# Supplemental Analysis

Supplementary Analysis for "Cellular Dialogs: cell-cell communication through diffusible molecules yields dynamic spatial patterns"

Yiteng Dang<sup>1,2</sup>, Douwe A.J. Grundel<sup>1,2</sup> and Hyun Youk<sup>1,2,3,\*</sup>

<sup>1</sup>Kavli Institute of Nanoscience,

<sup>2</sup>Department of Bionanoscience, Delft University of Technology, 2629HZ Delft, The Netherlands <sup>3</sup>CIFAR, CIFAR Azrieli Global Scholars Program, Toronto M5G 1M1 ON, Canada \*Correspondence to: h.youk@tudelft.nl

# Contents

| S1 Overview of self-organized patterns  | <b>2</b> |
|---|----------|
| S1.1 Static patterns  | . 2      |
| S1.2 Dynamic temporal patterns  | . 2      |
| S1.3 Dynamic spatial patterns   | . 4      |
| S2 Parameter-derived general constraints on the dynamics  | 6        |
| S2.1 Dynamical phases for each interaction  | 6        |
| S2.2 State diagrams $\ldots \ldots \ldots$          | . 8      |
| S2.3 Simplified dynamics  | . 10     |
| S2.4 Algorithm for computing state diagrams   | . 11     |
| S3 Analytic framework for traveling wave propagation  | 15       |
| S3.1 Features of traveling waves  | 15       |
| S3.2 Abundance of traveling waves   | 16       |
| S3.3 Traveling wave propagation conditions $\ldots \ldots \ldots$               | 20       |
| S3.3.1 General conditions for pattern propagation $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$   | 20       |
| S3.3.2 Straight and bent waves  | 21       |
| S3.3.3 Structure of traveling waves   | . 22     |
| S3.3.4 Traveling wave propagation conditions  | 24       |
| S3.3.5 Nearest-neighbor approximation $\ldots \ldots \ldots$   | 24       |
| S3.4 Explicit example   | 25       |
| S3.5 Performance of the analytic framework  | 28       |
| S3.6 Oscillatory traveling waves  | 31       |
| S4 Robustness and reliability of traveling waves  | 32       |
| S4.1 Robustness (Fig. S6A-C) $\ldots \ldots \ldots$ | 32       |
| S4.2 Robustness: analytical calculation (Fig. S9) $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$                                    | . 33     |
| S4.3 Reliability (Figs. 5F and S8) $\ldots$   | 38       |
| S4.4 Influence of initial conditions (Fig. S6D)   | 39       |
| S4.5 Stability of TWs (Fig. S14) $\ldots$   | 40       |
| S4.6 Persistence of traveling waves with noise (Fig. S12)   | 41       |
| References  | 43       |

# S1 Overview of self-organized patterns

In this section, we provide a detailed but mostly qualitative overview of the different types of self-organized patterns we observed in simulations of our model. The aim is give a general overview of the different possible morphologies and dynamic features of these patterns, and to understand basic features of these patterns in terms of the concepts we have introduced or will introduce. In section S3, we will analyze traveling waves - a subset of dynamic spatial patterns - more closely.

# S1.1 Static patterns

While the focus of this work is mainly on dynamic patterns, we also observed static patterns with a high degree of spatial organization in all two-gene networks studied. They commonly arise after a relatively short transient phase (10s-100s of time steps) and can have different shapes and compositions of cell states (i.e. the cells can have different combinations of gene expression). In terms of shape, most patterns consist of one or more islands or stripes of cells with a different cell state from the surrounding cells. Since there is no Turing mechanism in our system, we did not identify a natural length scale for the patterns, although changing parameters did seem to affect the size of structures such as islands. Patterns were most commonly observed to have two sets of cell states, where one group of cells has one cell state and the other group has another cell state. Patterns with three cell states are rare, but not impossible to generate. We did not observe any patterns where all four cell states existed concurrently. The most common static patterns are ones that also arise in the model with one signaling molecule and consist of one group of cells with a given gene ON and another group with that gene OFF. In the case of two molecules, it is common to find islands with both genes ON (or OFF) with the rest of the system consisting of cells with both genes OFF (or ON). We also observed patterns where the two genes were mutually exclusive, i.e. if a cell has gene 1 ON, it has gene 2 OFF, and vice versa. Finally, we occasionally found a boundary layer separating an region with similar gene expression (e.g. island or stripe) from the rest of the cells with a different cell state than either the region or the rest of the cells.

# S1.2 Dynamic temporal patterns

Dynamic temporal patterns are periodic steady states where the system returns to an earlier state after a finite number of time steps  $\tau > 1$  (the period of the oscillation), but do not propagate information across space.

**Single-cell oscillations** Oscillations can arise at the single-cell level in the case of one gene with a negative feedback loop. If certain parameter constraints are satisfied, the gene expression level of a single cell oscillates between ON and OFF indefinitely. These oscillations are the result of our adiabatic description, where we assume that cells respond slowly compared to the time for signaling molecule concentrations to reach a steady state. An ON-cell turns OFF because the concentration it senses is high enough to suddenly switch to the other state. The OFF-cell

then senses a low concentration and the cell switches to the other extreme immediately, without ever reaching the intermediate steady state.

With two genes, oscillations at the single cell level remain relatively simple and can only have periods up to four, since there are only four cell states with two genes. In practice, by examining all possible single-cell state diagrams (Sec. S2), we found that the vast majority of single-cell oscillations were of period 2 (see Fig. S4C). Single-cell period 2 oscillations arise in all networks that can generate dynamic patterns (temporal or spatial), while period 4 oscillations arise only in networks with an incoherent mutual feedback (i.e., for all networks generating dynamic spatial patterns in Fig. 3D as well as Network 14 in Fig. 3C).



Table 1: Two gene network motifs generating oscillations. The three core topologies for mutual interaction between the two genes are shown together with typical single-cell state diagrams showing oscillations. The state diagrams are for the case when the concentration of the regulator genes always surpass the threshold when the gene is ON and is below the threshold when it is OFF, i.e.  $C_{ON}^{(j)} > K^{(ij)} > C_{OFF}^{(j)}$  for all genes i, j.

We can interpret these results by looking at the three core network structure that give rise to most dynamic patterns (Table 1). For each of the motifs, the interpretation of the oscillations is straightforward. For mutual repression (Table 1 - top row), a cell is able to oscillate between (0,0) and (1,1), whereas (0,1) and (1,0) are stable states. When both genes are off, both are unrepressed and will turn on the next time step, after which they are both repressed and turn off again. However, if only one of the genes is on, it represses the other gene but is not repressed itself. For mutual activation (Table 1 - middle row), the oscillation is between the states (0,1) and (1,0). Each gene can turn on the other, but turns off when the other gene is on. However, if both genes are ON, they sustained each other, whereas if neither is ON, they also cannot turn ON. Finally, for a positive-negative loop (Table 1 - bottom row), the system undergoes a period 4 oscillation between the four states.

chosen, but it is intuitive that they should be possible for some set of parameters. Again, these results rely on the separation of time scales between the relaxation of the signaling molecule concentrations and the response of the cells.

**Synchronization of single-cell oscillations** Oscillations persist on a multicellular level, with cells synchronizing their oscillations depending on how strongly the cells interact. The degree of synchronization is reflected in the proportion of cells that oscillates together, as well as how they are spatially arranged (e.g. we would consider an oscillating island more synchronized than randomly distributed cells that oscillate in sync). Generally, the oscillations in the multicellular system can vary between completely autonomous (i.e., each cell independently oscillates) to completely synchronized (i.e., all cells oscillate synchronously; see Fig. 2I). Full autonomy is reached if and only if each of the interactions is in the autonomous (A01) phase (see Sec. S2). Full synchronization can be reached for a variety of other parameter conditions. In between, the system can partially synchronize and exhibit domains of cells oscillating together that do not extend over the entire lattice. Oscillatory cells can also co-exist with stationary cells that are in one of the stationary states (for negative-negative or positive-positive feedback).

**Complex dynamic temporal patterns** We distinguish between "simple" oscillations, which are superpositions of single-cell oscillations and thus have a period of at most four, and more complex oscillations, which we will describe here. Oscillations of a more complicated form arise in the networks that are capable of generating dynamic spatial patterns. Each of these networks produces a wide range of periods with  $\tau \geq 5$  (Fig. S4D), most of which correspond to dynamic temporal patterns. Typically, they feature oscillating domains that coexist with a background of static cells (e.g., the oscillating island in Fig. 2H), but where different cells in the domain undergo different cycles of gene expression states over time. This can give rise to complicated temporal patterns because the oscillatory sequences of individual cells may not line up, especially when they are incommensurable.

These complex periods are indeed associated with dynamic patterns, as we can verify by measuring their degree of spatial order using the spatial index  $I^{(i)}$ . Overall, steady states with a period  $\tau \geq 5$  tend to have higher values of  $I^{(i)}$  for both genes (yellow and purple lines in Fig. S4E), indicating that they tend to be more spatially ordered than oscillations with simple periods (red lines in Fig. S4E) and static patterns (blue lines in Fig. S4E).

#### S1.3 Dynamic spatial patterns

Dynamic spatial patterns are characterized by gene expression profiles that translate across space, thereby allowing propagation of information across the multicellular system. These can be rigid profiles of gene expression that move across the system without changing shape, but we also count patterns that move and morph (i.e. change shape) at the same time as dynamic spatial patterns. Note that these patterns require periodic boundary conditions to be sustained indefinitely.

5

**Traveling waves** Traveling waves are characterized by stripes of cells that translate across the lattice in a regular fashion (see Fig. 2A, 2C-E and 2G for examples). They typically consist of three types of cells (with different states) and travel on a background consisting of cells of the fourth type. When two traveling waves in opposite directions collide, they typically annihilate each other, leaving a void of cells with the background state. Characteristics of traveling waves and their propagation conditions will be discussed in subsection S3.

**Complex wavelets** In a number of cases, we observed complex wavelets that propagate indefinitely without repeating themselves, within the maximum simulation time (Fig. 2B). Since the total number of system states is limited to  $2^N$ , these waves will eventually settle down to a steady state. The transient wave patterns they generate look very similar to (less coordinated) traveling waves, and arise as transient states during the generation of all types of dynamic spatial patterns described here.

**Spiral and concentric waves** Spiral and concentric waves are similar to traveling waves with the main difference that their orientation is outward from a source or center rather than linear in a fixed direction (Fig. 2F). Locally, they typically look like traveling waves, with the same set of cell states as in traveling waves. Due to annihilation of colliding waves, spiral and concentric waves are less stable, since only particular configurations where the outcome of the collision is an earlier spiral wave pattern will be observed as persistent spiral waves. It is more common to observe spirals and concentric waves as transient patterns that are created and annihilated repeatedly, until the system settles down to a more stable configuration such as a traveling wave.

**Traveling pulses** We also observe small, localized patterns of a few cells that translate across the lattice in a regular way. They are similar to traveling waves, but the traveling pulses are small, localized patterns that do not span the entire size of the system.

**Oscillatory traveling waves** In networks 16, 20 and 43 - characterized by the incoherent mutual feedback motif without positive self-regulations (Fig. 3D) - we found oscillatory traveling waves where both the wave states and the background state oscillate over time (Examples in Fig. S3). At any given fixed time, these waves typically look similar to the non-oscillatory traveling wave, but due to the oscillations the dynamics is different. Perfectly aligned waves where each wave state occupies a single band of cells are relatively rare. Most waves have bands that occupy the width of more than one cell (see for instance Fig. S3B). The waves undergo a successive sequence of static oscillations followed by an translation (Fig. S3D). Details of their dynamics are further discussed in Sec. S3.6.

6

# S2 Parameter-derived general constraints on the dynamics

In this section, we derive a number of a methods to derive general constraints on the dynamics of the system from the parameters of that system. First, we introduce the concept of dynamical phase for each regulatory interaction between two (possibly identical) signaling molecules. Next, we introduce the concept of state diagram - a graphical way to represent transitions between cell states - and discuss their usefulness in deducing constraints on the system's dynamics. We then derive general constraints on the dynamics of a system with multiple signaling molecules, which arise as special combinations of these phases. Finally, we present a formal algorithm to calculate the dynamical constraints and represent them in a state diagram for any set of arbitrary parameters.

# S2.1 Dynamical phases for each interaction

The idea behind dynamical phases is based on the observation that for extreme parameter values, the behavior of the system becomes predictable. For instance, if the interaction between cells is very strong and the threshold values characterizing their response are very low, then we expect the cells to always exceed these thresholds regardless of the precise states of the cells. These ideas were made precise in our previous work for systems with one signaling molecule, which represented these dynamical constraints as "phenotype functions" on a "phenotype diagram" [1]. In this work, we extend this formalism to a more general framework applicable to multiple interactions.

Consider an interaction between two genes where a regulating gene j controls the expression of the regulated gene i (possibly i = j). The interaction is specified by the threshold  $K^{(ij)}$ for turning the gene ON/OFF (depending on whether the interaction is activating or repressive), and the ON-secretion rate  $C_{ON}^{(j)}$ . Suppose the cells have an effective distance  $a_0$  to their nearest neighbors, and lead to an interaction strength  $f_N^{(j)}(a_0)$ . Recall that the outcome of the interaction is specified by  $g^{(ij)}(\mathbf{X})$ . We then distinguish the following phases:

- 1. **P1**: sensed concentration permanently above threshold. The phase is defined by  $(1 + f_N^{(j)})C_{OFF}^{(j)} > K^{(ij)}$ . For an activating interaction, this implies that  $g^{(ij)}(\mathbf{X}) = 1$  for any system state  $\mathbf{X}$ . The interaction is always ON. For a repressive interaction, we have  $g^{(ij)}(\mathbf{X}) = 0$  and the interaction is always OFF. For a single activating interaction, this corresponds to the *all ON phase* all cells in the system turn ON in one time step and remain ON.
- 2. **P0**: sensed concentration permanently below threshold. The phase is defined by  $(1 + f_N^{(j)})C_{ON}^{(j)} < K^{(ij)}$ . For an activating interaction this implies  $g^{(ij)}(\mathbf{X}) = 0$  and for a repressive interaction  $g^{(ij)}(\mathbf{X}) = 1$ . For a single activating interaction, this corresponds to the *all OFF phase* all cells in the system turn OFF in one time step and remain OFF.

- 3. A1: autonomy whenever the regulating gene is ON. The phase is defined by  $C_{ON}^{(j)} + f_N^{(j)}C_{OFF}^{(j)} > K^{(ij)}$ . For an activating interaction, this implies that  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 1$ . This means that the interaction is always ON in a cell k whenever gene j is ON regardless of the rest of the cells. However, when gene j is OFF, whether the interaction will be ON depends on the state of other cells. For repression,  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 1$ . In this case, the interaction is always OFF in cell k whenever it has gene j ON. For a single activating interaction, this corresponds to the activation phase, whereby ON cells remain ON regardless of the rest of the system.
- 4. A0: autonomy whenever the regulating gene is OFF. The phase is defined by  $C_{OFF}^{(j)} + f_N^{(j)} C_{ON}^{(j)} < K^{(ij)}$ . This is analogous to the previous case with some roles switches. For activation, we get  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 0$ . For repression,  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 0$ . For a single activating interaction, this corresponds to the *deactivation phase*, whereby OFF cells remain OFF regardless of the rest of the system.
- 5. A01: autonomy regardless of whether the regulating gene is ON or OFF. This phase is defined by parameter values for which both inequalities of A1 and A0 hold. These conditions can only be met simultaneously if  $f_N^{(j)} < 1$ . For activation, it implies that  $g_k^{(ij)}(\mathbf{X}) = X_k^{(j)}$ . More explicitly, it means that  $g_k^{(ij)}(\mathbf{X}) = 1$  whenever  $X_k^{(j)} = 1$  and  $g_k^{(ij)}(\mathbf{X}) = 0$  whenever  $X_k^{(j)} = 0$ . Hence  $X_k^{(j)}$  fully determines fate of the interaction. For repression, the roles are reversed and  $g_k^{(ij)}(\mathbf{X}) = 1 - X_k^{(j)}$ . For a single activating interaction, this corresponds to the *autonomy phase* - ON cells remain ON and OFF cells remain OFF.
- 6. U: unconstrained. This phase is defined by parameter values for which the conditions of neither A0 nor A1 are true. Hence, we have  $C_{ON}^{(j)} + f_N^{(j)}C_{OFF}^{(j)} < K^{(ij)}$  and  $C_{OFF}^{(j)} + f_N^{(j)}C_{ON}^{(j)} > K^{(ij)}$ . These conditions can only be met simultaneously if  $f_N^{(j)} > 1$ . In this phase, we cannot deduce any general constraints on  $g_k^{(ij)}(\mathbf{X})$  and have to look at the specific system state  $\mathbf{X}$  to determine whether an interaction will be ON or OFF. For a single activating interaction, this corresponds to the *activation-deactivation phase*, whereby cells can both activate (turn ON) as well as deactivate (turn OFF) other cells.

The interpretation of these phases are best understood for a system with only one signaling molecule. For now, note that the phases **P0** and **P1** make the interaction trivial - the outcome is always known regardless of the state of the cell itself or its neighbors. The phases **A0**, **A1** and **A01** place constraints which are dependent on the current state of the cell, and the phase **U** does not place any constraints on the system's dynamics. For a system with multiple signaling molecules, each interaction will be characterized by one phase. The obvious next question is how to put together the constraints from these different interactions to derive general constraints on the system's dynamics.

For two molecules, we can represent the phase regions and phases of each interaction together on a "phenotype diagram" consisting of two plots, as shown in Fig. S2C. In this example, each of the interactions lies in a colored region representing the phase. Specifically, the activation  $1 \leftarrow 2$  is in the A0 phase (lower figure, circle with "1"). As this is the only regulatory interaction

8

on gene 1, we now immediately deduce features of the dynamics of gene 1. Specifically, we now know that whenever gene 2 is OFF in a cell, gene 1 will always be inactivated (OFF) the next time step. Similarly, the repression  $1 \dashv 2$  is in the A01 phase, meaning that if gene 1 is ON, then gene 2 will always be repressed. However, if gene 1 is OFF, then it does not repress gene 2, but whether the latter turns on depends also on the result of its own interaction. More generally, we can combine all these phase constraints into general constraints on the dynamics of the system. In the next subsection, we consider limits where the phase constraints completely constrain each cell's dynamics.

## S2.2 State diagrams

The basic idea of the state diagram is that it displays all the possible transitions between different cell states of a system. The concept of state diagram has been explored in earlier work on modeling genetic circuits with binary expression states [2], but has been limited to models of gene circuits at the single-cell level. In our case, the cell states are the binary states  $X = (X^{(1)}, \ldots, X^{(l)}) \in \{0, 1\}^l$  specifying for a given cell whether each gene is ON or OFF. For one signaling molecule, there are only two cell states, 0 and 1. For two signaling molecules, we distinguish the four cell states (0, 0), (0, 1), (1, 0) and (1, 1). If a cell in a given cell state X can adopt the state Y after one time step - either under its own influence or through sensing molecules secreted by other cells - then we draw an arrow between the states X and Y.

For a single interaction, the procedure is straightforward and the diagrams are simple to interpret (see Fig. S2A). For instance, the P1 phase with activation means that cells always turn ON after one time step, hence giving the diagram with all arrows going to the 1 state and no other arrows. For A01 with an activating interaction, the cells are autonomous, so 0 remains 0 and 1 remains 1. However, A01 with repressive interaction gives oscillations between 0 and 1. An ON-cell always turns OFF because it will always sense a concentration above the threshold, repressing its gene expression, while an OFF-cell will always turn ON because its sensed concentration will always be below the threshold, leaving the gene unrepressed. The collective set of possible transitions is then displayed as a diagram with the cell states and possible transitions.

For two genes, the interpretation is analogous. We draw a directed graph with the four cell states (0,0), (0,1), (1,0) and (1,1) as nodes and directed edges between these states to indicate (possible) transitions between the states. In Fig. 4E, in the left diagram for dialogue 14, the transitions are heavily constrained. Each of the four states leads only to one other possible state. This completely constrains the dynamics of the system, so that it becomes completely predictable - any cell's dynamics in the system is completely determined by the transitions in one diagram. As such, this state diagram does not allow for multicellular pattern formation, as all individual cells will oscillate individually. In contrast, if we take a diagram such as the one depicted for dialogue 15 (Fig. 4E), the dynamics of the system is not entirely constrained. The state (0,0) has two arrows leaving from it, indicating that either transition is possible.

9

depending on the exact concentration a cell senses. As such, it allows for a pattern such as a traveling wave to propagate, because cells of the same state do not always evolve in the same way, but evolve depending on the other cells in the system. In practice, the state diagram of dialogue 14 (Fig. 4E - left figure) can be realized under a diverse range of parameter conditions, whereas the diagrams for dialogues 15 and 19 (Fig. 4E - right figure) are never realized without additional transitions that we have not drawn in the diagrams here. Note that without additional knowledge, traveling waves are in principle compatible with all diagrams containing the diagrams in Fig. 4E - right figure as subdiagram. However, the wave can propagate if and only if the particular wave states that we identified (Fig. 4 and section S3.3.3) follow the transitions depicted in the figure. This is not information contained in the diagrams merely represent the set of possible cell state transitions given knowledge of the parameter of the system.

Two properties of the system are immediately evident from the graphical representation of the state diagram. To begin with, a state diagram tells us which cell states could potentially be stationary. Such states must have an arrow to themselves in the state diagram, which we call a *self-transition*. Should the system reach a non-oscillatory steady state, then that final state can only be composed of cell states which have a self-transition. If there are no self-transitions, then the system cannot generate stationary patterns - this happens for instance in dialogue 14 with certain parameters, which can produce the state diagram shown in Fig. 4E. If there is only one self-transition, then the only possible stationary steady state is a uniform system where all cells have the state with the self-transition. Conversely, not all cell states with a self-transition need to appear in a stationary system state. In other words, having a self-transition does not imply that the state appears in any stationary pattern. As an extreme example, the system could have a fully connected state diagram, where each transition between two cell states is in principle possible. However, this system could still generate a uniform lattice of cells as final state if the parameters are chosen appropriately.

Secondly, the state diagrams show whether periodic steady states (e.g. oscillations) are possible. For any periodic steady state, all cells must revisit their earlier state after  $\tau > 1$  number of time steps, where  $\tau$  is the period of the oscillatory state. This implies that the state diagram should permit cells to return to their initial states after a finite number of time steps, and after passing through other states (otherwise it would be a stationary pattern). This is only possible if the state diagram contains cycles, i.e. closed loops obtained by tracing the edges of the graph from some initial state. The presence of cycles is thus a necessary condition for oscillations. However, it is not a sufficient condition for generating dynamic temporal patterns. This is because it is not guaranteed that a cell can traverse the edges of any cycle one by one when there are possible "routes" on the graph. Each transition then corresponds to a specific condition which depends on the state of all cells of the system. We cannot directly deduce whether a sequence of such transitions is possible at the level of the entire system of N cells. This is only possible in the special case that all transitions of the cycle are deterministic, i.e. when each node is connected to a unique other node on the graph. We then obtain an oscillation, with a

period equal to the length of the cycle.

In summary, the state diagram allows us to deduce two basic properties of our system without running simulations: the set of stationary cell states and the capacity of generating dynamic patterns. These are not purely mathematical properties but have biological relevance. The former is an indicator of multistability and tells us whether a population of identical cells could potentially diversify, generating stable configurations with multiple gene expression profiles. This is known as *phenotypic heterogeneity*, a phenomenon that has seen many experimental studies dedicated to it and may have a wide range of biological functions [3]. The latter tells us whether a multicellular system could potentially sustain oscillations, waves or other dynamic patterns (consult the main text for biological examples).

#### S2.3 Simplified dynamics

There are two limits in which the dynamics of the system simplifies dramatically.

- All interactions are either extremely weak or extremely strong. To be precise, this is the case if each of the interactions is in either the P0 or the P1 phase (all ON/all OFF phase). The system homogenizes after one time step, because each of the interactions is either ON or OFF for all cells in the system. For a spatially uniform system, the dynamics is simple and predictable.
- 2. All interactions are moderately strong, and the interaction between cells is relatively weak. In more precise terms, suppose all interactions are in the A01 phase. In this case the dynamics of each cell becomes equivalent to that of a single cell. The system is fully autonomous and each cell evolves under its own influence.

In these limits the state diagrams are identical to single-cell state diagrams with rescaled parameters. Therefore, these phases contain only deterministic state diagrams. Thus we know the exact dynamics of the system without running any simulations, for any initial conditions.

**Formal derivation** Let us show these two limits more explicitly. Let k be arbitrary, and (i, j) be an arbitrary pair of genes with  $M_{\text{int}}^{(ij)} \neq 0$ . Suppose this interaction is in the P0 phase. Then as a result of  $(1 + f_N^{(j)})C_{OFF}^{(j)} < K^{(ij)}$ , we have

$$g_{P0}^{(ij)} \equiv g_k^{(ij)}(\mathbf{X}) = \begin{cases} 0 & \text{if } M_{\text{int}}^{(ij)} = 1\\ 1 & \text{if } M_{\text{int}}^{(ij)} = -1 \end{cases}$$

That is to say,  $g_k^{(ij)}(\mathbf{X})$  becomes independent of both  $\mathbf{X}$  and k. Likewise, in the P1 phase we get  $g_k^{(ij)}(\mathbf{X}) \equiv g_{P1}^{(ij)}$ , with  $g_{P1}^{(ij)} = 1 - g_{P0}^{(ij)}$ . Now suppose all interactions are either in the P0 or P1 phase. Then,  $X_k^{(j)}(t+1) = \prod_j g_k^{(ij)}(\mathbf{X}) = \prod_j g_{c_j}^{(ij)}$  with  $c_j \in \{P0, P1\}$ . Hence,  $X_k^{(j)}(t+1)$  is also independent of both X and k. Therefore, all cells become identical after one time step. For an identical lattice with all cells in a state X, we note that each cell senses a concentration  $Y^{(j)} = (1 + f_N^{(j)})C^{(j)}(X^{(j)})$ . However,  $K^{(ij)}$  and other parameters are unchanged. Therefore,

the evolution of a uniform lattice is equivalent to that of a single cell with a rescaled secretion rate  $C^{(j)}(X) \to (1 + f_N^{(j)})C^{(j)}(X)$ .

Next, consider the case that all interactions are in the A01 phase. Again, let k and (i, j) be arbitrary with  $M_{\text{int}}^{(ij)} \neq 0$ . Let **X** be an arbitrary state of the system, and write  $\mathbf{X} = (X_k, Z_k)$ , with  $Z_k = \{X_l\}_{l \neq k}$ . Then the A01 phase puts the following constraints on the system:

$$Y^{(j)}(X_k = 1, Z_k) \ge C_{ON}^{(j)} + f_N^{(j)} C_{OFF}^{(j)} > K^{(ij)},$$
  

$$Y^{(j)}(X_k = 0, Z_k) \le C_{OFF}^{(j)} + f_N^{(j)} C_{ON}^{(j)} < K^{(ij)}.$$
(S1)

We see that any cell with gene j ON will always satisfy the first constraint, regardless of the rest of the system. Likewise, any cell with gene j OFF will always satisfy the second constraint. As a result,  $g^{(ij)}(X_k; Z_k) = g^{(ij)}(X_k)$  becomes independent of  $Z_k$ , the states of all cells other than k in the system. Therefore,  $X_k^{(i)}(t+1) = \prod_j g^{(ij)}(X_k(t); Z_k(t)) = g^{(ij)}(X_k(t))$  depends only on  $X_k(t)$ . In other words, the evolution of any cell in the system is independent of the state of the other cells.

#### S2.4 Algorithm for computing state diagrams

In this subsection, we present a general method for computing the state diagram for a system of one or two genes, given an arbitrary set of system parameters. The construction for two genes can be readily generalized to systems with more than two genes.

For a single gene, we state diagrams follow straightforwardly from the definition of the phases (section S2.1). The end result can be represented as a directed graph with two nodes (representing ON and OFF state of the gene) and up to four edges, which we can describe using its adjacency matrix

$$A = \begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix}.$$
 (S2)

The adjacency matrix gives information on whether edges are present for each potential link between two nodes. The entries  $A_{ij} \in \{0, 1\}$  are for transitions from state *i* to state *j*. If  $A_{ij} = 1$ , a transition  $i \to j$  is possible and we draw an edge between the two nodes. If  $A_{ij} = 0$ , *i* cannot transition to *j* and we draw no edge.

As an example, consider cells with a single signaling molecule with negative feedback to itself. The graphs for positive feedback are deduced in a similar way. In the P1 phase, the system is permanently repressed, so all states go to the 0 state. Hence  $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ . By contrast, in the P0 state, both ON and OFF cells always turn ON at the next time step, so  $A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ . In the A1 state, ON cells always turn OFF, but we do not know anything about the OFF cells. Hence both transitions  $0 \to 0$  and  $0 \to 1$  are possible. Therefore,  $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ . Conversely, in the A0 phase, only OFF cells are constrained to always turn on, so  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . In the A01 phase, OFF cells turn ON and ON cells turn OFF, so  $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ . Finally, in the U phase, all transitions are unconstrained, so  $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ .

For systems with two or more genes, the procedure of deriving state diagrams is considerably more involved. We first outline the intuitive idea behind this derivation and then provide a formal, mathematical derivation of the construction. With two mutually interacting signaling molecules, the dynamics of a gene i depends in general on both regulation by itself and regulation by the other gene, which we label j. If we know the phases of both regulations  $i \leftarrow i$  and  $i \leftarrow j$ , then we can deduce the constraints they impose on the dynamics of i. To do this, we have to combine the constraints from both regulatory interactions  $i \leftarrow i$  and  $i \leftarrow j$ , for which we employ a three-valued logic operation. Intuitively, this three-valued logic system represents the fact that there are three possible outcomes of each interaction: the regulated gene is activated, the regulated gene is repressed or the outcome is unknown. Hence, we need to know what the final response of gene i is for each combination of the three outcomes for both of the two regulatory interactions. For instance, suppose that both i and j positively regulate i (i.e.  $M_{\rm int}^{(ii)} = M_{\rm int}^{(ij)} = 1$ ), but the interaction  $i \leftarrow i$  is always activating (i.e., the sensed concentration of i always exceeds the threshold  $K^{(ii)}$  while the interaction  $i \leftarrow j$  is unknown. Then the final outcome for gene i is unknown, because both positive interactions must be activating for the gene to turn on. Next, recall that the phases in general place dynamical constraints that depend on the state of the system. Concretely, this means that the constraint placed by i on itself depends on whether it is ON or OFF, and the same holds for the constraint placed by gene i. Therefore, for each combination of states for genes i and j - there are four of these, corresponding to the four cell states (0,0), (0,1), (1,0) and (1,1) - we could have a different set of constraints on the dynamics for gene i. Thus, we have to separately consider each of the four cell states and see which constraints they impose on the dynamics of both gene i and gene *j*. In this way, for each cell state, we obtain all possible cell states to which it can transition to and draw the corresponding edges on the graph. This eventually gives us our state diagram.

Formal derivation of the algorithm for computing state diagrams First, we note that the four-node graph with up to 16 edges is now represented by a  $4 \times 4$  adjacency matrix, which we will denote

$$A = \begin{pmatrix} A_{(0,0)\to(0,0)} & A_{(0,0)\to(1,0)} & A_{(0,0)\to(0,1)} & A_{(0,0)\to(1,1)} \\ A_{(1,0)\to(0,0)} & A_{(1,0)\to(1,0)} & A_{(1,0)\to(0,1)} & A_{(1,0)\to(1,1)} \\ A_{(0,1)\to(0,0)} & A_{(0,1)\to(1,0)} & A_{(0,1)\to(0,1)} & A_{(0,1)\to(1,1)} \\ A_{(1,1)\to(0,0)} & A_{(1,1)\to(1,0)} & A_{(1,1)\to(0,1)} & A_{(1,1)\to(1,1)} \end{pmatrix}.$$
(S3)

The interpretation is the same: the state (i, j) can transition into the state (k, l) if and only if  $A_{(i,j)\to(k,l)} = 1$ , and we represent this graphically by drawing a directed edge between (i, j) and (k, l). Our goal is then to combine the constraints imposed by the different phases for each interaction to compute this adjacency matrix.

Recall that the time evolution for a cell determined by  $X^{(i)}(t+1) = \prod g^{ij}(\mathbf{X}(t))$ . This is a deterministic equation for  $X^{(i)}(t)$  when we know the precise input system state  $\mathbf{X}(t)$ . Now suppose we only know the cell's own state  $X = (X^{(1)}, X^{(2)})$  and the phase of each interaction  $i \leftarrow j$ . We want to calculate the set of *possible* output cell states for X(t+1). To do this, we introduce a set of three-valued logic states  $S = \{0, 1, 2\}$  and a logic AND function  $\wedge : S \to S$ 

defined by the truth table

$$\begin{array}{l} a \wedge b & b = 0 & b = 1 & b = 2 \\ a = 0 \\ a = 1 \\ a = 2 \end{array} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{pmatrix}$$
 (S4)

This 3-valued logic system is known as the *Kleene logic*. It has a third logic value UNKNOWN in addition to TRUE and FALSE. In our notation, 0=FALSE, 1=TRUE and 2=UNKNOWN. The UNKNOWN value can be interpreted a state that can be either TRUE or FALSE. When combined with a FALSE value, we know for sure that FALSE  $\land$  UNKNOWN = FALSE, since both FALSE  $\land$  FALSE = FALSE and FALSE  $\land$  TRUE = FALSE. However, TRUE  $\land$  FALSE = FALSE and FALSE  $\land$  TRUE = FALSE. However, TRUE  $\land$  FALSE = FALSE and therefore TRUE  $\land$  UNKNOWN = UNKNOWN.

We employ the three-valued logic system as follows: whenever a cell state has uncertain transitions (i.e. can transition to multiple output states), we assign a value of 2 (UNKNOWN) to it. This also allows us to combine unknown outcomes from different interactions. Any remaining undetermined transitions imply that there are cell states for which multiple transitions are possible.

Concretely, define  $g^{(ij)}(X) \in S$  as the outcome of the interaction  $i \to j$  for a given input cell states X. Note that it takes value in S, indicating that the interaction is either on, off or the outcome is unknown. Let  $Z_{out}(X) = (Z_{out}^{(1)}(X), Z_{out}^{(2)}(X))$  be the three-valued output state given input state X. We construct the output state as follows:

$$Z_{out}^{(1)}(X) = g^{(11)}(X) \wedge g^{(12)}(X)$$
  

$$Z_{out}^{(2)}(X) = g^{(21)}(X) \wedge g^{(22)}(X).$$
(S5)

Here we have replaced ordinary multiplication by the  $\wedge$  operation that takes into account unknown outcomes. This three-valued output state needs to be translated to the actual possible output (binary) cell states of the system. Intuitively, if there is an unknown outcome, i.e.  $Z_{out}^{(i)} = 2$  for some *i*, then we should take into account all possible outcomes of that state. Hence we should consider states with both  $Z_{out}^{(i)} = 0$  and  $Z_{out}^{(i)} = 1$ .

Formally, let us denote the set of possible output cell states as  $\Sigma_{out}(X)$ , with elements in  $\{0,1\}^2$ . We construct the set  $\Sigma_{out}(X)$  through the construction of two maps. Let  $\mathcal{P}(\{0,1\}) = \{\emptyset, 0, 1, \{0,1\}\}$  denote the power set of  $\{0,1\}$ . First we define a map

$$\sigma_1 : S \to \mathcal{P}(\{0, 1\})$$

$$x \mapsto \begin{cases} x & x \in \{0, 1\} \\ \{0, 1\} & x = 2 \end{cases}$$
(S6)

This map constructs the set of possible output gene states, by deconstructing the element  $2 \in S$ into the set  $\{0, 1\}$  of possible outcomes. Extend the map to  $S^2$  by defining  $\sigma : S^2 \to \mathcal{P}(\{0, 1\})^2$ as  $\sigma(Z) = (\sigma_1(Z^{(1)}), \sigma_1(Z^{(2)})).$  Next, we have to put together the deconstructions to arrive at a set of output cell states. Recall that  $\mathcal{X}_1^{(2)} = \{(0,0), (0,1), (1,0), (1,1)\}$  is the phase space of a single cell with two genes. We define a map

$$\tau : \mathcal{P}(\{0,1\}) \times \mathcal{P}(\{0,1\}) \to \mathcal{P}(\mathcal{X}_1^{(2)})$$
$$(x_1, x_2) \mapsto x_1 \times x_2, \tag{S7}$$

where the × denotes an ordinary Cartesian product between sets. For instance, if  $x_1 = \{0, 1\}$ and  $x_2 = 1$ , then  $x_1 \times x_2 = \{(0, 1), (1, 1)\}$ . Hence, the second map constructs all the possible cell states from the possible states for each gene. Putting it together, we construct the set of output cell states as

$$\Sigma_{out}(X) = (\tau \circ \sigma \circ Z_{out})(X).$$
(S8)

Finally, once we have the output cell states for our input state X, we set

$$A_{X \to Y} = 1, \ \forall Y \in \Sigma_{out}(X).$$
(S9)

In other words, we draw edges from X to all cell states in the set of possible output states  $\Sigma_{out}(X)$ . Doing this for all input states X gives us the full adjacency matrix for the state diagram.

# S3 Analytic framework for traveling wave propagation

In this section, we provide a detailed analysis of traveling waves moving on a constant background of cells, which are found in cellular dialogues 15, 19, 33, 34 and 36. We first discuss features of traveling waves that characterize and distinguish different instances of traveling waves (Section S3.1). These features are used in an analytic estimate of the density of traveling wave states in the overall system in Section S3.2. The core of this section is composed of a derivation of a set of conditions for TW propagation (Section S3.3). We then discuss the performance of the analytic theory in terms of how well it recapitulates simulation results in Section S3.5. Finally, we sketch how to extend our method to dynamic patterns on an oscillatory background in Section S3.6.

## S3.1 Features of traveling waves

The traveling waves that we observe can be distinguished from each other through a number of features:

- 1. Orientation and direction of the wave. The waves can be oriented in different ways and for each cellular dialogue we observe waves of all different orientations. We distinguish between horizontally, vertically and diagonally oriented waves (see e.g. Figs. 2C, 2D, 2E and 2G). Horizontal waves wrap around the horizontal axis once, without wrapping around the vertical axis, and travel in the vertical direction. Vertical wave wrap around the vertical axis once, without wrapping around the horizontal axis, and travel in the horizontal axis, and travel in the horizontal axis, and travel in the horizontal axis once are around the vertical direction. Diagonal waves wrap around each of the two axes at least once and can travel in either direction. A more precise way of accounting for the geometry of the wave is through winding numbers, which will be introduced in section S3.2.
- 2. Presence of bends in the wave. We distinguish between straight and bent waves according to whether all cells in a band of the wave are aligned in the same direction. For a triangular lattice, there are three directions along which the cells can align themselves. In one case, we get straight horizontal waves (e.g. Fig. 2C), whereas in the other cases we get diagonally oriented waves (e.g. Fig. 2G). However, we can also get waves with one or more bends (e.g. Fig. 2E), points at which the alignment of the cells changes direction. Note that the cells located at the bends have a different set of nearest neighbors from the aligned cells. Furthermore, we can distinguish between bends that are in the direction of propagation (outward bends) and bends that are opposite to the direction of propagation (inward bends).
- 3. Number of waves. In the simplest case, the system self-organizes into a single wave on a uniform background (e.g. Figs. 2C and 2E). However, we also observe multiple coexisting waves, separated by each other by regions of cells with the background state (see e.g. Figs. 2A and 2D). Such waves have the same orientation and direction of travel, but are not necessarily aligned parallel with each other.

- 4. Number of different cell states in the wave. For almost all the waves we observed, we found wave made up of three different cell states. The background was made up of the fourth cell states. The exact states which make up the wave and their order varies from topology to topology, and sometimes also between different parameter sets of the same topology. In rare cases, we also found waves consisting of two types of cell states on a background of a third cell state.
- 5. Number of bands in the wave. In most cases, we find waves consisting of single bands of cells of the same state. Waves with bands with two or more layers of cells and waves where different cell states have different band widths have also been observed (see e.g. Fig. 2C).
- 6. Defects. In rare cases we may see waves which contain single-cell defects such as an additional cell of the same cell state attached to an otherwise normal wave.

# S3.2 Abundance of traveling waves

We now derive an estimate of the relative abundance of traveling waves of the forms we observe in the system. Due to the variety of morphologies these waves can take, we could expect them to take up a considerable portion of the total phase space. In this scenario, finding system conditions under which most of the simulations go to traveling waves would not be entirely surprising. On the other hand, if the relative abundance of traveling waves in the system is low, we could interpret this as a sign that there is a self-organizing mechanism that drives the system towards traveling wave formation.

First, we identify the key aspects of traveling waves and divide them into a limited number of categories. For each category, we then calculate the number of distinct shapes the waves can take, as well as the total number of distinct "snapshots" each wave form is made up of. This then gives us an estimate of the total number of states in the system that can be considered traveling waves.

We have previously provided a list of features that distinguish traveling waves from each other. While we have observed waves that differ in all these categories, we note that the vast majority of waves have the same features for a number of categories. In particular, most waves are composed of three cell states (with a fourth background state), are composed of a single band and have no defects. We also rarely observe more than two waves propagating simultaneously in moderately large systems (e.g. N = 256). Hence we only consider the orientation and direction, the presence of bends and the number of waves to account for the vast majority of observed wave forms. In the following we consider a generic single-banded wave with  $N = n^2$  cells.

The orientation of a wave can be made more precise by considering the number of times the wave wraps around each axis. Since the system is periodic, effectively we are dealing with wave that winds around each axis of a torus different numbers of times. Let  $W_x$ ,  $W_y$  be the winding

numbers around the horizontal and vertical axis. For a plane horizontal plane wave such as shown in Fig. 2A,  $W_x = 1$ ,  $W_y = 0$ . For a vertical wave such as in Fig. 2E,  $W_x = 0$ ,  $W_y = 1$ . The diagonal wave in Fig. 2G has  $W_x = 1$ ,  $W_y = 2$ , but we can also imagine diagonal waves that wrap around the system in different ways. The most common winding numbers are listed in Table 2. As is apparent from the table, we mostly observe simple waves that are either horizontal, vertical or diagonal, but wrap around the axes only a few times. Note that traveling waves are characterized by  $W_x + W_y \ge 1$ , i.e. a traveling wave always wraps at least once around one of the axes. (Smaller structures that do not wrap around either axis but do translate in space are referred to as traveling pulses).

| $W_x$ | $W_y$ | $N_C(W_x, W_y)$ | $N_{wf}(W_x, W_y)$   | $T(W_x, W_y)$ |
|-------|-------|-----------------|--|---------------|
| 1     | 0     | n               | 1  | n             |
| 0     | 1     | n               | $\left(\begin{array}{c}n\\\frac{1}{2}n\end{array}\right)$            | n             |
| 1     | 1     | $\frac{3}{2}n$  | $\left(\begin{array}{c}\frac{3}{2}n\\\frac{1}{2}n\end{array}\right)$ | 2n            |
| 1     | 2     | 2n              | 1  | 2n            |
| 2     | 1     | $\frac{5}{2}n$  | $\left(\begin{array}{c} \frac{5}{2}n\\n\end{array}\right)$           | 2n            |

Table 2: Main properties of most common types of waves. n is the linear grid size, with  $N = n^2$ . We assume that n is an even number, so that the system is a perfect hexagonal lattice on a torus. The data is based on empirical observations of self-generated traveling waves.  $W_x, W_y$  are the winding numbers,  $N_C$  is the number of cells of the wave,  $N_{wf}$  is the number of wave forms and T is the period of the wave.

Once we fixed the winding numbers, the precise form of the wave is often still unspecified. For instance, a vertical wave can have different number of bends in both directions. Neverthe the set of wave forms given  $(W_x, W_y)$ . Let us consider a single wave that travels in a fixed direction. Suppose we pick a random cell of the wave. Empirically, we find that each cell of the wave that has the same state has precisely two neighbors with the same state. This is even the case when there are complicated bends in the wave. Now pick one of the neighbors of our selected cell that has the same cell state. The nearest-neighbor vector that connects the two cells lies along one of the six directions one can travel in on a hexagonal lattice. These can expressed in terms of the basis vectors of the lattice as  $\vec{e_1}, -\vec{e_1}, \vec{e_2}, -\vec{e_2}, \vec{e_3} \equiv \vec{e_2} - \vec{e_1}, -\vec{e_3} = \vec{e_1} - \vec{e_2}$  (Fig. S10B - left figure). The second cell has a unique neighbor of the same state that we have not selected yet. The vector between the second and third cell defines a new direction that we record. We can therefore continue this procedure and pick subsequent cells in our wave, until we get back to our original cell. This is because the wave wraps around an axis at least once as noted before. For each step we take, we keep track of the direction we need to move in to get to the next cell. At the end, we count the number of steps in each of the six directions obtained through this procedure (illustrated in Fig. S5B - right figure). Let us denote these by  $\{n_{i,\alpha}\}$ , where  $1 \leq i \leq 3$  and  $\alpha \in \{-,+\}$ . For example,  $n_{2,-}$  gives the number of steps we took in the  $-\vec{e_2}$  direction.

Wave forms differ by their set of nearest-neighbor vectors that we obtain with this procedure. Nevertheless, once we fix the winding numbers, this constraints the possible sets of direction vectors in a way we can make precise. First, we note that empirically we find that waves with fixed winding numbers always have the same number of cells of a given state, which we will denote  $N_C(W_x, W_y)$ . Empirical results for commonly found waves are listed in Table 2. For instance, for a horizontal wave, we find that it always has  $N_c(1,0) = n$  cells of a given state, which make up exactly one row of the lattice. This constraints the total number of nearest-neighbor vectors to  $N_C(W_x, W_y)$ , such that our first constraint is

$$\sum_{i,\alpha} n_{i,\alpha} = n_{1,-} + n_{1,+} + n_{2,-} + n_{2,+} + n_{3,-} + n_{3,+} = N_C \left( W_x, W_y \right)$$
(S10)

Next, the winding numbers constrain the number of occurrences of each nearest-neighbor vector. For instance, for a horizontal wave the nearest neighbor vectors when added up must be align in the horizontal direction, with a magnitude equal to the grid size. However, a priori this does not imply that all nearest neighbor vectors are in the  $\vec{e_1}$  direction, since  $\vec{e_2}$  and  $\vec{e_3}$  also have horizontal components. In general, the constraints are that the number of steps taken in the horizontal and vertical directions must be equal to the  $\pm W_x n$  and  $\pm W_y n$  in order to return to the original cell. The sign degeneracy comes from the fact that starting from the initial cell we pick, we can traverse the wave in two different directions, which yield winding numbers that differ by a minus sign. Working out these conditions, we derive the following constraints:

$$n_{1,-} - n_{1,+} - \frac{n_{2,-}}{2} + \frac{n_{2,+}}{2} + \frac{n_{3,-}}{2} - \frac{n_{3,+}}{2} = \pm W_x \ n, \tag{S11}$$

$$-n_{2,-} + n_{2,+} - n_{2,+} + n_{3,+} = \pm W_y \ n \tag{S12}$$

We can now try to solve these constraints together with the general constraints  $0 \leq n_{i,\alpha} \leq n$ for all *i*,  $\alpha$  for given winding numbers  $W_x$ ,  $W_y$ . For all the winding numbers listed in Table 2, we obtained solutions of the form  $n_{i_1,\alpha_1} = n_{i_1,\alpha_1}(n) > 0$ ,  $n_{i_2,\alpha_2} = n_{i_2,\alpha_2}(n) > 0$  for some  $i_1, \alpha_1$  and  $i_2, \alpha_2, n_{i,\alpha} = 0$  for all other  $i, \alpha$ . The interpretation of this result is that in practice all waves are formed by traveling continuously in two directions, i.e. they never "bend back". Secondly, we found that the number of steps in each direction is a linear function n, so we get an explicit scaling of our results with system size.

We can now readily obtain the number of wave forms that satisfy these constraints. This reduces to a simple combinatorics problem where the wave forms differ by the order in which the directions  $i_1$ ,  $\alpha_1$  and  $i_2$ ,  $\alpha_2$  appear, in the procedure described above. Since there are  $N_C(W_x, W_y)$ such vectors (one for each cell of a wave layer), there are  $N_{wf}(W_x, W_y) \equiv \begin{pmatrix} N_C(W_x, W_y) \\ n_{i_1, \alpha_1} \end{pmatrix} = \begin{pmatrix} N_C(W_x, W_y) \\ n_{i_2, \alpha_2} \end{pmatrix}$  ways of ordering the vectors in either direction, corresponding to the number of possible wave forms with the given winding numbers. Finally, the sign degeneracy of  $W_x, W_y$ introduces an additional factor of 2 whenever both  $W_x, W_y > 0$ . This is because for the four possibilities  $\{(W_x, W_y), (W_x, -W_y), (-W_x, W_y), (-W_x, -W_y)\}, (W_x, W_y)$  and  $(-W_x, -W_y)$ give equivalent waves, whereas  $(-W_x, W_y) \equiv (W_x, -W_y)$  yields a different wave, corresponding to a different diagonal orientation.

The direction of a wave can in principle be in any of the six directions the hexagonal lattice allows. However, once the orientation of a wave is fixed, there are only two possible directions remaining. For instance, for a horizontal wave, the only directions are up and down.

The number of waves that can simultaneously propagate depends on the system size. For N = 256, we rarely observe more than two simultaneously propagating waves. Note that the waves need to have the same orientation and direction of motion, or else they would collide and annihilate or form new waves. Once the shapes of both waves are fixed, an additional variable is the spacing between the waves. Assume that both waves are horizontal, then the variable is the number of rows between the waves. For two waves, the distance between the waves lies in the range  $[1, \frac{n-6}{2}]$ , with  $\frac{n-6}{2}$  being an upper bound for when both waves are straight. This gives a degeneracy of roughly  $\frac{n-6}{2}$ . This is because both waves take up 3 rows, and from the arrangement of the remaining rows we take the shortest distance since the system is periodic.

Putting everything together, we now obtain our general estimate for the density of traveling waves in phase space. We estimate this to be in the order of

$$N n_{dir} \sum_{n_{waves}} n_{dist}(n_{waves}) \sum_{W_x, W_y} D(W_x, W_y) N_{wf}(W_x, W_y)^{n_{waves}}.$$
 (S13)

Here  $n_{dir} = 2$  signifies the two directions of propagation, the first sum is over the total number of waves in the system  $n_{waves}$ , the number of unique distances between the waves is denoted  $n_{dist}$ , the degeneracy after accounting for negative winding numbers is denoted  $D(W_x, W_y)$ , and the second sum is over the winding numbers  $W_x$ ,  $W_y \in \mathbb{N}_0$ . From the previous part we estimate  $n_{dist} (1) = 1$ ,  $n_{dist} (2) = \frac{n-6}{2}$ . The final term signifies the fact that in case of multiple waves, they can in principle have independent wave forms (with the same direction and winding numbers). The term N in front accounts for the possible positions of the wave on the lattice, obtained simply by counting the number of ways to place a given cell of the wave on the lattice. This is an upper bound since in case waves with symmetry different placements of this selected cell could still give the same configuration.

**Traveling wave formation time** The aim of doing the traveling wave density estimation is to provide more direct evidence for the intuitive idea that a self-organizing mechanism drives the formation of traveling waves (TWs), and that these do not simply arise by chance because TW states are abundant. To do so, we looked at the TW formation time, i.e. the number of time steps it takes to go from a random initial state to a TW. We compared empirically observed formation times from simulations with expected simulation times for a random process based on our wave density estimation. The random process can be seen as a null hypothesis stating that there is no self-organization, but rather TWs form by chance as the system randomly samples states in the system. To show that there is a self-organization process, we therefore demonstrate that the findings on TW formation times differ considerably from those of a random process.

More precisely, we considered a stochastic process whereby subsequent states of the system would be randomly sampled over the set of all  $2^N$  states with a uniform distribution, i.e. each next state would be drawn from the set of all states with equal probability for each of the  $2^N$ states. Under this assumption, the formation time of a TW would be equivalent to the first time of success in a Bernoulli process where the probability of success p equals the density of TWs derived above (Equation S13). Hence the waiting time - here the TW formation time would follow a geometric distribution with the same parameter p, with an expected waiting time of 1/p.

We first considered the expected waiting time as a function of the system size (measured by the grid size  $\sqrt{N}$ ). For the random process described above, we obtained a mean formation time that scales faster than exponentially with the grid size (Fig. S10C - left plot), and quickly reaches times that are orders of magnitude larger in our simulations. Conversely, our simulated data shows that the mean formation time scales linearly with grid size (Fig. S10C - left plot), and has typical values well within reasonable simulation time bounds.

Given this result, it is still possible that our simulations probe only the waves that form within the limited simulation time, while the majority of other TWs have formation times that are orders of magnitude larger. However, in this case we would obtain large fractions of simulations that do not reach any steady state at our maximum simulation time, which is not the case (see for instance Fig. S8A - here the fraction of simulations reaching  $t_{\text{max}}$  is low across most parameter sets capable of generating TWs). Furthermore, a second counterargument is that the TW formation time distribution would then be skewed towards the right, with relatively more TWs taking times closer to the maximum simulation time. This would be necessary to produce the geometric distribution of the Bernoulli process (Fig S10D - left plot). In contrast, we observed the opposite - the simulated formation time distribution is skewed toward the left, with relatively more wave forming within a short amount of time (Fig S10D - right plot).

In conclusion, we have provided evidence that traveling waves form through a self-organizing process whereby the system 'actively' converges onto traveling wave attractor states, rather than 'passively' exploring the state space of all possible configurations and randomly finding traveling wave attractors. This result should be intuitive given the relative ease of finding traveling waves in our simulations. Note that we have not aimed to give a rigorous definition of self-organization in this context or formally proving our claims, but merely want to point out that the self-organization process differs significantly from a purely random process.

#### S3.3 Traveling wave propagation conditions

#### S3.3.1 General conditions for pattern propagation

Before working out the case of traveling waves in detail, let us discuss what pattern propagation means in general. Suppose we have a pattern that periodically repeats itself in time. All information about the pattern is encoded in the states of the pattern over one period. Let us denote these by  $\{X(0), X(1), \ldots, X(\tau)\}$ , where  $X(t) = \{X_k^{(i)}(t)\}_{1 \le k \le N}^{1 \le i \le l}$  is the state of the system at time t as specified by the states of each of the genes of each cell. These can be considered a series of snapshots of the system that make up a movie of the dynamic pattern when played. In general, take an arbitrary state  $\xi(t)$  and suppose that the system is updated according to a rule

$$\xi(t+1) = f(\xi(t); P),$$
 (S14)

for some unspecified function f of the current state of the system that depends on the parameters of the system denoted by P. The condition that the pattern can propagate under the set of parameters P is precisely that f updates each snapshot of the pattern to the next snapshot of the system. In other words, X(t+1) = f(X(t); P) for all  $0 \le t \le \tau - 1$ .

In general, this would put constraints on each of the cells of the system, leading to a convoluted set of conditions for pattern propagation. However, in cases where the pattern exhibits a symmetry, these conditions can be drastically simplified. In one extreme case, if the pattern is a homogeneous collection of identical cells, at any time we would only need to check one set of conditions for an arbitrary cell of the system. Conversely, suppose the system is completely anisotropic for the whole duration of the pattern trajectory. Each cell then sees a different environment at any time. We then need to check all  $N \times l \times \tau$  conditions for each of the genes of each of the cells of the system, at each time step of the system.

The case of traveling waves allows us to exploit the symmetry of the pattern to drastically reduce the number of conditions for pattern propagation. First note that traveling waves are characterized by the fact that the state upon updating is related to the previous state by a simple translation in space. This means that rather than checking conditions for each time step, we only need to check that the wave propagates at one arbitrary time step. Furthermore, the spatial symmetry of the system allows us to check only a small number of cells of the system, as will be explained in the next section.

#### S3.3.2 Straight and bent waves

To derive conditions for the propagation of these waves, we look at straight (plane) waves (Fig. 4A, "straight wave"), waves with a single outward bend (Fig. 4A, "bent wave") and waves with a single inward bend (Fig. 4A, "bent wave" with reverse direction of propagation). In this way, we obtain results applicable to the vast majority of waves observed in the system, including more complicated waves which are typically locally still similar to these "simple" waves. For instance, the configuration in Fig. 1F contains two waves with multiple bends. However, the nearest neighbors of any cell is identical to the neighbor structure of a cell in one of the three prototype waves. Namely, the cells at the tip of the wavefront have nearest neighbors that is identical to the cell at the tip of the outward bent single wave. The cells that are bent towards the back of the wavefront have a neighbor that is identical to those at the bend of the inward bent wave.

Therefore, it suffices to study the conditions for propagation of each of these three simple types of waves. This gives a first approximation to the propagation conditions for more complicated waves, and is valid especially when the interaction between cells is not too strong and takes place mostly on a local scale (i.e. when  $a_0$  is sufficiently high). The types of waves which are not covered by this analysis are waves with multiple bands (because the cells of such waves have different nearest neighbors), and waves with defects (which are too rare to motivate analysis of each special case).

#### S3.3.3 Structure of traveling waves

**Wave structure** All waves consisting of three consecutive single bands of cells have a similar spatial structure. For any instance of such a wave, we can identify six types of cells that each have a unique set of nearest neighbors. Let us denote these six types of cells as follows (see Fig. 4B):

- 1.  $E_F$  front exterior
- 2. F front
- 3. M middle
- 4. B back
- 5.  $E_B$  back exterior
- 6. E exterior

Note that the types E,  $E_F$  and  $E_B$  all have the same cell state (the state of the white color in Fig. 2A). However, we divide the exterior cells up into three classes because they have different sets of nearest neighbors. A cell of type  $E_F$  in front of the wave neighbors F cells, whereas a cell  $E_F$  at the back of the wave neighbors B cells, while the rest of the E cells border only other E cells.

Hence, the six types of cells have four different cell states, which we denote as X(F), X(M), X(B) and  $X(E) = X(E_F) = X(E_B)$ . For binary cells, the possible cell states form the set  $S = \{(0,0), (0,1), (1,0), (1,1)\}.$ 

At any straight segment of a wave, the cells of the wave and those bordering the wave have exactly the same local structure (nearest neighbors). Concretely, this means that any cell of the straight segment borders the same number of cells of each state (Table 1). For instance, an F cell will always border two cells with state X(F), two cells with state X(M) and two cells with state X(E).

The set of nearest neighbors of a bent wave differs from that of plane waves only at the location of the bend. The rest of the cells have nearest neighbors identical to plane wave cells (Fig. 4B). We therefore take the propagation condition for the cells at the bend into account separately (Table 3).

| Plane wave |                           |           |           |                   |
|------------|---------------------------|-----------|-----------|-------------------|
| Wave state | Number of neighbor states |           |           |                   |
|            | $\overline{X(F)}$         | X(M)      | X(B)      | X(E)              |
| $E_F$      | 2                         | 0         | 0         | 4                 |
| F          | 2                         | 2         | 0         | 2                 |
| M          | 2                         | 2         | 2         | 0                 |
| B          | 0                         | 2         | 2         | 2                 |
| $E_B$      | 0                         | 0         | 2         | 4                 |
| E          | 0                         | 0         | 0         | 6                 |
| Wav        | ve with                   | outwar    | d bend    |                   |
| Wave state | Num                       | ber of ne | eighbor s | states            |
|            | $\overline{X(F)}$         | X(M)      | X(B)      | $\overline{X(E)}$ |
| $E_F$      | 1                         | 0         | 0         | 5                 |
| F          | 2                         | 1         | 0         | 3                 |
| M          | 3                         | 2         | 1         | 0                 |
| B          | 0                         | 3         | 2         | 1                 |
| $E_B$      | 0                         | 0         | 3         | 3                 |
| E          | 0                         | 0         | 0         | 6                 |
| Wa         | ve with                   | inward    | l bend    |                   |
| Wave state | Number of neighbor states |           |           |                   |
|            | X(F)                      | X(M)      | X(B)      | X(E)              |
| $E_F$      | 3                         | 0         | 0         | 3                 |
| F          | 2                         | 3         | 0         | 1                 |
| M          | 1                         | 2         | 3         | 0                 |
| В          | 0                         | 1         | 2         | 3                 |
| $E_B$      | 0                         | 0         | 1         | 5                 |
| E          | 0                         | 0         | 0         | 6                 |

Table 3: Cell states of nearest neighbors of the six types of cells  $(E_F, F, M, B, E_B, E)$  for straight waves and for the cells at the tip of waves with bends (Fig. 4B). Results are for a hexagonal lattice with coordination number z = 6.

#### S3.3.4 Traveling wave propagation conditions

For a traveling wave to propagate, we need a number of conditions to be satisfied. Since traveling waves have the property that the entire structure translates forward by one step, we can easily find these conditions. Basically, all cells of one layer take up the state of the next layer and the background cells remain constant. For example, an  $E_F$  cell right in front of the wave should become an F cell at the next time step. Hence we require that the cell obtains the state X(F) upon updating. Let  $\alpha \to X(\alpha')$  denote the condition that a cell of type  $\alpha$  acquires state  $X(\alpha')$  according to the update rule. Then we can succinctly write our set of conditions as:

$$E_F \to X(F)$$

$$F \to X(M)$$

$$M \to X(B)$$

$$B \to X(E)$$

$$E_B \to X(E)$$

$$E \to X(E)$$
(S15)

For a straight wave without bends, these conditions need to be checked only once, for cells that have the nearest neighbors as detailed in the first table in Table 3. For waves with at least one bend, both the condition at the location of the bend (either outward or inward) as well as the condition for plane waves (for the straight segments of the wave) need to be checked. For waves with a zig-zag pattern that have no straight segments (e.g. Fig. 2D), only the conditions for inward and outward bends need to be checked.

Mathematically, we can represent the propagation conditions through a general set of equations in terms of the network topology (specified by  $M_{\text{int}}^{(ij)}$ ), the sensed concentration ( $Y_{\alpha}^{(j)}$  for a cell of type  $\alpha$ ), and the sensing threshold  $K^{(ij)}$ ), which for AND-logic signal integration takes the following form:

$$X^{(i)}(\alpha) = \prod_{j} \left[ \theta \left( (Y_{\alpha}^{(j)} - K^{(ij)}) M^{(ij)} \right) + 1 - |M^{(ij)}| \right]$$
(S16)

Here  $X^{(i)}(\alpha)$  represents the state of gene *i* the cell with state  $\alpha$  should transition into. Here we define  $\theta(x)$  as the Heaviside function with convention  $\theta(0) = 0$ , i.e.  $\theta(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$ . Likewise, for OR-logic we have (see equation for OR-logic time evolution):

$$X^{(i)}(\alpha) = \sum_{j} \theta \left( (Y_{\alpha}^{(j)} - K^{(ij)}) M^{(ij)} \right) - \prod_{j} \theta \left( (Y_{\alpha}^{(j)} - K^{(ij)}) M^{(ij)} \right).$$
(S17)

#### S3.3.5 Nearest-neighbor approximation

Given an exact form of the wave and a specific interaction network of the two genes, we can work out the six conditions for traveling wave propagation, to obtain exact conditions in terms of system parameters. However, since the waves can have different features, we look for a more general approach that predicts propagation independent of the precise shape of the wave. To do this, we will apply a nearest-neighbor approximation (NNA). The idea is to only consider the immediate neighbors of a cell when calculating the concentration it senses, and take into account the rest of the cells through averaging and assuming they are randomly distributed.

Write  $Y_{\alpha}^{(i)}$  for the concentration of molecule i a cell of type  $\alpha$  senses. Then we can split the sensed concentration into terms of the cell itself, its neighbors and an approximation of the rest of the lattice, which we assume to be independent of the cell type  $\alpha$ :

$$Y_{\alpha}^{(i)} = Y_{self}^{(i)}(\alpha) + Y_{nei}^{(i)}(\alpha) + Y_{MF}^{(i)}.$$
(S18)

Recall that  $C^{(i)}(X)$  is the secretion rate for molecule *i*. Denote  $f_{nn}^{(i)} = f^{(i)}(a_0)$  as the nearest neighbor interaction strength, and  $n(X; \alpha)$  as the number of cells of state X that neighbor a cell of type  $\alpha$ . The sensed concentration due to neighbors can then be written as

$$Y_{nei}^{(i)}(\alpha) = \sum_{X \in S} f_{nn}^{(i)} n(X;\alpha) C^{(i)}(X).$$
(S19)

The contribution of rest of the lattice is estimated through a mean-field approximation. For a wave with  $N_w$  waves, each consisting of bands of width W, with winding numbers  $W_x$ ,  $W_y$ (Section S3.2), we can calculate the proportion of cells that have either of the genes on. This proportion depends on the cell states of the wave and background cells. In general, the fraction of cells with a given gene on is

$$p^{(i)} = \frac{N_W W}{N_C(W_x, W_y)} \left( X(F)^{(i)} + X(M)^{(i)} + X(B)^{(i)} \right) + \left(1 - 3\frac{N_W W}{N_C(W_x, W_y)}\right) X(E)^{(i)}$$
(S20)

Here  $X(S)^{(i)}$  denotes the state of gene *i* of a cell state X(S),  $N_C(W_x, W_y)$  is the number of cells of one layer given winding numbers  $(W_x, W_y)$  as used in the wave density estimation (Section S3.2). The mean-field contribution is then estimated to be the interaction strength of all cells excluding the nearest neighbors times the average secretion rate of the cells:

$$Y_{MF}^{(i)} = \left(f_N^{(i)} - 6f_{nn}^{(i)}\right) \left[ \left(C_{ON}^{(i)} p^{(i)} + C_{OFF}^{(i)} \left(1 - p^{(i)}\right) \right].$$
(S21)

In practice, we looked at single waves  $(N_W = 1)$  with layers of a single cell thick (W = 1), and only considered horizontal and vertical waves, for which  $N_C(W_x, W_y) = n = \sqrt{N}$  (Table 2). In this case, the expression for  $p^{(i)}$  simplifies to:

$$p^{(i)} = \frac{1}{\sqrt{N}} \left( X(F)^{(i)} + X(M)^{(i)} + X(B)^{(i)} \right) + \left( 1 - \frac{3}{\sqrt{N}} \right) X(E)^{(i)}.$$
 (S22)

#### S3.4 Explicit example

In this section, we work out an explicit example of the propagation conditions we derived. Consider cellular dialogue 15, which has interaction matrix  $M_{\text{int}} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ . From simulations we observed traveling waves with the composition

- X(F) = (1, 0),
- X(M) = (1, 1),
- X(B) = (0, 1),
- X(E) = (0, 0).

**Conditions for propagation** Let us denote  $E_F = (0,0)_F$  and  $E_B = (0,0)_B$ . Explicitly, the conditions for propagation can be expressed in terms of inequalities, as shown below. Here,  $\wedge$  denotes a logical AND and  $\vee$  denotes a logical OR.

| Condition for propagation | Condition on gene 1   | Condition on gene 2            |       |
|---------------------------|---|--------------------------------|-------|
| $(0,0)_F \to (1,0)$       | $Y^{(2)}_{(0,0)_F} < K^{(12)} \land Y^{(1)}_{(0,0)_F} > K^{(11)}$ | $Y^{(1)}_{(0,0)_F} < K^{(21)}$ |       |
| $(1,0) \to (1,1)$         | $Y_{(1,0)}^{(2)} < K^{(12)} \land Y_{(1,0)}^{(1)} > K^{(11)}$     | $Y_{(1,0)}^{(1)} > K^{(21)}$   |       |
| $(1,1) \to (0,1)$         | $Y_{(1,1)}^{(2)} > K^{(12)} \lor Y_{(1,1)}^{(1)} < K^{(11)}$      | $Y_{(1,1)}^{(1)} > K^{(21)}$   | (S23) |
| $(0,1) \to (0,0)$         | $Y_{(0,1)}^{(2)} > K^{(12)} \lor Y_{(0,1)}^{(1)} < K^{(11)}$      | $Y_{(0,1)}^{(1)} < K^{(21)}$   |       |
| $(0,0)_B \to (0,0)$       | $Y^{(2)}_{(0,0)_B} > K^{(12)} \lor Y^{(1)}_{(0,0)_B} < K^{(11)}$  | $Y_{(0,0)_B}^{(1)} < K^{(21)}$ |       |
| $(0,0)_R \to (0,0)$       | $Y_{(0,0)_R}^{(2)} > K^{(12)} \lor Y_{(0,0)_R}^{(1)} < K^{(11)}$  | $Y_{(0,0)_R}^{(1)} < K^{(21)}$ |       |

We can show that the last condition for  $(0,0)_R$  is redundant in general. Namely, we have  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_F}^{(i)}$  and  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_B}^{(i)}$  for both molecules i = 1, 2. This implies that if the condition for gene 2 for  $(0,0)_B$  is fulfilled, then the condition for  $(0,0)_R$  is automatically fulfilled, since  $Y_{(0,0)_R}^{(i)} \leq Y_{(0,0)_B}^{(i)} < K^{(21)}$ . For gene 1, the conditions  $Y_{(0,0)_F}^{(2)} < K^{(12)}$  (condition on  $(0,0)_F$ ) and  $Y_{(0,0)_R}^{(2)} > K^{(12)}$  (condition on  $(0,0)_R$ ) give a contradiction. Since the first condition has to be true due to the AND function, the second is necessarily false. This leaves  $Y_{(0,0)_R}^{(1)} < K^{(11)}$ . But since  $Y_{(0,0)_R}^{(1)} = Y_{(0,0)_B}^{(1)}$ , this becomes equivalent to the condition  $Y_{(0,0)_B}^{(2)} < K^{(12)}$  (condition on  $(0,0)_B$ ). Furthermore, we also have  $Y_{(0,0)_B}^{(2)} = Y_{(1,0)}^{(2)}$ , and since  $Y_{(1,0)_R}^{(2)} < K^{(12)}$ , the condition  $Y_{(0,0)_B}^{(1)} > K^{(12)}$  must be false. Thus,  $Y_{(0,0)_B}^{(1)} < K^{(11)}$  becomes the only condition for gene 1 for  $(0,0)_B$ . Note that these arguments are specific to the network and wave form under consideration and do not need to hold in general.

**Plane waves** We work out the equations explicitly for plane waves. From Equation S18-S21 and Table 3, we get the sensed concentrations

$$\begin{split} Y_{(0,0)_{F}}^{(1)} &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)}, \qquad Y_{(0,0)_{F}}^{(2)} = 1 + 6f_{nn}^{(2)}, \\ Y_{(1,0)}^{(1)} &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)}, \qquad Y_{(1,0)}^{(2)} = 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)}, \\ Y_{(1,1)}^{(1)} &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)}, \qquad Y_{(1,1)}^{(2)} = C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)}, \\ Y_{(0,1)}^{(1)} &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)}, \qquad Y_{(0,1)}^{(2)} = C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)}, \\ Y_{(0,0)_{B}}^{(1)} &= 1 + 6f_{nn}^{(1)}, \qquad Y_{(0,0)_{B}}^{(2)} = 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)}, \end{split}$$
(S24)

We can then write out the conditions from Equation S23 explicitly. For gene 1 this gives:

$$\begin{array}{ll} (0,0)_{F} \rightarrow (1,0): & 1+6f_{nn}^{(2)}+Y_{MF}^{(2)} < K^{(12)} \wedge 1+2C_{ON}^{(1)}f_{nn}^{(1)}+4f_{nn}^{(1)}+Y_{MF}^{(1)} > K^{(11)} \\ (1,0) \rightarrow (1,1): & 1+2C_{ON}^{(2)}f_{nn}^{(2)}+4f_{nn}^{(2)}+Y_{MF}^{(2)} < K^{(12)} \wedge \\ & C_{ON}^{(1)}+4C_{ON}^{(1)}f_{nn}^{(1)}+2f_{nn}^{(1)}+Y_{MF}^{(1)} > K^{(11)} \\ (1,1) \rightarrow (0,1): & C_{ON}^{(2)}+4C_{ON}^{(2)}f_{nn}^{(2)}+2f_{nn}^{(2)}+Y_{MF}^{(2)} > K^{(12)} \vee \\ & C_{ON}^{(1)}+4C_{ON}^{(1)}f_{nn}^{(1)}+2f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(11)} \\ (0,1) \rightarrow (0,0): & C_{ON}^{(2)}+4C_{ON}^{(2)}f_{nn}^{(2)}+2f_{nn}^{(2)}+Y_{MF}^{(2)} > K^{(12)} \vee \\ & 1+2C_{ON}^{(1)}f_{nn}^{(1)}+4f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(11)} \\ (0,0)_{B} \rightarrow (0,0): & 1+6f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(11)} \end{array}$$

For gene 2, the conditions are

$$\begin{split} &(0,0)_F \to (1,0): \quad 1+2C_{ON}^{(1)}f_{nn}^{(1)}+4f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(21)} \\ &(1,0)\to (1,1): \quad C_{ON}^{(1)}+4C_{ON}^{(1)}f_{nn}^{(1)}+2f_{nn}^{(1)}+Y_{MF}^{(1)} > K^{(21)} \\ &(1,1)\to (0,1): \quad C_{ON}^{(1)}+\ 4C_{ON}^{(1)}f_{nn}^{(1)}+2f_{nn}^{(1)}+Y_{MF}^{(1)} > K^{(21)} \\ &(0,1)\to (0,0): \quad 1+\ 2C_{ON}^{(1)}f_{nn}^{(1)}+4f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(21)} \\ &(0,0)_B\to (0,0): \quad 1+\ 6f_{nn}^{(1)}+Y_{MF}^{(1)} < K^{(21)} \end{split}$$

Since  $f_{nn}^{(1)} < 1$  (i.e. the interaction of a cell with itself should be larger than that with its nearest neighbor), we can simplify the equations to account for redundancy. After some algebraic manipulations, this reduces the conditions to the following set of inequalities:

$$1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(21)}$$

$$C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(21)}$$

$$1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)} + Y_{MF}^{(2)} < K^{(12)}$$

$$C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)} > K^{(12)}$$

$$1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)} < K^{(11)}$$

$$1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)} > K^{(11)}$$
(S27)

Note that for this particular example, the constraints reduce to a simple set of constraints for each of the three interactions in the system. Namely, the first two inequalities involve only parameters that affect the interaction  $2 \leftarrow 1$ , e.g.  $C_{ON}^{(1)}$  and  $K^{(21)}$ , whereas the second and third pair involve only the interactions  $1 \leftarrow 2$  and  $1 \leftarrow 1$  respectively. This does not need to be the case in general, since genes which are regulated by both genes will produce coupled constraints in terms of both interactions. Hence, in this case we can recast the conditions into

Cellular Dialogs: cell-cell communication through diffusible molecules yields dynamic spatial patterns 27

a concise set of equations:

$$K_{\min}^{(1,1)}(C_{ON}^{(1)}) \leq K^{(1,1)} \leq K_{\max}^{(1,1)}(C_{ON}^{(1)}),$$
  

$$K_{\min}^{(1,2)}(C_{ON}^{(2)}) \leq K^{(1,2)} \leq K_{\max}^{(1,2)}(C_{ON}^{(2)}),$$
  

$$K_{\min}^{(2,1)}(C_{ON}^{(1)}) \leq K^{(2,1)} \leq K_{\max}^{(2,1)}(C_{ON}^{(1)})$$
(S28)

where the  $K_{\min}^{(i,j)}$  and  $K_{\max}^{(i,j)}$  are functions of  $C_{ON}^{(j)}$  defining the minimal and maximal possible values of  $K^{(i,j)}$ . Explicitly, we have

$$\begin{aligned} K_{\min}^{(1,1)}(C_{ON}^{(1)}) &= 1 + 6f_{nn}^{(1)} + Y_{MF}^{(1)}, \\ K_{\max}^{(1,1)}(C_{ON}^{(1)}) &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)}, \\ K_{\min}^{(1,2)}(C_{ON}^{(2)}) &= 1 + 2C_{ON}^{(2)}f_{nn}^{(2)} + 4f_{nn}^{(2)} + Y_{MF}^{(2)}, \\ K_{\max}^{(1,2)}(C_{ON}^{(2)}) &= C_{ON}^{(2)} + 4C_{ON}^{(2)}f_{nn}^{(2)} + 2f_{nn}^{(2)} + Y_{MF}^{(2)}, \\ K_{\max}^{(2,1)}(C_{ON}^{(1)}) &= 1 + 2C_{ON}^{(1)}f_{nn}^{(1)} + 4f_{nn}^{(1)} + Y_{MF}^{(1)}, \\ K_{\max}^{(2,1)}(C_{ON}^{(1)}) &= C_{ON}^{(1)} + 4C_{ON}^{(1)}f_{nn}^{(1)} + 2f_{nn}^{(1)} + Y_{MF}^{(1)}. \end{aligned}$$
(S29)

Note also that  $Y_{MF}^{(i)}$  is a linear function of  $C_{ON}^{(i)}$  (Equation S21), so that these constraints reduce to linear relations between  $C_{ON}^{(i)}$  and  $K^{(ij)}$  for each interaction  $i \leftarrow j$ . The values  $(C_{ON}^{(i)}, K^{(ij)})$ together determine the relative strength of this interaction, and we find that its strength is constrained by two inequalities that determine a reduced but unbounded region of phase space. The boundaries of these regions together with the predicted TW conditions are plotted in Fig. S5B. This gives an alternative view of the parameter sets that can support TWs next to the spider charts, which only show that most parameters can span several orders of magnitude but do not directly reveal the structure of the set of TW parameters. In contrast, the analytic result reveals that each of the circuit parameters  $(C_{ON}^{(i)}, K^{(ij)})$  is unbounded from above, but is confined to a region such that each interaction can be neither too strong  $(C_{ON}^{(i)} \gg K^{(ij)})$ nor too weak  $(C_{ON}^{(i)} \ll K^{(ij)})$ , except in the case of the self-activation loop where we tend to have  $C_{ON}^{(i)} \gg K^{(ij)}$ . Similar results to Fig. S5B are obtained for the other networks that can support TWs. The set of inequalities (Equation S28-S29) also allows us to analytically calculate traveling wave robustness, as will be discussed in Section S4.

#### S3.5 Performance of the analytic framework

To assess the validity of the analytic framework derived in the previous sections, we directly compared the predictions from the theory to actual simulations of the waves. We quantified the degree to which these results match and considered the accuracy of the main approximation in the analytic framework, the nearest-neighbor approximation (Eqs. S18 and S21).

**Computational search for traveling waves** We verified with our analytic approach that the above wave forms are indeed the only possible wave forms for two-gene networks. To this end, we screened a large number of parameter sets for all distinct two-gene networks. Specifically, we checked the six conditions Equation S15 for wave propagation for a total of  $10^6$  parameter sets for each network. The parameter sets were generated by Latin hypercube sampling over all non-zero  $C_{ON}^{(i)}$  and  $K^{(ij)}$  parameters. We considered a network to be capable of generating a wave if for at least one of the  $10^6$  parameter sets all the conditions for traveling wave propagation were fulfilled. The results were consistent among the three types of waves (plane, with inward bends, with outward bends) that we examined: in all three cases exactly the same results were found.

**Statistical measures for performance** The performance of the analytic method we derived is determined by how well it predicts the conditions under which traveling waves can propagate. We can view our analytic theory as a binary classifier that predicts for a given gene network and given set of parameters whether TWs can propagate. The theory takes as input a set of parameters and gives as output a binary prediction about whether a TW can propagate or not. As such, we quantified its performance using well-established concepts for evaluating classifiers from machine learning. In particular, we look at the *accuracy, precision* and *recall* of the predictor for all the six cellular dialogues and corresponding waveforms we found. These are defined as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$precision = \frac{TP}{TP + FP}$$
$$recall = \frac{TP}{TP + FN}$$

Here TP = true positives, TN = true negatives, FP = false positives, FN = false negatives. True positives are parameter sets for which the TW propagates according to both theory and simulation. True negatives are parameter sets for which according to both theory and simulation TWs cannot propagate. False positives are predicted to be capable of sustaining TWs by the theory, but turn out not to do so in an actual simulation. False negatives are parameter sets that are capable of propagating TWs, but are missed by the theory.

Assessment of the analytic framework (Fig. S5) For all of the networks that yielded waves, we find that the theory correctly predicts plane TWs to an extremely high degree of accuracy, close to 100% (Fig. S5A). This means that the theory correctly predicts whether a wave can or cannot propagate in almost all cases. In contrast, the precision and recall take slightly lower scores, with a precision is between roughly 0.6-0.8 and a recall in the range of roughly 0.5-0.7. The interpretation is that roughly 60-80% of conditions predicted to allow TW propagation are indeed ones that can propagate a TW in an exact simulation, and that 50-70% of the all the conditions for TW propagation are correctly identified by the classifier. These lower values are caused by the low number of actual positives (conditions under which a TW can propagate), which is low compared to the total number of parameters we examined. This means that the few incorrect predictions that arise from the approximation have a relatively large impact on these performance metrics.

Overall, we thus obtain a good estimator for traveling wave propagation that is accurate except near the boundary of the regions permitting wave propagation. This also become apparent when we plot the interaction parameters of the predicted and actual waves together with the theoretical bounds (Fig. S5B). This shows that the false predictions (false positives and false negatives) are mainly due to slight misestimations of the boundaries of the regions allowing for TW propagation. In particular, both the upper bound and lower bound for  $K^{(ij)}$  are slightly underestimated, meaning that the estimations for the mean-field contribution  $Y_{MF}^{(i)}$  are underestimated. This is evident from the fact that most false positives are near the lower bound for  $K^{(ij)}$  and most false negatives are near the upper bound for  $K^{(ij)}$ .

Validity of the nearest neighbor approximation The accuracy of the nearest neighbor approximation depends on how much of the total interaction the nearest neighbors capture. The more the nearest neighbors contribute to the total interaction strength, the more accurate the approximation is. This is because all deviations between the theory and the exact model come from the mean-field approximation, which has only a marginal contribution if the sensed concentration is mostly due to the cell itself and its nearest neighbors. We can quantify this by comparing  $f_{NN}^{(i)} \equiv 6f_{nn}^{(i)}$  (since there are six nearest neighbors) to the total interaction strength  $f_N^{(i)}$  for signaling molecule *i*. If  $f_{NN}^{(i)} \approx f_N$ , then the cells beyond the direct neighbors have only marginal influence on the concentration a cell senses. However, if  $f_{NN}^{(i)} \ll f_N$ , then the nearest neighbor approximation will be comparatively inaccurate, because we take into account the rest of the cells in an averaged manner only and neglect their spatial positions.

Note that  $f_{NN}^{(i)}/f_N$  depends on the parameters N,  $a_0$  and  $\lambda^{(i)}$ . By examining how this quantity depends on these parameters, we get a picture of when the NNA is most accurate. For weak interaction (high  $a_0$ ), the nearest neighbor approximation matches closely with the actual system. For stronger interaction, the nearest neighbor approximation becomes worse (Fig. S5C). In this case, one possible solution would be to extend the analysis to next-to-nearest neighbors, as we will discuss in the next part. In contrast, the ratio is hardly dependent on system size N and diffusion length  $\lambda^{(i)}$  (Fig. S5C). Finally, longer diffusion length implies comparatively more influence from cells further away, leading to less accuracy for the nearest-neighbor approximation. However, this effect is also weak, accounting for less than 30% variation in interaction strength.

We also considered how taking next-to-nearest neighbors into account improved the accuracy of the analytic approach (Fig. S5C - dotted lines). The contribution from next-to-nearest neighbors can be quantified by an interaction parameter  $f_{NNN}^{(i)} \equiv 4f^{(i)}(\sqrt{3}a_0) + 8f^{(i)}(2a_0)$ , which takes into account the interaction with the twelve cells in the second layer surrounding a cell on a hexagonal lattice. There are four cells at a distance of  $\sqrt{3}a_0$  and eight cells at a distance of  $2a_0$  in this layer. We find that total contribution to the interaction strength from the two layers of cells closest to a given cell,  $(f_{NN}^{(i)} + f_{NN}^{(i)})$  indeed would improve accuracy, but the effects become less significant as  $a_0$  becomes larger. Thus, extending our framework to include next-to-nearest neighbors would improve the accuracy of our method, but as a NNA already perform well, it seems unnecessary to make our framework more complicated.

## S3.6 Oscillatory traveling waves

So far, we focused on traveling waves that are characterized by a propagating pattern on a fixed background. However, we also observed a variety of dynamic spatial patterns with oscillatory background cells in Networks 16, 20 and 43 (Fig. S3A). In particular, oscillatory traveling waves (Sec. S1.3) form a subset that we can analyze using our framework. Here, we outline how to adapt our framework to the analysis of these dynamic patterns. From simulations, we observe that the oscillations have period 3 and always follows a fixed pattern. Using the definition of the Exterior (E), Front (F), Middle (M) and Back (B) states of a wave (S3.3.3) and Fig. 4), we can trace out how these states transition on a state diagram (Section S2). We found that the cells of oscillatory waves follow a fixed pattern of cell state transitions (Fig. S3D). Networks 16 and 20 each have one distinct state diagram, and network 43 can generate waves that follow either of the two state diagrams (Fig. S3B). Each of the states undergoes a separate period 3 oscillation, but together they follow an regular pattern on the state diagram. At the transition with the dotted lines, the wave moves one step further (i.e., the entire pattern not only oscillates but also moves one cell layer ahead). This occurs for transitions where both genes are switched, either between (0,0) and (1,1) or between (0,1) and (1,0). Waves in network 43 can follow either the transitions of network 16 or those of network 20, or show more complicated patterns which fall outside these two standard cases. Hence, by imposing each of the transitions on either the same group of cells (when the wave oscillates but not moves) or a neighboring group of cells (when the wave translates), we could derive a more complicated set of constraints for the propagation of such waves if necessary.

# S4 Robustness and reliability of traveling waves

# S4.1 Robustness (Fig. S6A-C)

Biological robustness is typically referred to as the ability of a biological system to adapt to environmental perturbations by maintaining its function [4]. Control mechanisms such as feedback loops may play a role in maintaining robustness. In our system, different time scales allow us to study the concept of robustness at different levels. The dynamics of the parameters of the system occurs at an evolutionary time scale (unless the experimentalist intervenes), while the dynamics of the gene expression happens on a much shorter time scale (minutes to hours) and the dynamics of the signaling factors occurs on an even faster time scale. As such, we may consider perturbations at each of these levels of description to see how they affect the system's ability to perform a certain function - which in this case means it's ability to generate patterns. In this paper, we considered the system's response to changes in parameter values. Since we assume the parameters to stay constant during the entire simulation, we will use robustness as a static quantity obtained by comparing simulations at different sets of fixed parameters.

Specifically, we considered the robustness of traveling waves for two different situations. We considered the robustness of TW formation - how changing parameters impacts the system's ability to self-organize into a TW, as well as the robustness of TW propagation - how parameters influence the ability of an already formed TW to continue propagating. We quantified the robustness in both cases by the fraction of parameter sets, or Q-value, that can generate or propagate a TW [5, 6]. In the absence of further information about the parameters, this tells us how likely it is to find parameters which are compatible with a certain property or behavior of the system - formation or propagation of TWs in our case.

**Normalized Q-value** The Q-values we obtain as a fraction of parameter set compatible with TWs depend on the number of parameters m we sample over for each network. These values will tend to be higher for networks with fewer parameters than for networks with higher number of parameters. One method used to correct for this is to take the m-th root of the Q-value [5]. We will call this the "normalized Q-value". This value represents the chance for each of the m parameters to be compatible with TW formation over a specified range of values.

Calculation of Q-values from simulation data In the absence of any predictive theory, the phase space volume compatible with TWs can only be estimated through drawing random samples from the parameter space and determining whether TWs can form or propagate for each of these samples. Note that in principle our parameters are unbound, i.e.  $K^{(ij)} \ge 1$  and  $C_{ON}^{(j)} \ge 1$  with no upper bound. To calculate the robustness, we therefore specified a finite region defined by  $1 \le K^{(ij)} \le L$  and  $1 \le C_{ON}^{(j)} \le L$  for each signaling molecule j and each interaction  $i \leftarrow j$  (neglect the parameters for non-existent interactions). In practice, we took L = 1000 everywhere. We then used Latin hypercube sampling to generate a large number of parameter sets and tested whether TWs could form or propagate for each of the parameter sets.

For TW formation, we tested how likely it is to find self-organization of TWs in the five networks for which we found self-organized TWs (Fig. 3D). We used the same 10,000 sampled parameter sets as were used to generate the initial network classification (Figs. 3 and S4). For each parameter set, we considered it to be capable of self-organizing TWs if at least one simulation (out of 10 runs per parameter set) led to a self-organized TW. The Q-value we obtained for TW formation in this way is of the order of  $10^{-3}$  (Fig. S6B — upper figure). Alternatively, after correcting for the number of parameters, the normalized Q-value corresponds to a randomly generated parameter having around 30% chance of taking a value compatible with TW formation across a 1,000-fold range for each parameter (Fig. S6B — lower figure). The Q-values obtained in this way are in fact lower estimates as we perform only a finite number of simulations and would be higher if we could screen over all possible initial states (in which case the Q-values for TW formation and TW propagation would coincide, since we would also include the final pattern as initial state). Nevertheless, this approach mirrors the situation in wet lab experiments, where can only test a finite number of replicates before concluding that a particular result is highly unlikely to be reached.

For TW propagation, we tested the two types of TWs we found (Fig. 4D) for each of the networks in which we found them. We used the same data as obtained from Latin hypercube sampling which we used to quantify the performance of our analytic predictor (Fig. S5), as this contains precise information on whether each parameter set should be able to propagate TWs according to both the theoretical prediction as well as explicit simulations. This gave higher Q-values, in the order of  $10^{-2}$  for TW propagation (Fig. S6A — left figure), corresponding to normalized Q-values of about 40 - 50% for each parameter to be compatible (Fig. S6A — right figure). In comparison, the robustness of the *Drosophila* segment polarity gene network was quantified for a network with a far larger set of parameters, for which the authors found Q-values corresponding to normalized values of about 80 - 90% for each parameter [5].

#### S4.2 Robustness: analytical calculation (Fig. S9)

We can also obtain an estimate for the robustness of TWs through the analytically derived conditions (Section S3.3.3), which we will apply to the explicitly derived conditions for network 15 (Eqs. S28). The derived inequalities define a region  $\mathcal{U} \subset \mathcal{P}_L$  that is compatible with TW propagation. The volume of this region,  $V(\mathcal{U})$ , in relation to the total phase space volume  $V(\mathcal{P}_L)$  defines the robustness, i.e. we can express the Q-value as

$$Q = \frac{V(\mathcal{U})}{V(\mathcal{P}_L)}.$$
(S30)

The volume over the region compatible with traveling waves can be expressed as an integral over  $\mathcal{P}_L$ :

$$V(\mathcal{U}) = \int_{1}^{L} dC_{ON}^{(1)} \int_{1}^{L} dC_{ON}^{(2)} \int_{1}^{L} dK^{(1,1)} \int_{1}^{L} dK^{(1,2)} \int_{1}^{L} dK^{(2,1)} \mathbb{1}_{TW},$$
(S31)

where the function  $\mathbb{1}_{TW}$  takes values 1 on the domain for which the inequalities are satisfied and 0 elsewhere. We have shown that the derived conditions can be reduced to sets of independent conditions on  $K^{(ij)}$  and  $C^{(j)}_{ON}$  for each interaction  $i \leftarrow j$ . This means that the integral decomposes into two separate integrals

$$V(\mathcal{U}) = V^{(1)}V^{(2)}$$

$$V^{(1)} = \int_{1}^{L} dC_{ON}^{(1)} \int_{1}^{L} dK^{(1,1)} \int_{1}^{L} dK^{(2,1)} \mathbb{1}_{TW}$$

$$V^{(2)} = \int_{1}^{L} dC_{ON}^{(2)} \int_{1}^{L} dK^{(1,2)} \mathbb{1}_{TW}$$
(S32)

To evaluate  $V^{(1)}$  and  $V^{(2)}$ , one must take into account the various ways in which the borders defined by the inequalities intersect with the boundaries of the box. The integrals over  $C_{ON}^{(j)}$  can then be split into integral over the various regions defined by these intersections. Concretely, let us define  $C_{min}^{(ij)}$  and  $C_{max}^{(ij)}$  as values of  $C_{ON}^{(j)}$  at which  $K_{min}^{(ij)} = L$  and  $K_{max}^{(ij)} = L$  respectively. Note that  $C_{min}^{(ij)} > C_{max}^{(ij)}$  by definition. For  $V^{(2)}$ , we have three different cases to distinguish: If  $L < C_{max}^{(12)} < C_{min}^{(12)}$ , we have a single integral

$$V^{(2)} = \int_{1}^{L} dC_{ON}^{(2)} \left( K_{max}^{(12)} - K_{min}^{(12)} \right)$$

If  $C_{max}^{(12)} < L < C_{min}^{(12)}$ , we have to split up the integral into two parts:

$$V^{(2)} = \int_{1}^{C_{min}^{(12)}} dC_{ON}^{(2)} \left( K_{max}^{(12)} - K_{min}^{(12)} \right) + \int_{C_{min}^{(12)}}^{L} dC_{ON}^{(1)} \left( L - K_{min} \right)$$

Finally, if  $C_{max}^{(12)} < C_{min}^{(12)} < L$ , we have to split up the integral into three parts, where the integrand over the third part is zero:

$$V^{(2)} = \int_{1}^{C_{min}^{(12)}} dC_{ON}^{(2)} \left( K_{max}^{(12)} - K_{min}^{(12)} \right) + \int_{C_{min}^{(12)}}^{C_{max}^{(12)}} dC_{ON}^{(1)} \left( L - K_{min}^{(12)} \right) + \int_{C_{max}^{(12)}}^{L} dC_{ON}^{(1)} 0$$

In the case of  $V^{(1)}$ , the procedure is a bit more involved, but can be summarized as follows: we first order the values  $C_{min}^{(11)}, C_{max}^{(11)}, C_{max}^{(21)}, C_{max}^{(21)}$ , so that we can split the integration domain of  $C_{ON}^{(j)}$  up into  $m \leq 4$  parts  $[b_0 = 1, b_1], [b_1, b_2], \ldots, [b_{m-1}, b_m = L]$  by taking all of the ordered values which are less than or equal to L. For each integration segment  $[b_{s-1}, b_s]$ , we then distinguish between the same three cases as in the previous calculation, but now with an integrand that is the product of two factors, which we will denote  $F_s^{(11)}$  and  $F_s^{(21)}$  for now (for  $K^{(11)}$  and  $K^{(12)}$  and segment number s). The cases to distinguish are then similar:

1. If 
$$b_i < C_{max}^{(i1)} < C_{min}^{(i1)}$$
, then  $F_s^{(i1)} = \left(K_{max}^{(1j)} - K_{min}^{(1j)}\right)$ .

2. If 
$$C_{max}^{(i1)} < b_i < C_{min}^{(i1)}$$
, then  $F_s^{(i1)} = \left(L - K_{min}^{(1j)}\right)$ .

3. If 
$$C_{max}^{(i1)} < C_{min}^{(i1)} < b_i$$
, then  $F_s^{(i1)} = 0$ .

The integral is then calculated as

$$V^{(2)} = \sum_{i=0}^{m} \int_{b_i}^{b_{i+1}} dC_{ON}^{(1)} F_s^{(11)} F_s^{(21)}.$$
 (S33)

**Maximal robustness** Equipped with an explicit expression for the robustness of TWs, we can now study how the robustness changes with parameters, and in particular how to maximize the robustness by tuning parameters of the system. We quantified the robustness through a Q-value calculated over parameters describing the parameters of the gene circuit  $(K^{(ij)} \text{ and } C_{ON}^{(j)})$ , but this value will depend on other parameters of the system. In particular, the expressions of the boundaries  $C_{min}^{(ij)}$  and  $C_{max}^{(ij)}$  already reveal that the Q-value directly depends on  $f_{nn}^{(i)}$  and  $f_N^{(i)}$  (Eqs. S29 and S21). These quantities in turn depend on N,  $a_0$  and  $\lambda^{(i)}$  (i = 1, 2). Hence, we will study how the Q-value changes with these four parameters, and in particular whether there are global maxima for the Q-value as a function of these parameters. Note that we can reduce the number of parameters to three by normalizing the diffusion lengths,  $\lambda^{(i)}$  (see STAR Methods).

We first numerically sampled over N,  $a_0$  and found a consistent dependence of the robustness on  $a_0$  for different values of the system size (Fig. S9A-B). The Q-value first rapidly increases with  $a_0$ , reaches a maximum and then decays towards zero (Fig. S9B). Intuitively, this trend can be understood as a balance between only self-signaling (at large  $a_0$ ) and excessive communication (at small  $a_0$ ). Namely, at large  $a_0$ , the cells only sense their own signaling molecules. But we have seen that traveling waves are emergent phenomena that rely on cell signaling to realize different dynamics for cells with the same states, based on their location in relation to the wave. Thus, as  $a_0 \to \infty$ , necessarily the robustness of the waves goes to zero. Mathematically, this is evident from the fact that  $f_{nn} \to 0$  as  $a_0 \to \infty$ , so that locally different neighborhoods now exhibit the same dynamics. Conversely, when  $a_0$  approaches zero, the Q-value also goes to zero (for large system sizes at least). This is likely because the interaction between the cells becomes too strong, impeding transitions required to turn off genes, which can only occur if the sensed concentrations are low enough.

For very small system sizes (grid size  $\leq 10$ ), the trend differs at low values of  $a_0$ , where interactions beyond nearest-neighbors become important (Fig. S9D). But since traveling waves are only of interest in large enough systems, we will not consider their behavior in very small systems. Conversely, in the  $N \to \infty$  limit,  $f_N$  approaches a constant value and therefore the Q-values also converge to a constant, as is already evident from the trend at the largest grid sizes shown in Fig. S9D.

We then examined the effect of varying the diffusion lengths. We will work with the normalized lengths  $l^{(1)}, l^{(2)}$ , so that our results are independent of  $a_0$ . Since we have previously seen that the results have only weak dependence on N, effectively we can consider the robustness as a function of these two parameters only. A numerical screening across these two parameters reveals the presence of a single global maximum for the Q-value (Fig. S9A), at a value of  $l^{(1)} \approx 2.2, l^{(2)} \approx 0.7$ . This suggests that the robustness of traveling waves can be optimized by appropriately choosing the signaling molecules (in particular their diffusion lengths) the cells use to communicate with. In the following, we will try to explain the presence of this single maximum through a more detailed analysis.

Area fractions Whereas the Q-value is a measure for overall robustness, we can focus on individual interactions and look at how likely it is to find parameters compatible with TWs for each interaction. We first project the parameter set onto a 2D-plane described by  $K^{(ij)}$ ,  $C_{ON}^{(j)}$ (for an interaction  $i \leftarrow j$ ), as these two parameters together specify the relative strength of the interaction. We then determine area spanned by  $K_{min}^{(ij)}$  and  $K_{max}^{(ij)}$  in this plane and calculate the fraction of this area with respect to the total phase space area considered  $(L-1)^2$  (recall that we assumed that  $K^{(ij)}$  and  $C_{ON}^{(j)}$  span a similar range of values as they have the same units). Formally, we can therefore write

$$a^{(ij)}(L) = \frac{1}{(L-1)^2} \int_1^L dK^{(ij)} \int_1^L dC^{(j)}_{ON} \mathbb{1}_{TW}.$$
(S34)

These area fractions represent the probability of randomly picking the right parameters for each interaction. We then calculated how these area fractions depend on the signaling lengths  $l^{(1)}, l^{(2)}$ . By plotting how these individual area fractions depend on these variables, we obtain a better understanding of how the maximum Q-value arises (Fig. S9E). First, we note that since each area fraction depends only on one signaling length, we can study how they vary with this single parameter (Fig. S9F). We see that the area fraction for i = 1, j = 1 increases with  $l^{(1)}$ , whereas the area fractions for i = 1, j = 2 and i = 2, j = 1 first slightly increase and then decrease with  $l^{(2)}$  and  $l^{(1)}$  respectively. Although the Q-value does not directly decompose into a product of area fractions, we still expect it to be high if and only if all area fractions are relatively high. Thus, to obtain optimal robustness, for the second signaling molecule we simply have to tune its signaling length  $l^{(2)}$  to the maximum of the area fraction  $a^{(12)}$ . For the first signaling molecule, there is a competition between maximizing the area fractions of  $a^{(11)}$  and  $a^{(21)}$ , which generally have opposing trends. Hence, optimal robustness is likely to be found at an intermediate value of  $l^{(1)}$  that is neither to low or too high.

The trend in the area fractions can be further interpreted by explicitly examining the projected areas and analyzing limiting cases explicitly. We first note that in the limits  $l^{(1)} \rightarrow 0$  and  $l^{(1)} \rightarrow \infty$ , the interaction strength becomes

$$\lim_{l^{(i)} \to 0} f^{(i)}(\rho) \to 0,$$

$$\lim_{l^{(i)} \to \infty} f^{(i)}(\rho) \to \frac{r_{cell}}{\rho},$$
(S35)

where  $\rho = r/a_0$  as we recall. Indeed,  $l^{(1)} \to 0$  means that the signaling molecules hardly diffuse anymore, so the interaction between cells becomes negligible. Conversely, if  $l^{(1)} \to \infty$ , it means that the signaling molecule is basically never broken down as it diffuses away from its source. In this case,  $f^{(i)}(\rho)$  reaches a constant that is still distance-dependent, since the same concentration is spread across a larger and larger area (or volume in 3D) as the molecules diffuse away from the source. The  $1/\rho$  dependence implies that the concentration on annulus of inner radius  $\rho$  and width  $d\rho$ ,  $f^{(i)}(\rho) \rho d\rho$ , is independent of distance.

With this knowledge, we are now in a position to understand the dependence of the area fractions on the signaling lengths of the two molecules. Let us do this by considering each of the interactions separately.

First, consider the interaction  $1 \leftarrow 1$ . The bound  $K_{min}^{(11)}$  represents the lowest  $K^{(11)}$  value at which the background cells remain off, so the transition  $(0,0)_E \rightarrow (0,0)_E$  can occur. The bound  $K_{max}^{(11)}$  represents the highest value at which gene 1 can turn ON or remain ON, which is required for the transitions  $(0,0)_F \rightarrow (1,0)$  and  $(1,0) \rightarrow (1,1)$ .

- 1. In the limit  $l^{(1)} \to 0$ , we have  $K_{min}^{(11)} \to 1$  and  $K_{max}^{(11)} \to 1$ , and therefore the area fraction  $a^{(11)}$  goes to  $a^{(11)} \approx 0$  (left plot of Fig. S9F). The system cannot simultaneously keep gene 1 off in the background state (0,0) and turn on gene 1 in the  $E_F$  cells in front of the wave.
- 2. In the limit  $l^{(1)} \to \infty$ , we have  $K_{min}^{(11)} \to 7 + Y_{MF}^{(1)}$  and  $K_{max}^{(11)} \to 5 + 2C_{ON}^{(1)}Y_{MF}^{(1)}$ , and therefore the area fraction  $a^{(11)}$  reaches a constant value (left plot of Fig. S9F). Since both boundaries have relatively large slope in  $C_{ON}^{(1)}$ , the area is moderately small.

Then, consider the interaction  $1 \leftarrow 2$ . Since this is a repressive interaction, the role of the bounds have reversed.  $K_{min}^{(12)}$  represents the lowest value of  $K^{(12)}$  for which gene 1 can be ON (unrepressed), whereas  $K_{max}^{(12)}$  represents the highest value at which gene 1 can still be repressed.

- 1. In the limit  $l^{(2)} \to 0$ , we have  $K_{min}^{(12)} \to 1$  and  $K_{max}^{(12)} \to C_{ON}^{(2)}$ , and therefore the area fraction goes to  $a^{(11)} \approx 1/2$  (middle plot of Fig. S9F). To turn or keep on gene 1,  $K^{(12)}$ cannot be too low. But this is required only when gene 2 is off (for the transitions  $(0,0)_F \to (1,0)$  and  $(1,0) \to (1,1)$ ), in this limit the sensed concentration of gene 2 is very low for these states. Conversely, to turn off gene 1 through repression by gene 2,  $K^{(12)}$  cannot be too high. However, since turning off is only required in states where gene 2 is ON (namely, for the transitions  $(1,1) \to (0,1)$  and  $(0,1) \to (0,0)$ ), and the cell senses a concentration  $C_{ON}^{(2)}$  by default in these states, the threshold  $K^{(12)} \ge C_{ON}^{(2)}$ . Together, this leaves a relatively large area permitted.
- 2. In the limit  $l^{(2)} \to \infty$ , we have  $K_{min}^{(12)} \to 5 + 2C_{ON}^{(2)} + Y_{MF}^{(2)}$  and  $K_{max}^{(12)} \to 2 + 5C_{ON}^{(2)} + Y_{MF}^{(2)}$ , and therefore the area fraction  $a^{(12)}$  reaches a constant value (middle plot of Fig. S9F). Since both boundaries have relatively large slope in  $C_{ON}^{(2)}$ , the area is moderately small.

Altogether, this implies that the interaction  $1 \leftarrow 2$  should favor a relatively small diffusion length  $l^{(2)}$ , as is observed. The analysis does not explain the existence of a maximum between these extremes.

Finally, for the interaction  $2 \leftarrow 1$ , the equations are identical to those for  $1 \leftarrow 2$  and therefore the analysis is similar, with the difference that this is an activating interaction and therefore the interpretations of turning ON and OFF should be reversed. In summary, although this technical analysis is rather involved, in essence the existence of a single maximum in robustness can be understood through a competition between self-interaction and neighbor interaction. Transitions which occur mostly due to self-interaction favor weak interaction with neighbors (low  $l^{(1)}$  and  $l^{(2)}$ ). Conversely, transitions that rely on neighbor interactions (e.g. neighbor-induced gene activation for the interaction) favor strong interaction or high signaling lengths. The competition between these two effects is responsible for creating a single optimal robustness in terms of the signaling lengths.

## S4.3 Reliability (Figs. 5F and S8)

While robustness deals with the probability of finding parameter sets compatible with TWs, we can also ask what the chance of finding a TW is once the parameter set has been fixed. We define the *reliability* of TW formation as the percentage of simulations with varying initial conditions that generate TWs given a set fixed parameters. For each of the sets of parameters that yielded self-organized TWs, we determined the reliability by running a large set of simulations and counting in how many of those TWs spontaneously formed. Overall, we found an average reliability of 0.2-0.4 across all networks, indicating that we expect TWs to form in roughly 20-40% of the time for these parameters sets (Fig. 5F). However, upon closer inspection we find that this average results from a considerable variability between different parameter sets, indicating that the precise choice of parameters has a large influence on the reliability of TW formation (Fig. S8A). While for many parameter sets the reliability is exceedingly low (5-10%), there is a continuum of reliability values all the way up to about 80% (Fig. S8A).

This finding raises the question of whether we could identify any source of this variability in reliability values between different parameter sets. To address this question, we took a large set of parameter sets (n = 2534) capable of sustaining a TW once it has formed, as tested explicitly in simulations starting with a TW. For each of these parameter set, we then ran a large number of simulations (100) to see whether it could also self-organize into traveling waves if we set up the initial configuration to be random. Surprisingly, we found that a large set of these parameter sets did not yield self-organized waves at all (Fig. S8B). This indicates the system may be able to propagate a pattern, but have only few ways of generating such a pattern in the first place. Between the parameter sets that were found to self-organized TWs, the reliability varies dramatically along a continuum between 0 - virtually no simulations become TWs - to close to 1 - almost all simulations become TWs. When we plot the distribution of the 578 parameter sets found to generate TWs, we find that the probability to find a given reliability decays nearly monotonically (Fig. S8C), indicating that parameter sets with higher reliability increasingly rare. However, when we examined the reliability of these different parameter sets as a function of the parameters, we observed no clear trend or correlation in two different projections of the parameter sets (Fig. S8D-E). The parameters sets with extremely high reliability values are scattered around the entire region in which TWs are possible (Fig. S8D). Furthermore, for any of the parameters, there are parameter sets with high reliability for both very high and very low values of that parameters, and the same applies to low reliability (Fig. S8E).

# S4.4 Influence of initial conditions (Fig. S6D)

In our deterministic model, the initial state of the system fully determines whether a TW forms or not, even though the link between initial and final state is typically unclear unless one runs an actual simulation. It is clear that not all initial states lead to TWs even when suitable parameters are chosen. As a counterexample, consider initiating the system as a uniform lattice of cells that each have the same state. Since all cells sense the same concentration, each cell will evolve to the same state. More generally, in a deterministic cellular automaton there is no symmetry breaking mechanism that can produce a pattern with a set of symmetries from an initial state with a different set of symmetries.

We therefore studied whether certain features of the initial states had a significant impact on whether TW formed or not. In particular, we looked at the contribution of a few statistical variables characterizing the initial states - the initial mean expression level of the genes  $(p^{(1)}, p^{(2)})$ and the initial spatial order of both genes  $(I^{(1)}, I^{(2)})$ . This is because even for moderately large systems, the number of states exceeds the computational limits of ordinary computers (e.g. for a lattice with N = 100 cells, there are  $4^N \approx 10^{60}$  states), making it impossible to exhaustively simulate all initial states. We found that the fractions of cells with either of the genes ON had a significant impact on whether TWs formed or not, while the initial spatial order had a notable but much smaller impact.

Initial fractions of active genes (Fig. S6D — left plot) We studied the influence of the initial fractions of ON-cells for both genes by generating states with fixed values of  $(p^{(1)}, p^{(2)})$  - this can be easily done by randomly drawing a fixed number of cells to be ON for both genes. Our first observation is that no TWs form for extreme values of  $p^{(1)}, p^{(2)}$ . For instance, if we start with very low initial p values we cannot produce a TW. Both fractions of ON-genes will go to zero and the system will go to the homogeneous state. In contrast, the highest probabilities to produce a TW was found for intermediate values of  $p^{(1)}, p^{(2)}$ , and can reach a maximum of more than 80%. This is remarkable, since it implies that for the given circuit is capable of generating TWs almost certainly if one initiates the system with the given parameters and fractions of genes that are ON, regardless of how the cells that have these genes ON are placed in space. In particular, it suggests that one can improve the maximal reliability of the system by placing constraints on the initial states.

These findings were confirmed in simulations using other parameters sets capable of generating TWs (not shown). While the exact numbers differ between parameter sets, we consistently observed that more moderate levels of  $p^{(1)}, p^{(2)}$  close to (0.5, 0.5) had higher probabilities to generate TWs.

Initial spatial order (Fig. S6D — right plot) In a similar vein, we studied the effect of the initial amount of spatial clustering on TW formation by running simulations where the initial  $I^{(1)}$  and  $I^{(2)}$  were varied (using the approach illustrated in Figure S1 and described in STAR Methods). We observed highly similar results across the range of initial I values, but found that fraction of TWs formed tended to be lower at lower values of I.

| Metric | Accuracy | Precision | Recall |
|--------|----------|-----------|--------|
| Value  | 0.562    | 0.565     | 0.835  |

Table 4: Performance metrics of the logistic model fitted to the data shown in the right plot of Figure S6D. The logistic model is used as classifier on the same data as it was fitted on.

In order to draw more statistically rigorous conclusion about the results displayed in Fig. S6D (right plot), we fitted a logistic regression model to the simulation data using the statistical programming language R. A logistic regression model uses one or multiple predictor variables to calculate a probability that a sample belongs to one out of two possible classes. In this case we can use the initial values of  $I^{(1)}$  and  $I^{(2)}$  to predict if a TW forms (classes: TW, no TW). First, we determined that the logistic regression model using the information of both I values (the proposed model) was significantly better than the null model, which takes uses only information about the values of  $(p^{(1)}, p^{(2)})$ . This indicates that there is information in the I values, and knowing the initial I has a non-negligible influence on whether TWs form or not. However, the *residual deviance* (indication of goodness of fit of the model based on the log-likelihood of the data given the model) is hardly reduced when going from the simple null model to the more complicated proposed model. This means that including the initial I only marginally improves the quality of the predictor.

Next, the proposed model is used as a classifier. For each of the simulations the logistic model uses the initial *I*-values to predict whether the simulation will result in a TW or not. More specifically, the logistic model gives a probability that initial *I* values ( $I^{(1)}$  and  $I^{(2)}$ ) will result in a TW. If this probability exceeds 0.5, then we say that the model predicts a TW.

The classifier always predicted a wave, except for the relatively low I values that correspond to the lower wave fractions observed in simulations. The overall performance of the classifier was further assessed and the results are in Table 4, using the same metrics as used in S3.5. The performance metrics indicate that the logistic classifier performed slightly better than random (accuracy > 0.5). Most TW were correctly predicted (high recall), but this is only because the classifier mostly predicts waves. Many simulations that did not yield a TW were incorrectly labeled as TW (low specificity). It should be noted that the classifier was used to predict the data it was trained on, leading in general to overestimation of the performance metrics. Altogether, this implies that the influence of initial spatial order on the formation of TWs is only marginal.

#### S4.5 Stability of TWs (Fig. S14)

We also investigated the robustness of TWs to perturbations in individual cell states and the sensitivity to initial conditions for TW formation. In both cases, we perturbed a randomly chosen set of cells, by assigning them a random other state with equal probabilities for all states. We then looked at whether a TW could still propagate or form and the time required to reform the TW.

We first took TW configurations as initial conditions and applied perturbations of varying numbers of cells to the system. The results show that the fraction of simulations that reform into TWs drops with the number of perturbed cells, but retain relatively high values (typically > 0.5) even when a majority of cells are perturbed (Fig. S14A). This shows that when the parameter conditions are right, random perturbations to an already formed wave can still lead to reforming of TWs, even when the perturbations are large. The time it takes to reform a wave is small when the perturbations are small, retaining a value far below the original formation time of the wave for small perturbations, for the cases studied. From both the probability of reforming a wave as well as the reforming times, we deduce that for our case studies, waves are relatively stable up for perturbations up till about 10% (or roughly 20 out of N = 225) of the cells.

The same holds true when we perturb an initial condition that leads to the formation of a TW (Fig. S14B). The probability of forming a wave drops less rapidly in these cases and also retains high values when a large portion of the cells are perturbed. The formation time can be both lower or higher than that of the original wave and does not change appreciably with the number of cells changed.

## S4.6 Persistence of traveling waves with noise (Fig. S12)

Recall that we observe that traveling waves persist with complex elements up to a certain degree (Fig. 6), but that decreasing Hill coefficient or increasing noise, lattice disorder or cell motility beyond certain thresholds will cause the waves to stop propagating. We obtained these results from running simulations with different values for the new parameters the extended model introduces (i.e. noise, Hill coefficient, etc.). While this is a feasible way to obtain information about persistence of TWs, ideally we would like to be able to estimate these effects without running any simulations. To this end, we derived an analytical method to compute the effect of noise on traveling wave persistence.

We study the propagation conditions derived in Section S3.3 under the influence of noise in the form of fluctuating thresholds  $K^{(ij)} \to K^{(ij)} + \delta K^{(ij)}$  with  $\delta K^{(ij)} \sim \mathcal{N}(0, \sigma K^{(ij)})$ , as defined previously. Under our wave decomposition scheme, we have six different transitions that need to be satisfied by a given number of cells. Consider the transition  $\alpha \to \beta = X(\alpha)$  for a transition of a wave state  $\alpha \in \{E_F, F, M, B, E_B, E\}$  to a cell state  $\beta = (\beta^{(1)}, \beta^{(2)}), \beta^{(i)} \in \{0, 1\}$ . The condition under which this transition occurs can be written as

$$\beta^{(i)} = \prod_{j} g^{(ij)}(X(\alpha)), \tag{S36}$$

$$g_{\alpha}^{(ij)} = \begin{cases} \theta(Y_{\alpha}^{(j)} - K^{(ij)}) & M^{(ij)} = 1\\ \theta(K^{(ij)} - Y_{\alpha}^{(j)}) & M^{(ij)} = -1\\ 1 & M^{(ij)} = 0 \end{cases}$$
(S37)

Here  $g_{\alpha}^{(ij)} = g^{(ij)}(X(\alpha))$  and  $Y_{\alpha} = (Y_{\alpha}^{(1)}, Y_{\alpha}^{(2)})$  is the sensed concentration of the cell with wave state  $\alpha$ , which we calculate in the nearest-neighbor approximation (Section S3.3.5). The persistence of the TW requires that the transition conditions are met for each of the N cells

of the system. We next derive the probability that this occurs for a given value of the noise strength  $\sigma$ .

First, we derive the probability that  $g^{(ij)}(X(\alpha)) = 1$  for any interaction and  $\sigma$ . The probability that this holds depends on the interaction type specified by  $M^{(ij)}$ . For  $M^{(ij)} = 0$ , this condition is trivially met, so we consider  $M^{(ij)} \neq 0$ . We then have the general expression

$$P(g_{\alpha}^{(ij)} = 1) = P\left(Y_{\alpha}^{(j)} - K^{(ij)} - \delta K^{(ij)}\right) M^{(ij)} > 0\right)$$

$$= \begin{cases} P\left(\delta K^{(ij)} < Y_{\alpha}^{(j)} - K^{(ij)}\right) & M^{(ij)} = 1\\ P\left(\delta K^{(ij)} > Y_{\alpha}^{(j)} - K^{(ij)}\right) & M^{(ij)} = -1 \end{cases}$$

$$= \begin{cases} D(Y_{\alpha}^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)}) & M^{(ij)} = 1\\ 1 - D(Y_{\alpha}^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)}) & M^{(ij)} = -1 \end{cases}$$

$$= \left(\frac{1 + M^{(ij)}}{2}\right) D(Y_{\alpha}^{(j)} - K^{(ij)}) + \left(\frac{1 - M^{(ij)}}{2}\right) \left(1 - D(Y_{\alpha}^{(j)} - K^{(ij)})\right).$$
(S38)

Here  $D(Y_{\alpha}^{(j)} - K^{(ij)}; 0, \sigma K^{(ij)})$  is the cumulative distribution function of the normal distribution with mean 0 and standard deviation  $\sigma K^{(ij)}$  evaluated at  $Y_{\alpha}^{(j)} - K^{(ij)}$ . Because of the AND-logic we impose, and because the noise terms are independent for each (i, j), the probabilities for the final state can be written as

$$P_{\alpha}(\beta^{(i)}) \equiv P(\beta^{(i)}\alpha) = \begin{cases} 1 - P(g_{\alpha}^{(i1)} = 1)P(g_{\alpha}^{(i2)} = 1)) & \beta^{(i)} = 0\\ P(g_{\alpha}^{(i1)} = 1)P(g_{\alpha}^{(i2)} = 1)) & \beta^{(i)} = 1 \end{cases}$$
(S39)

$$= \beta^{(i)} \prod_{j} P(g_{\alpha}^{(ij)}) + (1 - \beta^{(i)})(1 - \prod_{j} P(g_{\alpha}^{(ij)})).$$
(S40)

By inserting Equation S38 into Equation S40 we obtain transition probabilities at the single-cell level in terms of the sensed concentrations  $Y_{\alpha}$ , thresholds  $K^{(ij)}$  and noise levels  $\sigma$  (for a given transition  $\alpha \to \beta$ ). In order for the entire wave to propagate, the transitions must be satisfied for each gene *i* and each of the cells in the system. Let  $n_{\alpha}$  be the number of cells of type  $\alpha$  in our system. For a single, straight plane wave, we have  $n_{\alpha} = \sqrt{N}$  for  $\alpha = E_F, F, M, B, E_B$  and  $n_{\alpha} = N - 5\sqrt{N}$  for  $\alpha = E$ . We can then write the probability that the wave survives for one time step as

$$P_{survival} = \prod_{\alpha} P(\alpha \to \beta)^{n_{\alpha}} = \prod_{\alpha} \prod_{i=1}^{L} P_{\alpha}(\beta^{(i)}) \right)^{n_{\alpha}}.$$
 (S41)

The probability that the wave survives for t time steps is then  $(P_{survival})^t$ .

Accuracy of computed survival probability The accuracy of the computed survival probability mostly depends on the validity of the nearest-neighbor approximation. Recall that we estimate the signal molecule concentration from all cells beyond nearest neighbors through a mean-field approximation term  $Y_{MF}$  (Equation S21). This approximation is the reason why the computed TW propagation conditions are not exact, and as a result the survival probabilities are also not exact. However, this only leads to significant discrepancies when either (1) the

interaction strength is very high, i.e. when cells interact strongly with each other  $(a_0 \text{ small})$ , or (2) when the original parameter set is close to the boundary of the region where TWs can propagate. In the second case, the problem arises when the calculated boundary of the 'TW propagation phase' is not accurate. In this case, the noise strength required to perturb the system beyond the boundary cannot be accurately estimated, which in turn leads to the survival probability being inaccurately computed. When the system is far away from the boundary, these deviations are proportionally smaller and as a result the order of magnitude estimation of the required noise strength to perturb the wave is more accurate.

Application to other complex elements Similar arguments can in principle be derived to study the effect of disordered cell positions and cell motility, while for finite Hill coefficient at this point we lack an analytical framework for wave propagation. In essence, when only the positions of the cells are altered, this reflects only in the interaction terms  $f^{(i)}(r_{ij})$  which are functions of the distance between cells  $r_{ij}$ . At first approximation, for small deviations only the nearest neighbor terms  $f_{nn}^{(i)} = f^{(i)}(a_0)$  are affected. Hence one can estimate the effect of spatial rearrangement of the cells through the effect on the nearest-neighbor interaction strengths  $f_{nn}^{(i)}$ , which now typically become different for individual cells if the rearrangement process is stochastic. These in turn can be directly computed from distributions of nearest-neighbor distances obtained from the stochastic rearrangement process.

However, in practice, even for simple Brownian motion of the cells the expressions for distributions of nearest-neighbor distances yield unwieldy mathematical expressions in terms of special functions such as Bessel functions. Furthermore, while the effects are straightforward to estimate for a pair of moving cells, it is not trivial whether the pair-wise calculation can be extended to a full lattice. As such, we have not attempted to fully derive analytical results regarding the effect of moving cell positions on traveling wave propagation, but merely want to point out that such a calculation is in principle possible.

# References

- Maire, T. & Youk, H. Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations. *Cell Systems* 1, 349 –360 (2015).
- 2. Thomas, R. & D'Ari, R. Biological Feedback (CRC Press, Inc., 1990).
- 3. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology* **13**, 497 (2015).
- 4. Kitano, H. Biological robustness. *Nature Reviews Genetics* 5, 826–837 (2004).
- 5. Dassow, G. v., Meir, E., Munro, E. M. & Odell, G. M. The segment polarity network is a robust developmental module. *Nature* **406** (2000).
- Ma, W., Lai, L., Ouyang, Q. & Tang, C. Robustness and modular design of the Drosophila segment polarity network. *Molecular Systems Biology* 2 (2006).