpha_2 r)^{\eta_2} + (\alpha_1 g)^{\eta_1} + (\alpha_2 r)^{\eta_2}}{[1 + (\alpha_1 g)^{\eta_1}][1 + (\alpha_2 r)^{\eta_2}]} \quad (4)$$

$$\text{Double negative regulation (Fig. 2E)} : \frac{\partial b}{\partial t} = \frac{1}{[1 + (\alpha_1 g)^{\eta_1}][1 + (\alpha_2 r)^{\eta_2}]} \quad (5)$$

$$\text{Double positive and negative regulation (Fig. 2F)} : \frac{\partial b}{\partial t} = \frac{(\alpha_1 g)^{\eta_1}}{[1 + (\alpha_1 g)^{\eta_1}][1 + (\alpha_2 r)^{\eta_2}]} \quad (6)$$

where η_i is the Hill coefficient and α_i is the binding constant in the regulatory interaction i ($\eta_i = 10$ and $\alpha_i = 2$ for all regulations shown in Fig. 2).

In this way, the rate of change of a gene can integrate any combination of interactions. For example, the following PDE describes a gene a regulated by two activators b and c as sufficient, two activators d and e as necessary, and an inhibitor f :

$$\frac{\partial a}{\partial t} = \rho_a \frac{\beta_a + [(\alpha_1 b)^{\eta_1} (\alpha_2 c)^{\eta_2} + (\alpha_1 b)^{\eta_1} + (\alpha_2 c)^{\eta_2}] (\alpha_3 d)^{\eta_3} (\alpha_4 e)^{\eta_4}}{[1 + (\alpha_1 b)^{\eta_1}][1 + (\alpha_2 c)^{\eta_2}][1 + (\alpha_3 d)^{\eta_3}][1 + (\alpha_4 e)^{\eta_4}]} - \lambda_a a + D_a \nabla^2 a \quad (7)$$

where ρ_a is the gene production constant, β_a is the gene basal expression level, η_i are the regulation Hill coefficients, α_i are the regulation binding constants, λ_a is the gene decay constant, and D_a is the gene diffusion constant. All parameter units are arbitrary. The basal expression level is zero when a gene receives at least one positive regulation; otherwise, it is one to model a constitutive promoter that can be inhibited by other genes.

GRM fitness error

The error of a developed expression pattern is calculated at the last time step of the simulation (100 steps with $dt=1$ in this work) by comparing it to the input target expression pattern with a kernel-based method. In this approach, both the developed and target patterns are first blurred by a box blur kernel convolution. The error of a candidate GRM is then calculated as the sum of the average log difference between each domain location (pixel) in the simulated pattern from the candidate GRM and the target pattern, plus the maximum concentration change at the last time step, which penalizes patterns not in equilibrium. In this way, the fitness function represents an approximate comparison between the developed and target images, ignoring small details. Hence, the error between a developed expression pattern D at the last time step in the simulation and the input

target expression pattern T is calculated as:

$$\text{error}(D, T) = \frac{1}{w \cdot h} \sum_{i=1}^w \sum_{j=1}^h \log \left(1 + \left(|(\omega * D)_{i,j} - (\omega * T)_{i,j}| - \alpha \right)^+ \right) + (\Delta_D - \beta)^+ \quad (8)$$

where w and h are the simulation domain width and height, respectively, k is the kernel size, ω is the box blur kernel defined as $\omega = \frac{1}{k} J_k$, where J_k is the unit matrix, α is the error concentration threshold, Δ_D is the maximum concentration change in D at the last time step in the simulation, and β represents the equilibrium penalty threshold. $(x)^+$ indicates the positive part function of x , which outputs 0 if it is negative and x if x is nonnegative. $f * g$ represents the convolution of f and g .

Machine learning method

The machine learning method is based on evolutionary computation to automatically design a GRM that can produce a given pattern. The algorithm evolves a population of candidate regulatory mechanisms iteratively by reproduction, fitness calculation, and selection to find an optimal GRM. The population follows an island distribution approach⁸⁹ to maximize parallelism and ensure robustness and diversity of the candidate mechanisms. The method produces new GRMs by stochastically mixing existing ones and adding random mutations in each generation.

A crossover creates two new children GRMs by randomly combining two existing regulatory mechanisms in the population. For this, the genes of both parents along with their regulatory links are distributed randomly to the two children GRMs (without gene duplication or changes in kinetic parameters). Next, mutations are applied randomly so that genes and regulations can be added, removed, and their parameters replaced with new ones within their ranges from a random uniform distribution. The input and output genes cannot be removed. Deletion mutations were set to a higher probability than duplication mutations to bias the algorithm towards simpler mechanisms and prevent bloating⁹⁰. A deterministic crowding method⁹¹ was used to select new offspring when their errors (fitness) are equal or better than their closest parents. The algorithm runs until a GRM with zero error is found and 250 additional generations have passed without a decrease in its complexity (number of edges plus three times the number of genes of the simplest GRM with zero error).

The parameter ranges for the mutation operator uniform distributions were set as follows: Hill coefficient (1,10), binding constant (1100), decay constant (0.1,1), and production constant (0, 0.1), except for the input morphogens, which is set to 0. Diffusion constants were set to zero for all products, in which case they are considered exclusively intracellular. The meta-parameters of the machine learning method were set as follows: crossover rate 75%, mutation rate 1%, link/gene duplication rate 1%, link/gene deletion rate 7.5%, equilibrium penalty threshold 10^{-3} . All search runs used 32 subpopulations (islands) with 64 individuals each. Islands are randomly paired, and their regulatory mechanisms are shuffled every 250 generations.

Implementation

The methodology was implemented in C++ with the standard, Eigen (Gaël Guennebaud, Benoît Jacob, and others), Qt (The Qt Company Ltd.), and Qwt (Uwe Rathmann and Josef Wilgen) libraries. We implemented the Euler finite difference method⁹² to numerically solve the system of PDEs in a 64×64 domain. The method used 48 parallel threads and was run on a computer server with two 24-core Intel Xeon Gold 6240 R CPUs at 2.4 GHz and 192 GB DDR4 RAM to evaluate its performance.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Supplementary Information includes the system of equations for all the GRMs presented and the Supplementary Movies file includes their simulations.

Code availability

The source code for the method is freely available in GitHub (<https://github.com/lobolab/grm-design>).

Received: 21 November 2023; Accepted: 19 March 2024;

Published online: 02 April 2024

References

- Kicheva, A. & Briscoe, J. Developmental pattern formation in phases. *Trends Cell Biol.* **25**, 579–591 (2015).
- Lobo, D., Lobikin, M. & Levin, M. Discovering novel phenotypes with automatically inferred dynamic models: a partial melanocyte conversion in *Xenopus*. *Sci. Rep.* **7**, 41339 (2017).
- Santos-Moreno, J. & Schaerli, Y. Using synthetic biology to engineer spatial patterns. *Adv. Biosyst.* **3**, 1800280 (2019).
- Zarkesh, I. et al. Synthetic developmental biology: engineering approaches to guide multicellular organization. *Stem Cell Rep.* <https://doi.org/10.1016/j.stemcr.2022.02.004> (2022).
- Barbier, I., Perez-Carrasco, R. & Schaerli, Y. Controlling spatiotemporal pattern formation in a concentration gradient with a synthetic toggle switch. *Mol. Syst. Biol.* **16**, 1–15 (2020).
- Barbier, I., Kusumawardhani, H. & Schaerli, Y. Engineering synthetic spatial patterns in microbial populations and communities. *Curr. Opin. Microbiol.* **67**, 102149 (2022).
- Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
- Kim, H., Jin, X., Glass, D. S. & Riedel-kruse, I. H. Engineering and modeling of multicellular morphologies and patterns. *Curr. Opin. Genet. Dev.* **63**, 95–102 (2020).
- Liu, C. et al. Sequential establishment of stripe patterns in an expanding cell population. *Science* **334**, 238–241 (2011).
- Appleton, E., Madsen, C., Roehner, N. & Densmore, D. Design automation in synthetic biology. *Cold Spring Harb. Perspect Biol.* **9**, a023978 (2017).
- Buecherl, L. & Myers, C. J. Engineering genetic circuits: advancements in genetic design automation tools and standards for synthetic biology. *Curr. Opin. Microbiol.* **68**, 102155 (2022).
- Nielsen, A. A. K. et al. Genetic circuit design automation. *Science* **352**, (2016).
- Dasika, M. S. & Maranas, C. D. OptCircuit: An optimization based method for computational design of genetic circuits. *BMC Syst. Biol.* **2**, 24 (2008).
- Hiscock, T. W. Adapting machine-learning algorithms to design gene circuits. *BMC Bioinf.* **20**, 214 (2019).
- Huynh, L., Tsoukalas, A., Köppe, M. & Tagkopoulos, I. SBROME: a scalable optimization and module matching framework for automated biosystems design. *ACS Synth. Biol.* **2**, 263–273 (2013).
- Marchisio, M. A. & Stelling, J. Automatic design of digital synthetic gene circuits. *PLoS Comput. Biol.* **7**, e1001083 (2011).
- Rodrigo, G. & Jaramillo, A. AutoBioCAD: full biodesign automation of genetic circuits. *ACS Synth. Biol.* **2**, 230–236 (2013).
- Rodrigo, G., Carrera, J. & Jaramillo, A. Computational design of synthetic regulatory networks from a genetic library to characterize the designability of dynamical behaviors. *Nucleic Acids Res.* **39**, e138 (2011).
- Ko, J. M., Mousavi, R. & Lobo, D. Computational Systems Biology of Morphogenesis. in *Computational Systems Biology in Medicine and Biotechnology: Methods and Protocols* (eds. Cortassa, S. & Aon, M. A.) 343–365 (Springer US, New York, NY). https://doi.org/10.1007/978-1-0716-1831-8_14 (2022)
- Stillman, N. R. & Mayor, R. Generative models of morphogenesis in developmental biology. *Semin. Cell Dev. Biol.* <https://doi.org/10.1016/j.semcdb.2023.02.001> (2023)
- Cotterell, J. & Sharpe, J. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Mol. Syst. Biol.* **6**, 425 (2010).
- Schaerli, Y. et al. A unified design space of synthetic stripe-forming networks. *Nat. Commun.* **5**, 4905 (2014).
- Delgado, F. M. & Gómez-Vela, F. Computational methods for Gene Regulatory Networks reconstruction and analysis: a review. *Artif. Intell. Med.* **95**, 133–145 (2019).
- Zhou, X. & Cai, X. Inference of differential gene regulatory networks based on gene expression and genetic perturbation data. *Bioinformatics* **36**, 197–204 (2020).
- Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
- Yin, L., Huang, C.-H. & Ni, J. Clustering of gene expression data: performance and similarity analysis. *BMC Bioinf.* **7**, S19 (2006).
- Faith, J. J. et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
- Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinf.* **7**, S7 (2006).
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776 (2010).
- Petralia, F., Wang, P., Yang, J. & Tu, Z. Integrative random forest for gene regulatory network inference. *Bioinformatics* **31**, i197–i205 (2015).
- Meyer, P. E., Kontos, K., Lafitte, F. & Bontempi, G. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP J. Bioinf. Syst. Biol.* **2007**, 1–9 (2007).
- Hauray, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. TIGRESS: trustful inference of gene REgulation using stability selection. *BMC Syst. Biol.* **6**, 145 (2012).
- Razaghi-Moghadam, Z. & Nikoloski, Z. Supervised learning of gene-regulatory networks based on graph distance profiles of transcriptomics data. *npj Syst. Biol. Appl.* **6**, 21 (2020).
- Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J. & Ragan, M. A. Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief. Bioinf.* **15**, 195–211 (2014).
- Mordelet, F. & Vert, J.-P. SIRENE: supervised inference of regulatory networks. *Bioinformatics* **24**, i76–i82 (2008).
- Gillani, Z., Akash, M. S. H., Rahaman, M. M. & Chen, M. CompareSVM: supervised, Support Vector Machine (SVM) inference of gene regularity networks. *BMC Bioinf.* **15**, 395 (2014).
- Yang, Y., Fang, Q. & Shen, H.-B. Predicting gene regulatory interactions based on spatial gene expression data and deep learning. *PLoS Comput. Biol.* **15**, e1007324 (2019).
- Wu, S. et al. Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. *Proc. Natl Acad. Sci. USA* **113**, 4290–4295 (2016).
- Durant, F., Lobo, D., Hammelman, J. & Levin, M. Physiological controls of large-scale patterning in planarian regeneration: a molecular and computational perspective on growth and form. *Regeneration* **3**, 78–102 (2016).
- Eskandari, M. & Kuhl, E. Systems biology and mechanics of growth. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **7**, 401–412 (2015).
- Marcon, L. & Sharpe, J. Turing patterns in development: what about the horse part? *Curr. Opin. Genet. Dev.* **22**, 578–584 (2012).
- Holland, J. H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to *Biology, Control, and Artificial Intelligence*. (Michigan Univ. Press, 1975).

43. Mousavi, R. & Eftekhari, M. A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches. *Appl. Soft Comput.* **37**, 652–666 (2015).
44. Mousavi, R., Eftekhari, M. & Haghighi, M. G. A new approach to human microRNA target prediction using ensemble pruning and rotation forest. *J. Bioinf. Comput. Biol.* **13**, 1550017 (2015).
45. Mousavi, R., Eftekhari, M. & Rahdari, F. Omni-ensemble learning (OEL): utilizing over-bagging, static and dynamic ensemble selection approaches for software defect prediction. *Int. J. Artif. Intelligence Tools* **27**, 1850024 (2018).
46. Reali, F., Priami, C. & Marchetti, L. Optimization algorithms for computational systems biology. *Front. Appl. Math. Stat.* **3**, (2017).
47. Jaeger, J. et al. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics* **167**, 1721–1737 (2004).
48. Jaeger, J. et al. Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371 (2004).
49. Verd, B., Crombach, A. & Jaeger, J. Dynamic maternal gradients control timing and shift-rates for *drosophila* gap gene expression. *PLOS Comput. Biol.* **13**, e1005285 (2017).
50. Francois, P. & Siggia, E. D. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* **137**, 2385–2395 (2010).
51. Henry, A., Hemery, M. & François, P. φ -evo: a program to evolve phenotypic models of biological networks. *PLoS Comput. Biol.* **14**, e1006244 (2018).
52. Lobo, D. & Levin, M. Inferring regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. *PLoS Comput. Biol.* **11**, e1004295 (2015).
53. Lobo, D., Morokuma, J. & Levin, M. Computational discovery and in vivo validation of *hnf4* as a regulatory gene in planarian regeneration. *Bioinformatics* **32**, 2681–2685 (2016).
54. Lobo, D. & Levin, M. Computing a Worm: Reverse-Engineering Planarian Regeneration. in *Advances in Unconventional Computing. 2: Prototypes, Models and Algorithms* (ed. Adamatzky, A.) 637–654 (Springer International Publishing, Switzerland). https://doi.org/10.1007/978-3-319-33921-4_24 (2017)
55. Noman, N., Palafox, L. & Iba, H. Evolving genetic networks for synthetic biology. *New Gener. Comput.* **31**, 71–88 (2013).
56. Francois, P. & Hakim, V. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl Acad. Sci. USA* **101**, 580–585 (2004).
57. Rodrigo, G., Carrera, J. & Jaramillo, A. Genetdes: automatic design of transcriptional networks. *Bioinformatics* **23**, 1857–1858 (2007).
58. Smith, R. W., van Sluijs, B. & Fleck, C. Designing synthetic networks in silico: a generalised evolutionary algorithm approach. *BMC Syst. Biol.* **11**, 118 (2017).
59. Otero-Muras, I., Henriques, D. & Banga, J. R. SYNBADm: a tool for optimization-based automated design of synthetic gene circuits. *Bioinformatics* **32**, 3360–3362 (2016).
60. Otero-Muras, I. & Banga, J. R. Multicriteria global optimization for biocircuit design. *BMC Syst. Biol.* **8**, 113 (2014).
61. Sequeiros, C., Vázquez, C., Banga, J. R. & Otero-Muras, I. Automated design of synthetic gene circuits in the presence of molecular noise. *ACS Synth. Biol.* **12**, 2865–2876 (2023).
62. Otero-Muras, I. & Banga, J. R. Automated design framework for synthetic biology exploiting pareto optimality. *ACS Synth. Biol.* **6**, 1180–1193 (2017).
63. Mousavi, R., Konuru, S. H. & Lobo, D. Inference of dynamic spatial GRN models with multi-GPU evolutionary computation. *Brief. Bioinf.* <https://doi.org/10.1093/bib/bbab104> (2021)
64. Stapornwongkul, K. S. & Vincent, J.-P. Generation of extracellular morphogen gradients: the case for diffusion. *Nat. Rev. Genet.* **22**, 393–411 (2021).
65. Tkačik, G. & Gregor, T. The many bits of positional information. *Development* **148**, dev176065 (2021).
66. Grant, P. K. et al. Orthogonal intercellular signaling for programmed spatial behavior. *Mol. Syst. Biol.* **12**, 849 (2016).
67. Baig, H. et al. Synthetic biology open language visual (SBOL visual) version 3.0. *J. Integr. Bioinf.* **18**, 20210013 (2021).
68. Perkins, T. J., Jaeger, J., Reinitz, J. & Glass, L. Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput. Biol.* **2**, e51 (2006).
69. Verd, B., Monk, N. A. & Jaeger, J. Modularity, criticality, and evolvability of a developmental gene regulatory network. *Elife* **8**, e42832 (2019).
70. Andreas, E., Cummins, B. & Gedeon, T. Quantifying robustness of the gap gene network. *J. Theor. Biol.* **580**, 111720 (2024).
71. Jaeger, J. Shift happens: The developmental and evolutionary dynamics of the gap gene system. *Curr. Opin. Syst. Biol.* <https://doi.org/10.1016/j.coisb.2018.08.004> (2018)
72. Dickmann, J. E. M., Rink, J. C. & Jülicher, F. Long-range morphogen gradient formation by cell-to-cell signal propagation. *Phys. Biol.* **19**, 066001 (2022).
73. Oliver Huidobro, M., Tica, J., Wachter, G. K. A. & Isalan, M. Synthetic spatial patterning in bacteria: advances based on novel diffusible signals. *Microbial. Biotechnol.* **15**, 1685–1694 (2022).
74. Vittadello, S. T., Leyshon, T., Schnoerr, D. & Stumpf, M. P. H. Turing pattern design principles and their robustness. *Phil. Trans. R. Soc. A* **379**, 20200272 (2021).
75. Marcon, L., Diego, X., Sharpe, J. & Müller, P. High-throughput mathematical analysis identifies Turing networks for patterning with equally diffusing signals. *Elife* **5**, e14022 (2016).
76. Sohka, T., Heins, R. A. & Ostermeier, M. Morphogen-defined patterning of *Escherichia coli* enabled by an externally tunable band-pass filter. *J. Biol. Eng.* **3**, 10 (2009).
77. Bird, J. E., Marles-Wright, J. & Giachino, A. A user's guide to golden gate cloning methods and standards. *ACS Synth. Biol.* **11**, 3551–3563 (2022).
78. Martínez-García, E. et al. SEVA 4.0: an update of the Standard European Vector Architecture database for advanced analysis and programming of bacterial phenotypes. *Nucleic Acids Res.* **51**, D1558–D1567 (2023).
79. Scholes, N. S., Schnoerr, D., Isalan, M. & Stumpf, M. P. H. A comprehensive network Atlas reveals that turing patterns are common but not robust. *Cell Syst.* **9**, 243–257.e4 (2019).
80. Liu, H. L., Chen, L., Deb, K. & Goodman, E. D. Investigating the effect of imbalance between convergence and diversity in evolutionary multiobjective algorithms. *IEEE Trans. Evolut. Comput.* **21**, 408–425 (2017).
81. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
82. Jones, T. S., Oliveira, S. M. D., Myers, C. J., Voigt, C. A. & Densmore, D. Genetic circuit design automation with Cello 2.0. *Nat Protoc* **17**, 1097–1113 (2022).
83. Hwang, J., Hari, A., Cheng, R., Gardner, J. G. & Lobo, D. Kinetic modeling of microbial growth, enzyme activity, and gene deletions: An integrated model of β -glucosidase function in *Cellvibrio japonicus*. *Biotechnol. Bioeng.* **117**, 3876–3890 (2020).
84. Davies, J. & Levin, M. Synthetic morphology with agential materials. *Nat. Rev. Bioeng.* **1**, 46–59 (2023).
85. Ko, J. M. & Lobo, D. Continuous Dynamic Modeling of Regulated Cell Adhesion: Sorting, Intercalation, and Involution. *Biophys. J.* **117**, 2166–2179 (2019).
86. Herath, S. & Lobo, D. Cross-inhibition of Turing patterns explains the self-organized regulatory mechanism of planarian fission. *J. Theor. Biol.* **485**, 110042 (2020).

87. Lobo, D. Formalizing Phenotypes of Regeneration. in *Whole-Body Regeneration: Methods and Protocols* (eds. Blanchoud, S. & Galliot, B.) 663–679 (Springer US, New York, NY). https://doi.org/10.1007/978-1-0716-2172-1_36 (2022)
88. Roy, J., Cheung, E., Bhatti, J., Muneem, A. & Lobo, D. Curation and annotation of planarian gene expression patterns with segmented reference morphologies. *Bioinformatics* **36**, 2881–2887 (2020).
89. Whitley, D., Rana, S. & Heckendorn, R. B. The island model genetic algorithm: on separability, population size and convergence. *J. Comput. Inf. Technol.* **7**, 33–47 (1999).
90. Luke, S. & Panait, L. A comparison of bloat control methods for genetic programming. *Evolut. Comput.* **14**, 309–344 (2006).
91. Mahfoud, S. W. Crowding and preselection revisited. in *Parallel Problem Solving from Nature 2* (eds. Manner, R. & Manderick, B.) 27–36 (Elsevier, 1992).
92. Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. *Numerical Recipes*. (Cambridge University Press, New York, 1986).

Acknowledgements

We thank the members of the Lobo Lab for helpful discussions. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM137953. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Computations used the UMBC High Performance Computing Facility (HPCF) supported by the NSF MRI program grants CNS-1920079 and OAC-1726023.

Author contributions

Both authors designed and implemented the methods and wrote the manuscript. RM produced and analyzed the data. DL secured funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-024-00361-5>.

Correspondence and requests for materials should be addressed to Daniel Lobo.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024