A comprehensive network atlas reveals that Turing patterns are common but not robust

Natalie S. Scholes,[1,4] David Schnoerr,[1,4] Mark Isalan,[1,2] Michael P.H. Stumpf ,[1,3,5,*]

[1]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.
[2]Imperial College Centre for Synthetic Biology, Imperial College London, London, SW7 2AZ, UK.
[3]School of BioScience and School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia.
[4]These authors contributed equally
[5]Lead contact
[*]Correspondence: m.stumpf@imperial.ac.uk

Summary

Turing patterns (TPs) underlie many fundamental developmental processes, but they operate over narrow parameter ranges, raising the conundrum of how evolution can ever discover them. Here we explore TP design space to address this question and to distill design rules. We exhaustively analyze 2- and 3-node biological candidate Turing systems, amounting to 7,625 networks and more than 3×10^{11} analysed scenarios. We find that network structure alone neither implies nor guarantees emergent TPs. A large fraction (>61%) of network design space can produce TPs, but these are sensitive to even subtle changes in parameters, network structure and regulatory mechanisms. This implies that TP networks are more common than previously thought, and evolution might regularly encounter prototypic solutions. We deduce compositional rules for TP systems that are almost necessary and sufficient (96% of TP networks contain them, and 92% of networks implementing them produce TPs). This comprehensive network atlas provides the blueprints for identifying natural TPs, and for engineering synthetic systems.

Introduction

Pattern formation is an essential aspect of development in biology and we have a wealth of examples how complex, structured, multi-cellular organisms develop from single fertilized cells. Many organisms develop complex spatial features with exquisite precision and robustness, and this has been the subject of extensive molecular and theoretical study (Green & Sharpe, 2015; Maini, Woolley, Baker, Gaffney, & Lee, 2012).

Various mechanisms have been proposed to explain developmental patterning processes, ranging from maternally inherited cues (Wolpert, 1969), to mechanical forces (Howard, Grill, & Bois, 2011) and chemical reaction-diffusion networks or Turing patterns (TPs) (Gierer & Meinhardt, 1972; Turing, 1952). The latter were first proposed by Alan Turing in 1952 (Turing, 1952), and were later independently described by Gierer and Meinhardt (Gierer & Meinhardt, 1972). TPs are particularly intriguing because they are capable of generating entirely self-organized, complex, repetitive patterns of gene expression (Figure 1A, B).

TPs generally alter local concentrations of biochemical components, resulting in self-organized spatial patterns such as spots, stripes and labyrinths (Kondo & Miura, 2010). These patterns have unique and useful biological properties: perturbing them results in recovery and re-organization of the patterns ("healing"), as an intrinsic property of the dynamical biochemical interactions. This also implies that if there is variability in size across individuals, the TPs will automatically re-scale themselves, simply adding or subtracting pattern segments in response to different field sizes. This is a valuable property to support changes in size, both within existing populations and over evolutionary time. In addition, TP networks are extremely parsimonious, often employing just two or three biochemical species. This implies that they might be an economical solution for evolution to employ, wherever repetitive self-organizing patterns are needed.

Given these advantages, it is perhaps not surprising that TPs are regarded as the driving morphogenetic patterning mechanisms in many biological systems. These include bone and tooth formation, hair follicle distribution and the patterns on the skins of animals, such as fish and zebras (Raspopovic, Marcon, Russo, & Sharpe, 2014; Sick, Reinker, Timmer, & Schlake, 2006; Jung et al., 1998; Nakamasu, Takahashi, Kanbe, & Kondo, 2009; Economou et al., 2012). However, despite several experimentally verified examples (Raspopovic et al., 2014; Sick et al., 2006), the underlying complexity in biological systems has often prevented identification of the precise molecular mechanisms governing the potentially large number of TPs in nature. A second key problem is that there is a paradox between the apparent widespread distribution of natural TPs and the observation — from mathematical analyses (Gaffney, Yi, & Lee, 2016; Iron, Wei, & Winter, 2004; Palmer, 2004; Meinhardt & Gierer, 2000; Gierer & Meinhardt, 1972) — that kinetic parameters need to be finely tuned for TPs to arise. This raises the questions of how evolution could ever discover such tiny islands in parameter space and, even if it could, how would the resulting developmental mechanisms still occur robustly under noisy real conditions.

One approach to resolve these apparent contradictions is to explore TP systems mathematically. While a rich mathematical literature on Turing patterns exists, the vast majority of studies analyze single, idealized networks with fixed parameters (Gaffney et al., 2016; Iron et al., 2004; Liu, Shi, Wang, & Feng, 2013). Although these studies have significantly increased our understanding of patterning mechanisms, they do not provide general guidelines for either the identification of naturally evolved Turing networks in biological systems, or for synthetic engineering of TP networks (Scholes & Isalan, 2017; Borek, Hasty, & Tsimring, 2016; Carvalho et al., 2014; Duran-Nebreda & Solé, 2016; Boehm, Grant, & Haseloff, 2018; Cachat et al., 2016; Diambra, Senthivel, Menendez, & Isalan, 2015; Cachat et al., 2016). A recent approach which has proved very successful in increasing our understanding of biological design principles is the "network atlas" approach (Babtie, Kirk, & Stumpf, 2014; Ma, Trusina, El-Samad, Lim, & Tang, 2009). In this approach, biological networks that execute a particular function are modeled exhaustively: for example, all 2 and 3-node inducible genetic networks that achieve adaptation behaviour were modeled by Ma et al. (Ma et al., 2009). Similarly, all 3-node networks that form a central stripe pattern in a developmental morphogen signaling gradient were modeled by the group of Sharpe (Cotterell & Sharpe, 2010; Schaerli et al., 2014). Such approaches allow one to compare network and parameter design space (Barnes, Silk, Sheng, Stumpf, & Stumpf, 2011) with the resulting phenotypic map, resulting in an atlas or guidebook for the design principles behind that function. A guidebook of potential mechanisms and design rules for discovering TP networks would help towards solving the problem of characterizing molecular players in natural TP systems. Furthermore, a comprehensive Atlas of Turing network space might shed new light on the problem of how evolution could ever discover and stabilize systems which only ever function in tiny islands of parameter space.

In terms of progress towards creating a TP network atlas, two recent studies have begun to analyze larger sets of TP network topologies and parameters and have made important progress in our global understanding of Turing systems (Zheng, Shao, & Ouyang, 2016; Marcon, Diego, Sharpe, & Müller, 2016). Marcon et al., 2016 find only a small number of networks being stable, and only a tiny fraction exhibiting Turing patterns. However, they study only linear reaction systems with all steady states located at zero. While this may be mathematically convenient, it is known to systematically bias stability analysis and to misrepresent the stability properties of realistic systems (frequently >90% misclassification, Kirk, Rolando, Maclean, and Stumpf, 2015a; Maclean, Kirk, and Stumpf, 2015). Furthermore, the simplified modelling framework allowed "negative concentrations". It is therefore unclear to what extent these results apply to real biological systems with positive concentrations and non-linear regulatory functions. In a further study, Zheng et al. did analyze all 2- and 3-node networks for a more realistic model with positive concentrations and non-linear regulatory functions (Zheng et al., 2016), considering activating competitive and inhibiting non-competitive interactions as regulatory mechanisms. These authors find many more Turing networks than Marcon et al., showing the importance of analyzing biologically realistic models. However, in our current work we find that Zheng et al. still identified only a fraction of the existing Turing networks. This is presumably due to their relatively sparse sampling of potential scenarios.

In this work, we therefore carry out an exhaustive analysis and comparison of different, biologically relevant regulatory functions, as well as comprehensive sensitivity/robustness analyses of Turing networks with respect to parameter variations. While the analysis of a given system w.r.t. its Turing generating capability is conceptually straightforward, it comes with various numerical difficulties and pitfalls, making an efficient and robust automated analysis non-trivial.

For this study we developed code that allows us to perform efficient and robust TP analysis. The code is user-friendly and allows the automated analysis of networks of arbitrary size and types of regulatory functions. We utilised this code together with significant computational resources to perform an extensive analysis of $7,625$ different networks with up to three reacting species, testing them for their ability to form TPs for different parameters and regulation types on a much larger scale than has previously been done (we analyzed approximately 3×10^{11} different scenarios - amounting to 8 CPU years computing time). As summarized in Figure 1D, these are analyzed in terms of (1) the network topology; (2) the regulatory function; (3) the kinetic parameters; (4) the diffusion constants of the different species. We consider both competitive and non-competitive regulatory mechanisms (Figure 1E) and study their quantitative and qualitative differences.

In this way, we systematically explore what proportion of network topologies are capable of generating

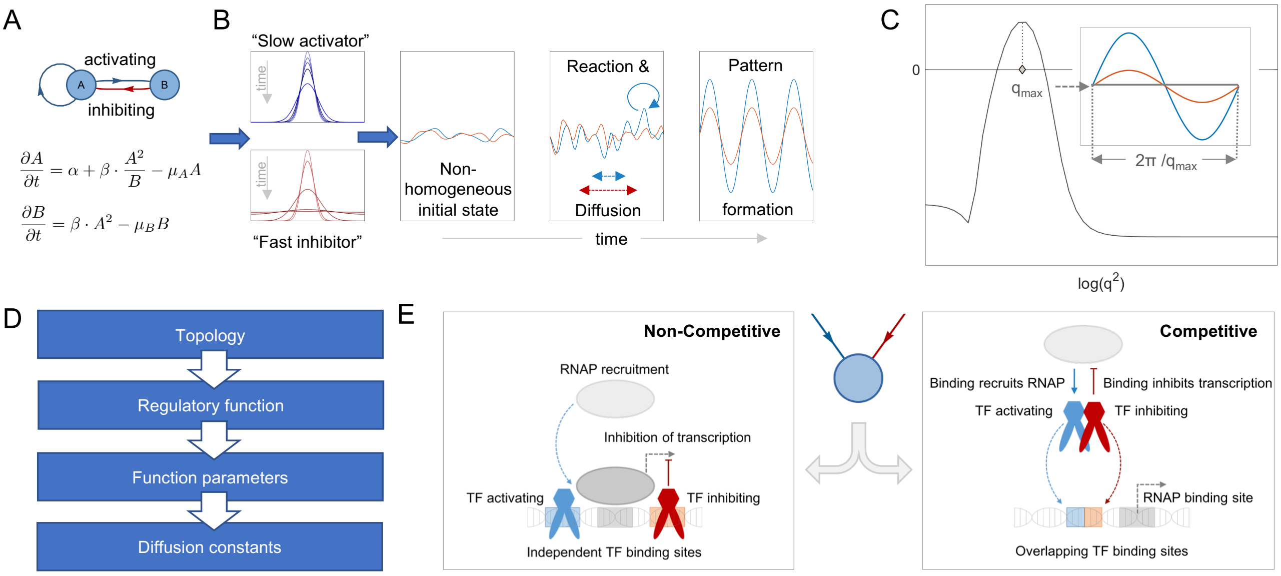


Figure 1: Reaction networks and Turing instabilities. (A) A network graph of the Gierer-Meinhardt model (Gierer & Meinhardt, 1972) as an example for a 2-node Turing network (top), with the corresponding ordinary differential equations below (bottom). Blue and red arrows indicate activating and inhibiting regulations, respectively. Species A activates both itself and species B, while species B inhibits species A. (B) The top left panel represents the diffusion profiles for species A (blue, slow activator) and the bottom panel for species B (red, fast inhibitor). Over time, small deviations in a noisy, non-homogeneous initial condition (Panel 2) can get amplified by the interplay of reactions and diffusion (Panel 3). For the given system this can lead to the formation of stable patterns (Panel 4). (C) An exemplary dispersion relation (real part of the largest eigenvalue of the linearized system as a function of wavenumber q) of the system shown in (A). The wavenumber q_{max} for which the dispersion relation is maximal becomes amplified the strongest. This leads to the formation of a pattern with wavelength $2\pi/q_{max}$ as shown in the inset. (D) In this article we analyze four hierarchical factors determining a network's pattern forming properties: the topology — the species and types of interactions between them; the regulatory function — the functional form of the interactions; kinetic parameters — parameters in the regulatory functions; and the diffusion constants of the different species. (E) Visualization of the two regulatory mechanisms analyzed in this study on a transcriptional level. A species regulated by two other components is represented as a node with two incoming edges (middle panel). We model such multiple regualatory inputs either as competitive or non-competitive: non-competitive regulation describes the case where transcription factors (TF) bind to independent TF sites and thus regulate the recruitment of RNA polymerase (RNAP) and transcription independently (left panel). In the competitive case, in contrast, TFs directly compete for the binding site.

TPs. Moreover, we rank these networks with respect to their robustness to variations in network topology, kinetic parameters and diffusion rates, allowing us to determine which kinds of networks are most robust.

We thus identify a set of irreducible or minimal networks from which all Turing networks can systematically be constructed. This in turn allows us to distill and test compositional rules for predicting whether a given network will support TPs.

With these results in hand we can identify the networks that are most suitable for downstream synthetic engineering under different physiological conditions. There is growing interest in synthetic biology to engineer patterning systems from first principles (Basu, Gerchman, Collins, Arnold, & Weiss, 2005; Schaerli et al., 2014; Borek et al., 2016; Carvalho et al., 2014; Duran-Nebreda & Solé, 2016; Boehm et al., 2018; Cachat et al., 2016). Artificial TPs are expected to have eventual applications in nanotechnology, tissue engineering and regenerative medicine (Scholes & Isalan, 2017; Tan, Chen, Peng, Zhang, & Gao, 2018). Despite much effort in this area, engineered TPs remain elusive. Our comprehensive Turing network atlas contains the "blueprints" required for identifying natural TPs, and guiding the engineering of synthetic systems.

Results

Developing a tractable model to search for TPs

We generate a network atlas in order to explore the design space of TP formation. Our goal is to study the dynamical behavior of spatially-distributed molecule concentrations, and their capability to form stable spatial patterns, across the complete range of 2- and 3-node network topologies. Furthermore, we consider different regulatory functions (competitive and non-competitive Hill functions), over as wide a range of parameters as possible (e.g. diffusion, activation, and repression, etc.) and necessary; we thus forgo mathematical convenience and tackle biologically more realistic models using state-of-the-art computationally intensive, but robust methods.

A network's capability to form patterns is determined by four factors (Figure 1D). First, the network topology needs to be defined. Next, the functional form of the interactions f_i need to be specified. Subsequently, the kinetic parameters need to be specified. Finally, we have to include diffusion.

To define network structure (Figure 2) we only consider connected networks that cannot be split into non-interacting sub-networks. We further exclude networks with nodes that have no incoming or outgoing edges as these are not involved in feedback processes. The influence of such nodes on the rest of the network would thus not have any impact on spatial patterning. Finally, we reduce the number of networks using symmetry arguments (Ma et al., 2009; Babtie et al., 2014), where simply relabeling nodes maps one onto the other.

This network pruning results in 21 and 1934 networks with two and three nodes, respectively. For 3-node systems, we further distinguish which species/nodes diffuse. For a system with two diffusing entities (3N2D networks) this results in 5670 networks (see SI document 6 for a complete list of network graphs), where A and B denote diffusing nodes and C is assumed to be stationary, and in 1934 networks where all three nodes can diffuse.

Having defined the network topologies, we next specify the regulatory mechanisms (Figure 2B). In this work, we use Hill-type functions for the regulation as these have been found to fit well with experimental measurements of gene regulatory networks (Estrada, Wong, DePace, & Gunawardena, 2016; Becskei, Séraphin, & Serrano, 2001; Rosenfeld, Elowitz, & Alon, 2002; Burrill & Silver, 2010; Gardner, Cantor, & Collins, 2000; Ferrell & Machleder, 1998; Ferrell, Tsai, & Yang, 2011; Klumpp, Zhang, & Hwa, 2009). We consider fully non-competitive and fully competitive regulation (c.f. Figure 1E, Figure 2B and STAR methods for equations); the former corresponds to the situation where transcription factors independently regulate the corresponding target gene (Figure 1E, left panel); for the latter case transcription factors compete for shared or overlapping binding sites (Figure 1E, right panel). We also include basal production and a linear degradation term for each species (see STAR Methods).

Regarding the choice of parameter values, the degradation rates and the Hill coefficients were varied over a range of $0.01-1$ and $2-4$, respectively. All other intracellular parameters were varied over the range $0.1 - 100$. Because the resulting number of parameter combinations for a fully-connected 3-node network (3^P networks, where $P = 18$ is the number of parameters) exceeded feasible running times, 3-node systems were further restricted by setting $V = 100$ and $b = 0.1$. These values were chosen because they lead to the largest robustness of 2-node networks w.r.t. the other parameters. Finally, the diffusion rates (i.e. the extracellular parameters) were set to 1 for node A and varied between 10^{-3} and 10^3 for node B and C. The intracellular parameters were varied over three values, while the diffusion constants were varied over seven values. Therefore, we are able to determine with a high degree of certainty whether a system can exhibit the hallmarks of a TP mechanism: (i) stability of the non-spatial dynamics; with (ii) simultaneous instability of the corresponding spatial dynamics (see STAR Methods).

To determine property (i), we need to find stable states of the non-spatial model of a given system. We do so by numerically solving the corresponding ordinary differential equations on a grid of initial conditions (see STAR methods for details). We subsequently verify, by means of linear stability analysis, if the found steady state is indeed stable.

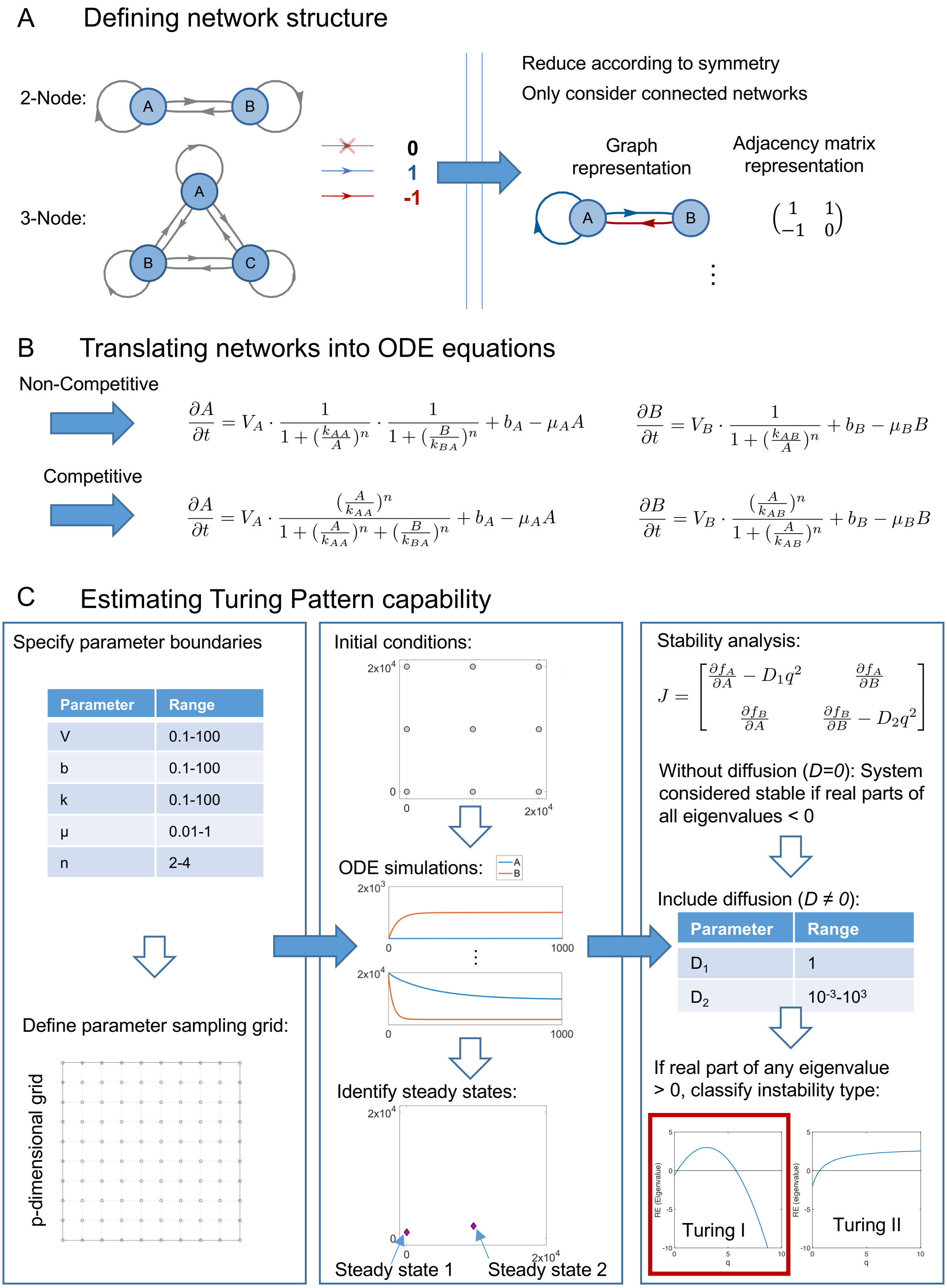
Next, to determine if criterion (ii) is satisfied, we need to assess whether the stable steady state of the non-spatial model becomes unstable if we consider a spatial model with diffusion. To this end, the homogeneous steady state is perturbed by a harmonic wave of a certain wavevector q. The stability can then be determined by the *dispersion relation*: the behaviour of the real part of the largest eigenvalue of the system's Jacobian as a function of q (see Figure 2C and STAR methods for details). If the dispersion relation becomes positive for some finite wavevector q, the system becomes unstable. If it assumes a positive maximum for a finite q_{max}, the system typically gives rise to a pattern of wavelength $2\pi/q_{max}$, and we speak of a "Turing I" instability.

We note that due to the rotational symmetry of spatially homogeneous systems, the dispersion relations of systems of arbitrary dimension are equivalent to the one-dimensional case, which means the systems possess the same stability behaviour. It is hence sufficient to study one-dimensional dispersion relations.

Turing topologies are common but sensitive to regulatory mechanisms

Analyzing all 2-node topologies (21 networks) and 3-node topologies with two diffusors (5670 networks) we find that more than 61% can exhibit Turing I Instabilities and thus we expect them to be capable of generating TPs (Figure 3A). This large number of potential Turing networks is many fold higher than the number of networks identified in the literature to date (≈ 700 topologies) (Zheng et al., 2016; Marcon et al., 2016).

We observe that subtle features beyond network structure influence a network's pattern generating capability. The first difference appears in the choice of



A Defining network structure

Reduce according to symmetry

Only consider connected networks

Graph representation

Adjacency matrix representation

$$\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$$

B Translating networks into ODE equations

Non-Competitive

$$\frac{\partial A}{\partial t} = V_A \cdot \frac{1}{1+(\frac{k_{AA}}{A})^n} \cdot \frac{1}{1+(\frac{B}{k_{BA}})^n} + b_A - \mu_A A \qquad \frac{\partial B}{\partial t} = V_B \cdot \frac{1}{1+(\frac{k_{AB}}{A})^n} + b_B - \mu_B B$$

Competitive

$$\frac{\partial A}{\partial t} = V_A \cdot \frac{(\frac{A}{k_{AA}})^n}{1+(\frac{A}{k_{AA}})^n + (\frac{B}{k_{BA}})^n} + b_A - \mu_A A \qquad \frac{\partial B}{\partial t} = V_B \cdot \frac{(\frac{A}{k_{AB}})^n}{1+(\frac{A}{k_{AB}})^n} + b_B - \mu_B B$$

C Estimating Turing Pattern capability

Specify parameter boundaries

Parameter	Range
V	0.1-100
b	0.1-100
k	0.1-100
μ	0.01-1
n	2-4

Define parameter sampling grid:

Initial conditions:

ODE simulations:

Identify steady states:

Steady state 1 Steady state 2

Stability analysis:

$$J = \begin{bmatrix} \frac{\partial f_A}{\partial A} - D_1 q^2 & \frac{\partial f_A}{\partial B} \\ \frac{\partial f_B}{\partial A} & \frac{\partial f_B}{\partial B} - D_2 q^2 \end{bmatrix}$$

Without diffusion (*D=0*): System considered stable if real parts of all eigenvalues < 0

Include diffusion (*D ≠ 0*):

Parameter	Range
D_1	1
D_2	10^{-3}-10^3

If real part of any eigenvalue > 0, classify instability type:

Figure 2: **Definition and analysis of networks**

(A) Definition of network structure. Left: 2- and 3-node networks containing all possible edges. The latter can be specified as either 0's (no regulation), 1's (activation) and -1's (inhibition). From the set of all possible networks, we identify groups of networks that are equivalent to each other (redundant) and remove all but one network per group. Furthermore, we exclude networks containing any unconnected node(s). The resulting networks can be represented as a network graphs or by their adjacency matrices. The network shown here corresponds to network #8 (see Figure 3A).

(B) Generation of ODEs equations. Each edge in a network corresponds to a Hill term in the ODE equations. The way these terms get combined depends on the regulatory mechanism. In the equations, V denotes the maximum level of expression and b the basal production rate. n is the Hill coefficient, indicating the "steepness" of the regulation. We include a linear degradation term with degradation constant μ for each species.

(C) Workflow for estimating steady states and identifying Turing instabilities. Left panel: for each parameter, a range (see table) is specified and a logarithmic grid generated with three values per parameter. The range of the parameters represents closely the biophysical properties of proteins and samples four orders of magnitude for each of the sampled parameter dimensions. For 3-node systems the parameters V and b were fixed to 100 and 0.1, respectively. These parameters were chosen as they maximized the likelihood of finding a TP solution for all studied 2-node systems. Even then, for fully connected 3-node networks, the remaining parameter combinations amount to 531441. Middle panel: for each parameter set, the corresponding ODEs are solved numerically until time $t = 1000$ for several different initial conditions. We use k-means clustering on the endpoints of the trajectories to find the steady states of the system. Right panel: the final step of the algorithm calculates the eigenvalues of the Jacobian for each steady state. For the Jacobian evaluated at zero diffusion, the real parts of all eigenvalues are required to be smaller than zero, corresponding to a stable steady state. Subsequently, diffusion is taken into account. A Turing instability exists if the real part of the largest eigenvalues becomes positive for some finite wavenumber q. Depending on the behavior of the eigenvalue as a function of the wavenumber q, we classify the instability into two types. In this article we only consider Type I instabilities as only these generate patterns with finite wavelengths.

regulatory mechanism. We find that there are fewer competitive Turing topologies overall among the 2-node networks (Figure 3A). Five networks are detected for non-competitive mechanisms, of which only three are found for competitive interactions (Figure 3B). However, we find that the rarer competitive interactions are more robust to parameter variations than non-competitive interactions, which results in more TP solutions within a given topology. Network #8 constitutes the classical Turing network, which was analyzed by Alan Turing in 1952 (Turing, 1952). The other 2-node networks have also been reported elsewhere (Zheng et al., 2016; Marcon et al., 2016).

For 3-node systems we similarly find Turing topologies that are shared by both regulatory mechanisms (2356 networks, Figure. 3A). While a large fraction of topologies exhibit a Turing I instability, this is again mainly for non-competitive interactions. This is in sharp contrast to the existing literature which mainly highlights the network topology as the deciding factor for TP capability (Diego, Marcon, Müller, & Sharpe, 2017). By contrast, our more detailed analysis shows that network structure alone does not suffice: choice of regulatory function and parameters also critically determine a network's Turing capability.

So far, we have considered systems with either competitive or non-competitive interactions. However, real biological systems may employ combinations of the two regulation types. While it would be instructive to study all possible combinations, this increases the number of networks by two orders of magnitude and we leave this analysis for future work.

Minimal topologies define key properties such as pattern phasing

We next set out to determine key features/topologies required to generate TPs. A similar analysis, for example, revealed that the key motifs to achieve single stripe patterns mediated by external cues (French Flag patterns) are incoherent feed-forward loops (Ingolia & Murray, 2004; Schaerli et al., 2014; Cotterell & Sharpe, 2010).

To identify minimal TP motifs, we create a network atlas in which all networks that differ by a single edge are connected (see atlas for 2-node systems in Figure 3C). Each connection between a network represents the addition/deletion of a single edge, or a change of sign of a single edge. Networks are subsequently sorted hierarchically according to their complexity (here defined as network size). From all networks that can generate TPs, we identify two minimal or "core topologies" for 2-node networks: #8 for the competitive case and #8 and #9 for the non-competitive case. All other Turing networks can be constructed from these by the addition of one or more edges.

In Turing patterns the concentration maxima of the different molecular species are either in phase or out of phase. "In phase" refers to systems in which the maximal concentrations of all species coincide (Figure 1B). By simulating the spatial behaviour (see STAR Methods: PDE simulations), we find that all 2-node Turing networks with competitive regulations give rise to in-phase patterns. In the competitive case, it appears that the networks #15 and #20 inherit the patterning phase from the core network #8 to which they can be reduced.

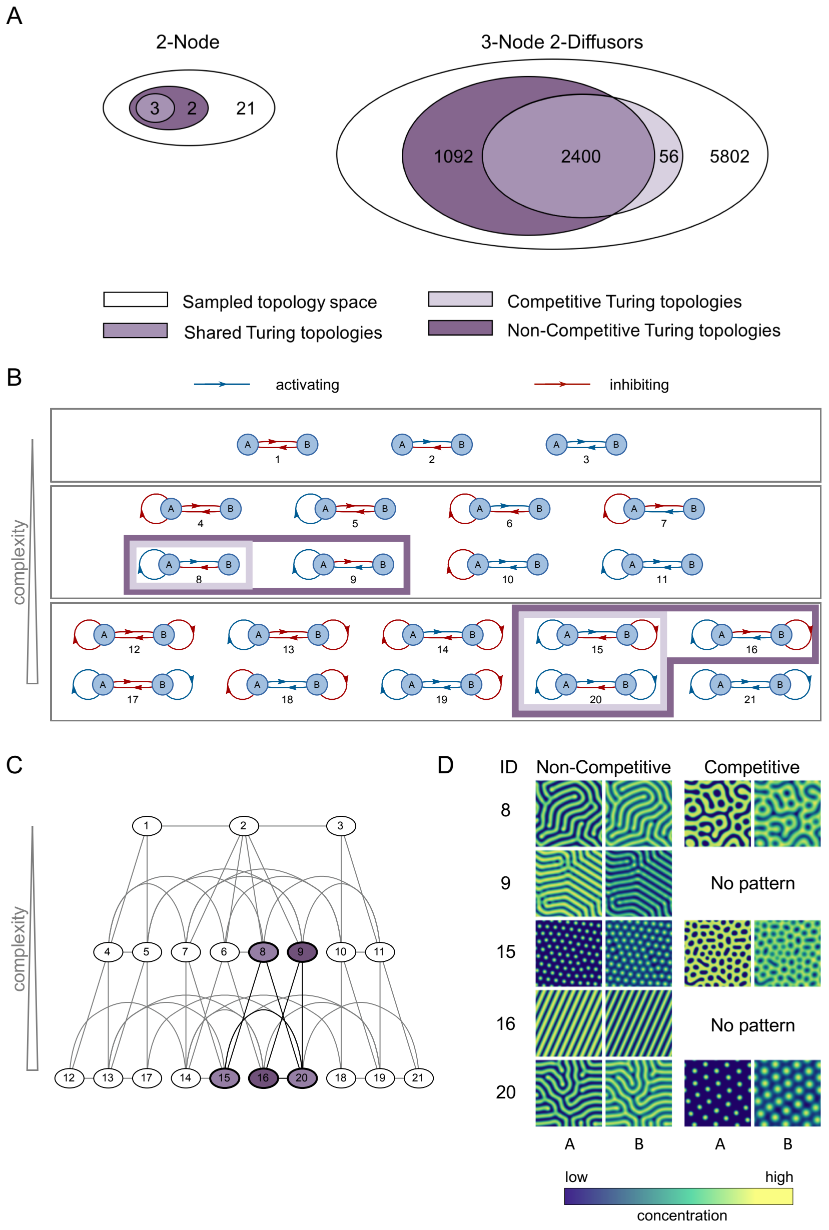


Figure 3: Results for 2-node networks. (A) Visualization of the sampled network topology space and fractions of Turing pattern generators (Turing topologies) for 2- and 3-node networks with 2 diffusing molecules. (B) Considered 2-node networks selected according to criteria described in Figure 2A. Blue and red edges indicate activation and inhibition, respectively. The networks are arranged according to their complexity (that is, the number of edges) with increasing complexity towards the bottom. Each network is given an ID number (1-21). Lilac boxes indicate the networks identified as Turing pattern generators. Dark lilac indicates non-competitive regulatory mechanisms, whereas the lighter shade indicates competitive ones. (C) Hierarchical graph of 2-node networks. Each node represents a network of a given ID number. Networks are connected by an edge whenever they can be transformed into each other by addition, deletion or modification (change of sign) of a single edge. For example network #1 can be transformed into network #4 by addition of a negative self-interaction on node A, and network #1 into network #2 by changing the sign of one edge from inhibition to activation (see networks in (B)). The nodes are colored according to the legend shown in (A), with different shades of lilac indicating for which regulatory mechanism the networks exhibit Turing I instabilities. (D) Representative 2-dimensional patterns for the identified 2-node Turing generating networks. Note that for visualisation purposes the figure shows the patterns generated by species A and B next to each other for each system. The parameters for which the patterns were generated are given in SI document 2.

For non-competitive regulation network #8 again exhibits only in-phase patterning, whereas the second core topology, network #9, shows out-of-phase patterning. One might thus expect network #20 (which can be reduced to either network #8 or network #9 by removal of one edge) to give rise to both in and out-of-phase patterns. Our analysis shows that this is indeed the case and phase can be controlled by the diffusion constants: when A diffuses faster than B, in-phase patterning is observed, whereas if B diffuses faster than A, patterning is out of phase. Key qualitative properties of TPs such as phasing appear to be mediated — but are not determined — by the underlying core topologies; which node corresponds to the "fast" or "slow" species also profoundly affects the TP.

Core topologies also specify phase properties for 3-node networks

In order to identify core topologies, we next apply the complexity reduction procedure to the 5670 different 3-node networks with two diffusing nodes. In the following, these networks will be referred to as 3N2D (numbers refer to amount of either the nodes = N, or the diffusing molecules = D). The resulting hierarchical graph can no longer be depicted due to its large size (> 2400 nodes and $> 10^4$ dependencies). But within this atlas, we find 12 and 20 core topologies for competitive and non-competitive regulation, respectively, all of which have four edges (Figure 4A). As in the 2-node case, the competitive core topologies constitute a subset of the non-competitive systems.

To test the pattern's phasing, we perform PDE simulations for all systems. In-phase mechanisms (Figure 4A bottom) are observed less commonly than out-of-phase mechanisms (Figure 4A top). One might expect that in-phase mechanisms are rarer with three nodes, since more species now have to be in phase. In contrast to the 2-node case, however, some competitive systems now also form out-of-phase patterns. Across the core topologies, we only observe either out-of-phase or in-phase patterns, which suggests that these minimal topologies exhibit unique behaviors.

Analyzing 3-node networks with three diffusing nodes, we find that all such 3N3D Turing networks are also a 3N2D Turing network. We further find the same core topologies for two and three diffusing molecules. This suggests that networks in which three molecules diffuse are mere expansions of the Turing parameter set, but that the instability driving component already exists in the two-diffuser systems. In order to understand the core TP mechanisms for 3-node systems, we can therefore focus on those with two diffusing species.

Two core motifs account for more than 92% of Turing topology space

Analyzing all 3-node core TP topologies, we identify two common core motifs: a positive feedback on at least one of the diffusing nodes, consisting of one or two edges (Figure 4B); and a diffusion-mediated negative feedback loop on both diffusing nodes (Figure 4C).

For the first core motif three possible configurations exist: a direct positive feedback (e.g. network $\#45$) or an indirect positive feedback consisting of either two positive or two negative edges to another node. The interaction can either be mediated via the other diffusing node (e.g. network $\#60$) or the non-diffusing node (e.g. network $\#123$). The second core motif consists of a negative feedback loop on one of the diffusing nodes that is mediated through the other diffusing node. This motif can be realized in different ways and some examples are shown in Figure 4C.

Having identified these two core motifs, we reanalyze all 2-node networks with respect to these motifs. We find that all 2-node Turing networks do indeed possess both core motifs, and all networks containing both motifs exhibit Type I instabilities in the non-competitive case (Figure 3B). For the 3-node networks, we find that 96% of all networks containing both core motifs do exhibit Turing I instabilities. Since we can only sample a finite number of parameters, we cannot categorically rule out that the missing 4% might also exhibit a Turing I instability for certain parameters (and for each case we analyzed another 10^5 randomly sampled parameter sets per network, which again failed to result in TPs). Analyzing all networks exhibiting Turing I instabilities, we observe that 92% possess both core motifs. We confirm by simulation that the remaining 8% topologies do indeed give rise to TPs despite the absence of the core motifs. Therefore, the two identified core motifs together constitute an *almost necessary* and *almost sufficient* criterion for Turing I instabilities for 2-node and 3-node systems.

To assess if the core motifs are also meaningful for larger networks we analysed randomly generated 4-node networks with respect to their TP capability. For 25 networks possessing both core motifs, we find that 84% indeed exhibit TPs. In contrast, for 25 networks not possessing both core motifs, only 24% give rise to Turing patterns. This suggests that our compositional rules are still meaningful for larger networks, even though they appear to be less restrictive than for 2- and 3-node networks.

Differentiating Turing instabilities

The dispersion relation of a system is typically related to the resulting pattern: the wavenumber $q_{\max}$ for which the dispersion relation assumes a global maximum (see STAR Methods) experiences the largest amplification. We thus expect to see a pattern with wave-

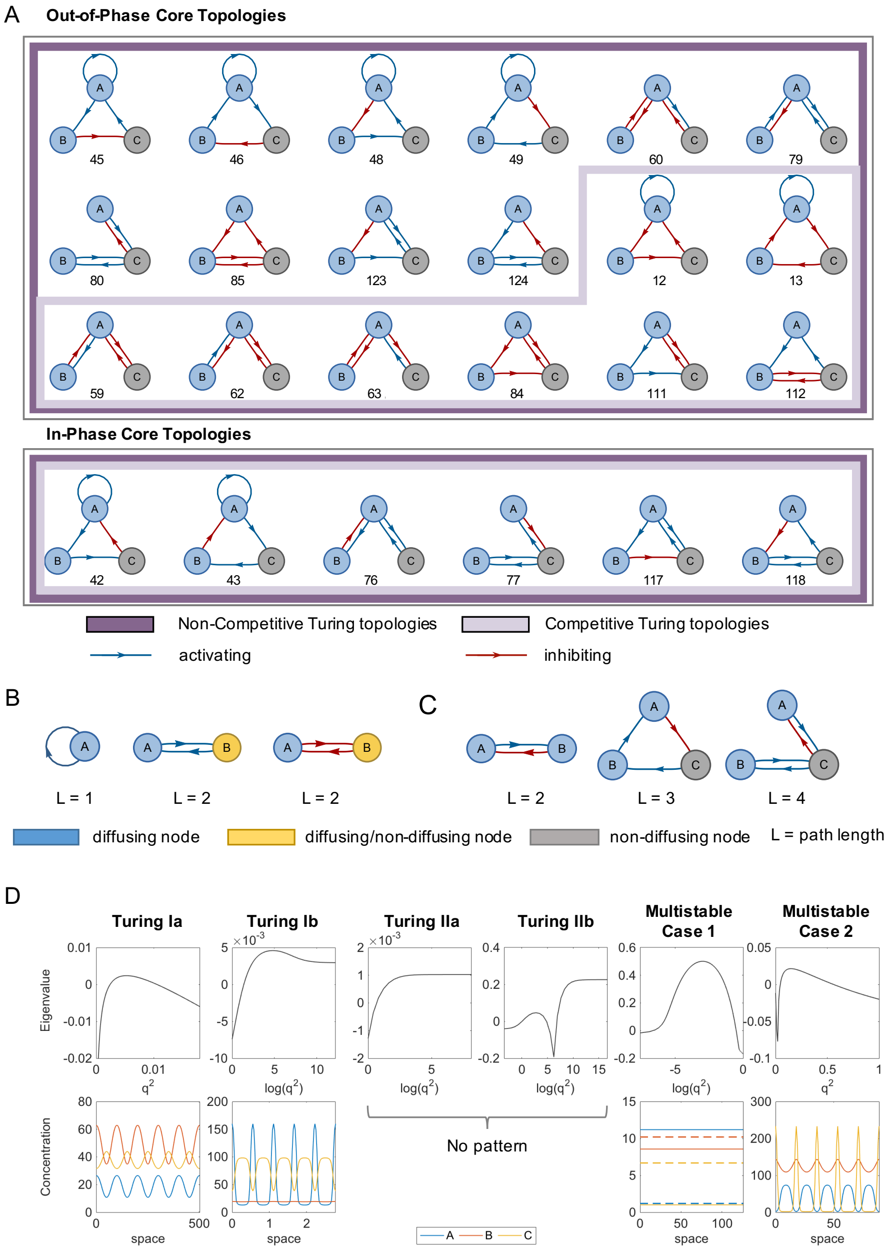


Figure 4: Core networks and core motifs for 3-node networks. (A) Top: 3-node core topologies that generate out-of-phase patterns (one species' maximum concentration is shifted by half a period with respect to the other two). Bottom: 3-node core networks exhibiting in-phase patterns (all concentration maxima aligned). Colored frames indicate the regulatory mechanism (competitiveness) for which the networks exhibit Turing I instabilities. (B) Realisations of first core motif: positive feedback loop of length one or two on one of the diffusing nodes. The figure shows the different possibilities to achieve this. If the path length is two it can be mediated by either the other diffusing or the stationary node. (C) Examples for the second identified core motif: diffusion-mediated negative feedback on a diffusing node. One diffusing node has to have a negative feedback loop whose path includes the other diffusing node. This interaction can consist of two, three or four edges and the figure shows one example each. The core motifs in (B) and (C) are almost sufficient and almost necessary for TPs. (D) Examples of different Turing instabilities and resulting patterns. Note that for Turing IIa and Turing IIb instabilities no patterns are formed. Note also that in the multistable case 1, no pattern is generated despite the existence of a Turing I instability. Instead, starting from a perturbation around the stable steady state with Turing I instability (indicated as dashed lines) this moves to a second homogeneous steady state. This behavior is observed for 4% (non-competitive) and 14 % (competitive) of multistable network-parameter combinations exhibiting a Turing I instability. The parameters for which the dispersion relations and patterns were generated are given in SI document 3.

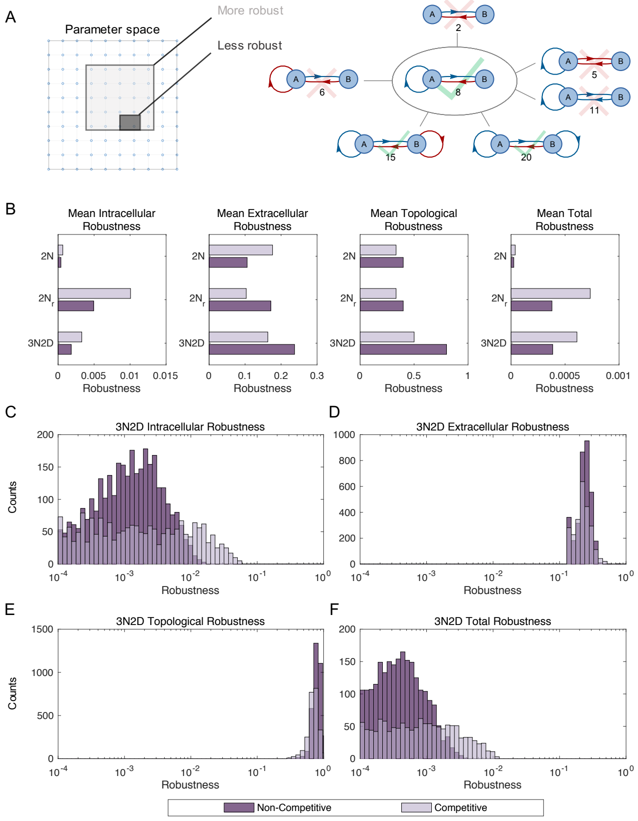


Figure 5: Quantifying a network's Turing capability: defining four measures of robustness. (A) Left: definition of intracellular (or extracellular) robustness as the fraction of sampled kinetic (or diffusion) parameter sets leading to Turing I instabilities. Right: definition of topological robustness as the fraction of neighboring networks (that is, networks into which a given network can be transformed by addition, deletion or modification of a single edge) that exhibit Turing I instabilities. The example shows that two out of six neighbors of network #8 are Turing networks, leading to a robustness of $2/6$. The total robustness is defined as the product of intracellular, extracellular and topological robustness. (B) Mean values for the different robustness measures for non-competitive (dark lilac) and competitive systems (light lilac). $2N$ ($2N_r$) denotes 2-node systems for which V and b are varied (restricted to 100 and 0.1, respectively). 3N2D represents the 3-node systems with two diffusing nodes. (C-F) Histogram plots of the four robustness measures comparing both regulatory mechanisms for 3N2D networks. We find that competitive systems are on average more robust for intracellular processes, whereas non-competitive systems are topologically more robust. In terms of total robustness, a subset of 3N2D competitive networks constitute the best performing networks.

length $2\pi/q_{\text{max}}$ (see Figure 1C).

Because of the unprecedentedly large number of networks considered here, we also find dispersion relations that have not previously been discussed. We can distinguish four different groups of dispersion relations. First, the "classical" Turing I instability fulfills three criteria: for $q = 0$, the system should be stable (i.e. $(\lambda) < 0$); for a finite q we have $(\lambda) > 0$; and $(\lambda) < 0$ for $q \to \infty$. This type of instability is the most commonly discussed mechanism underlying TPs.

For a Turing II instability the steady state is stable for $q = 0$, and the dispersion relation becomes positive for some q but becomes maximal for $q \to \infty$. Consequently, short wavelengths get amplified the most, which does not lead to a stable pattern with a well-defined wavelength.

The previous cases are referred to as Type Ia and IIa instabilities. However, other types of dispersion relations exist for diffusion driven instabilities. Figure 4D shows examples for all four categories that we find. Type Ib is similar to Type Ia and is capable of producing stable patterns. But the dispersion relation of Type Ib does not fulfill the third criterion, $(\lambda) < 0$ for $q \to \infty$, and remains positive instead.

Analogously Type IIb is similar to Type IIa and does not produce patterns, but in contrast to the original Type IIa case, the dispersion relation becomes negative again for some intermediate q, see Figure 4D.

We note that in the analyses described in the previous sections we have distinguished systems according to their patterning capability and referred to Ia and Ib jointly as "Type I", and IIa and IIb jointly as "Type II". Both types of instabilities are important when considering TP systems. Ib could be mistaken for a non-patterning system as it does not fulfill a classical criterion for a Turing I instability, which may lead to underestimating the robustness of a system. Similarly, IIb can be mistaken for a Type I instability (if q is sampled over an insufficient scale) consequently leading to an overestimation of Turing robustness.

Turing I instabilities are not a sufficient criteria for patterning

In development, cell-fate decisions are often governed by systems in which multiple stable steady states exist (Harrington, Azogui, Yalahom-Ronen, Plotnikov, & Stumpf, 2014; Harrington et al., 2013; Moris, Pina, & Martinez Arias, 2016). We therefore explore how multi-stability affects TP formation. Systems with multiple stable steady states can exhibit Type I instabilities (Ia or Ib), either for a single or for several steady states. The former has also been reported in a recent preprint (Smith & Dalchau, 2018). We investigate the PDEs for all multistable systems for which either one or multiple Turing I instabilities are present. Even though most systems indeed show the expected pattern formation (86% and 96 %, for competitive and non-competitive systems, respectively), some do not (14% for competitive and 4% for non-competitive). Rather than observing patterning, we find that these systems transition from a perturbation around the steady state with a Turing I instability to one of the other stable steady states, see Figure 4D Panel 5. We thus conclude that a Type I Turing instability is not a sufficient criterion for pattern formation in multi-stable systems.

Defining quantifiable measures of robustness

It is crucial not only to identify networks that exhibit Turing instabilities, but also to assess their sensitivity with respect to the uncertainties that pervade systems biology modelling (Kirk, Babtie, & Stumpf, 2015b). To this end, we define four measures of robustness: robustness to intracellular processes, robustness to extracellular processes, topological robustness and total robustness (Figure 5A). As intracellular parameters, we define all kinetic parameters of the ODE equations (see Figure 2B) that describe the chemical interactions between species within cells, and we define "intracellular robustness" as the fraction of the analyzed parameter combinations that is capable of Turing pattern formation.

In addition to intracellular processes, the speed of extracellular diffusion of molecules determines if a network possesses a Turing instability. We accordingly define the "extracellular robustness" as the robustness of a Turing network to changes in the diffusion constants, given that the intracellular parameters are fixed to values that can give rise to a Turing instability.

Moreover we are typically not even certain about the network topology of biological systems (Babtie et al., 2014). Accordingly we define "topological robustness" as follows: for a given network, consider all networks that can be generated by adding, removing or changing one edge (see Figure 5A). Then the topological robustness is defined as the fraction of generated networks that are capable of exhibiting Turing I instabilities.

Finally, the "total robustness" is defined as the product of the intracellular, extracellular and topological robustness (Figure 5).

Competitive 3-node systems are the most robust Turing networks

Turing networks with competitive interactions are on average more robust than non-competitive ones, in particular with respect to intracellular parameters. This is consistent for 2-node systems (~ 1.7-fold), and 3-node systems with two diffusing species (~ 1.8-fold) (Figure 5B). As the intracellular robustness varies over about two orders of magnitude more than the extracellular and topological robustness (Figure 5 B-D), competitive systems are also in total more robust ($\sim 1.5 - 1.9$-fold). Choice of regulatory interactions can have significant influence on a network's Turing capability: non-competitive topologies are more likely to be able to generate TPs whereas competitive systems that do generate TPs are on average more robust. We note that this is only the case on average; there exist some individual topologies that are more robust for non-competitive regulations.

Due to computational cost, the parameters V and b were restricted for the analysis of 3-node systems (see STAR Methods). Consequently, to compare like-for-like, we calculated the robustness for 2-node systems under the condition that V and b were fixed to the same values (denoted 2N$_r$ in Figure 5B). With this restriction on parameter space, 2-node systems are on average more robust to intracellular variations than 3-node systems ~ 3-fold for competitive and ~ 2.7-fold for non-competitive regulations). On the other hand, 3-node systems are on average more robust to extracellular ($\sim$ 1.5 (competitive) and $\sim$ 1.4-fold (non-competitive)) and topological ($\sim$ 1.5-fold (competitive) and $\sim$ 2-fold (non-competitive)) variations than 2-node systems. Even though in total the average robustness of 2 compared to 3-nodes does not differ appreciably, among the top (most robust) networks, 3-node systems are more than 4-fold more robust in total than the top 2-node systems (data not shown). It is thus likely that TP networks in nature will be composed of at least three interacting species.

Robustness maps of 3N2D topology space to reveal the most robust networks and their neighborhoods

Due to the large number of 3-node Turing topologies we visualize their total robustness in a "robustness map" shown in Figure 6A,E. We group networks according to their complexity, and depict the most robust network for each class (right panel). Overall, networks with complexity of 5-6 (competitive) and 6-7 (non-competitive) are the most robust topology groups. However, we find no clear relationship between robustness and topology complexity. This further suggests that increasing complexity does not generally lead to a larger robustness. All robustness measures (intracellular, extracellular, topological and total robustness) are provided in SI document 7; network identifiers correspond to those in the network graphs provided in SI document 6.

Due to the large number of topologies, we are not able to illustrate the full network atlas to show how these networks are related. Instead, we provide a local neighborhood atlas for the single most robust network (Figure 6C,G) and a corresponding two-dimensional PDE solution (Figure 6D,H). The local neighborhood atlas contains all the networks that can be generated from the central network by adding, deleting or modifying one edge. Most edge changes lead to a pronounced drop in Turing robustness, although it is still possible for evolution to "walk" from one topology to another while still maintaining a TP.

Identifying and engineering Turing systems in biological systems

Identifying Turing networks in real systems is a challenging task since it is typically difficult to identify all interactions in a given system, exclude influences of unobserved species, and/or to accurately estimate the system parameters. Consequently, only few Turing systems have been identified in natural systems to date and we believe that the results presented here will support future findings. Given partial knowledge of a hypothetical Turing system, the network atlas presented her may help to dismiss the hypothesis, or indicate what additional interactions are needed, thereby guiding further experiments. The identified core motifs should be particularly valuable in this respect as they provide general guidelines for the composition of potential TP networks.

Similarly, we believe that our results will prove to be useful for the challenging task of synthetically engineering TP networks. Given a synthetic system with certain components considered as suitable to produce TPs, our network atlas allows to identify missing interactions and necessary parameter values, allowing to tune the system torwards TP behaviour, or vice versa, allow to dismiss a the potential model all together. Crucially, we do not only provide a comprehensive atlas of networks capable of producing TPs, but a complete list of robustness values w.r.t. different measures, further supporting the identification of potential systems.

One important result of this study are the significant differences between competitive and non-competitive interactions w.r.t. TP generation. When searching for TPs in biological systems, it is hence not sufficient to merely identify the topology in terms of species and activating and inhibiting interactions, but the specific types of regulatory mechanisms need to be considered. While these are known for some biological networks, they are often difficult to identify. Our results may help to refine further experiments or dismiss potential models.

We found that fewer competitive TP systems exist than non-competitive ones, while the former are on average more robust than their non-competitive counterparts. This raises an interesting question from an evolutionary perspective: is it more likely for nature to find systems by varying the whole topology or by varying exploring parameter values for given topologies? More experiments on natural TP systems are needed to answer this question.

Interestingly, w.r.t. robustness in diffusion parameters the opposite is true: non-competitive systems are more robust than their competitive counterparts. In particular in the context of engineering of TPs where differential diffusion is often difficult to achieve, this may have implications on the choice of potential building blocks.

The most suitable networks for engineering purposes are those that are the most robust, both to intracellular parameters, as well as supporting a wide range of extracellular differential diffusion combinations. In this regard, the top networks in Figure 6 are the most interesting for synthetic biology engineering projects. The most robust network is $\#1275$ (Figure 6F). However, this requires competitive interactions, which may be challenging from a synthetic biology perspective. For example, activators and inhibitor connections inside cells (e.g. transcription factors) would need to interact with the same operator site to enable competition. Although this is not impossible, the non-competitive network $\#3954$ (Figure 6B) might therefore be easier to engineer. Notably, $\#3954$ supports out-of-phase expression of nodes A and B, which thus would give better resolved colour reporter outputs (e.g. GFP and RFP).

Discussion

Turing's pattern generating mechanism, later independently rediscovered by Gierer and Meinhardt, is an elegant way in which purely biochemical mechanisms can give rise to reproducible and self-organizing spatial patterns. Despite initial unease (and sometimes outright hostility) over them being relevant and robust mechanisms of patterning, TPs are now widely accepted and have become an important cornerstone of modern developmental biology.

Given their provenance, it is perhaps not surprising that TPs have also received close attention by mathematical modelers interested in biological pattern formation. But these have typically focused on single models, exploring them in great detail. The large-scale *in silico* surveying of potential TP models is a much more recent phenomenon. There are two potential pitfalls in such analyses: (i) computational cost may require simplified models or prohibit exhaustive analysis; (ii) automating any mathematical analysis, but in particular something as subtle as pattern formation is non-trivial unless we have very precise criteria by which stability, robustness and patterns can be scored. Our approach addressed these issues explicitly and from the outset, and the algorithm employed here is capable of analyzing a wide range of different network structures and is independent of functional choices for regulatory mechanisms and rate functions. The code is freely available and allows the automated analysis of arbitrary networks in a user-friendly manner.

In combination with extensive computational resources, we conducted a thorough and comprehensive analysis of 2-node and 3-node candidate Turing network models (amounting to 8 CPU years computing time). Our results demonstrate that the structure of the network alone cannot determine whether a TP ex-

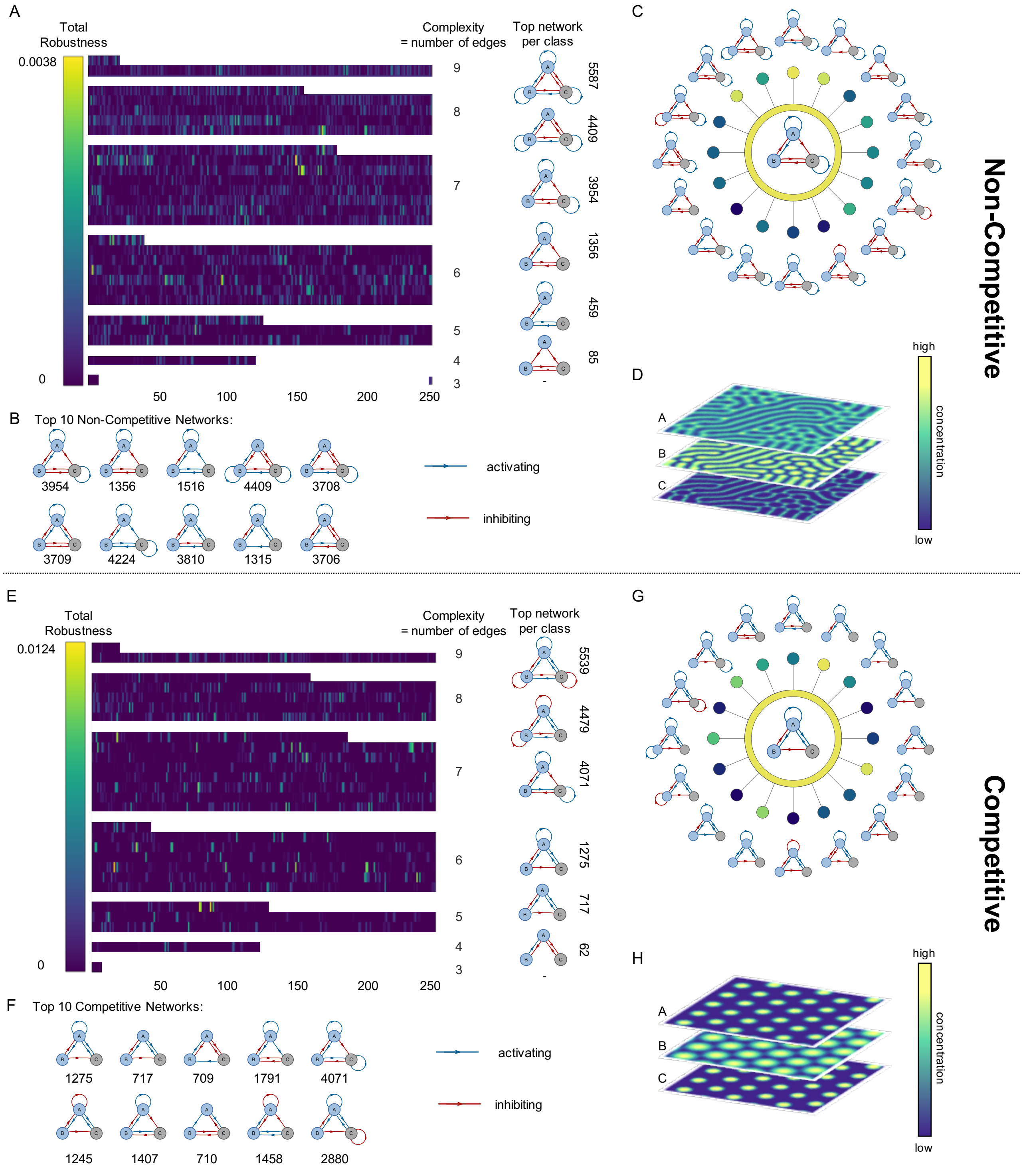


Figure 6: Most robust 3-node Turing networks with non-competitive (A-D) and competitive (E-H) regulation and two diffusing nodes. (A,E) Robustness map: The figure indicates the total robustness of all analyzed 3-node networks arranged according to complexity. For each complexity class we show the most robust network, together with its ID number. (B,F) Ten most robust networks with numbers indicating the corresponding network ID. Total robustness decreases towards the right and bottom, i.e. for competitive regulation (F) network #1275 is the most robust network. The robustness values are given in SI document 7. (C,G) Local neighborhood atlas for the most robust competitive Turing network #1275 (center). Colored nodes indicate the total robustness value according to the colorscale shown in (A). (D,H) 2-dimensional pattern for the the most robust non-competitive ((D) #3954) and competitive Turing topology ((H)#1275). Yellow and blue indicate high and low species concentrations, respectively, for nodes A, B and C. The parameters for which the pattern was generated are given in SI document 4.

ists; this is in line with a large body of work on network dynamics (Ingram, Stumpf, & Stark, 2006), but seems to directly contradict results from other studies on TP mechanisms (Zheng et al., 2016; Marcon et al., 2016). These, however, (i) fail to fully assess the dependence of TP on regulatory mechanisms and reaction rate parameters (largely because a mathematically more convenient model structure was imposed); and (ii) their models are special cases of the more general and more comprehensive treatment here.

Perhaps the most surprising result of this analysis is how common networks capable of producing TPs are. Due to the high sensitivity of TP networks to changing parameters, it may seem unlikely that these can ever exist in nature. How they are stabilised in real systems remains an important open question in the field. However, the large number of TP networks identified here (more than 61% of networks considered here can produce TPs) may go some way towards explaining how natural systems evolved to produce them, since it gave evolution many opportunities to 'stumble' across a potential Turing network.

Being common but not very robust to parametric and structural changes could suggest that many different architectures are used in nature to generate TPs (as was already hinted at in some of Meinhardt's work (Meinhardt, 2013)). Once a structure is in place with suitable regulatory interactions and reaction rate parameters, and the resulting TP confers an evolutionary advantage, natural selection is likely to maintain this mechanism. However, it remains an open question how systems in nature would be able to fine-tune kinetic parameters to form stable TP networks.

While network structure by itself neither guarantees nor implies the existence of a TP, the overwhelming majority of TP generating mechanisms embed the hallmarks encapsulated by two core motifs: a positive feedback on at least one of the diffusing nodes and a diffusion-mediated negative feedback loop on both diffusing nodes. It is tempting to speculate that larger systems will also reflect these compositional rules and have these core motifs embedded; a basic TP generating motif could thus, for example, be regulated in a more nuanced manner.

Finally, our exhaustive analysis reveals a spectrum of Turing-like instabilities, and subtle dependencies between these and the eventual pattern formation. These are naturally easily missed when analyzing individual Turing systems, or applying simplifying measures in large-scale surveys.

The analysis presented here provides us with a set of blueprints for TP generating mechanisms that can guide the search for naturally evolved Turing systems, as well as the *de novo* engineering of biosynthetic systems. For the latter, in particular, competitive networks should be a safer bet, due to their increased (intracellular) parametric robustness. This tension between commonality and robustness is perhaps one of the most fascinating (or vexing) features of Turing mechanisms.

Acknowledgements

This work was funded by the Volkswagen Foundation. NS was partially supported by a BBSRC DTP PhD studentship and by a Boehringer-Ingelheim-Fonds PhD Fellowship. We like to thank the members of the *Theoretical Systems Biology Group* in London and Melbourne for helpful discussions. The manuscript has benefited from the comments of three anonymous referees. And we are grateful to Neil Dalchau for pointing out additional symmetry considerations in counting candidate Turing pattern generating motifs.

Author Contributions

N.S.S. and D.S. performed the research. All authors designed the research and wrote the paper. All authors read and approved the final manuscript.

Declaration of interest

The authors declare no competing interests.

References

Babtie, A. C., Kirk, P, & Stumpf, M. P. H. (2014). Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18507–18512.

Barnes, C. P., Silk, D., Sheng, X, Stumpf, M. P. H., & Stumpf, M. P. H. (2011). Bayesian design of synthetic biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15190–15195.

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature*, *434*(7037), 1130.

Becskei, A., Séraphin, B., & Serrano, L. (2001). Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *The EMBO journal*, *20*(10), 2528–2535.

Boehm, C. R., Grant, P. K., & Haseloff, J. (2018). Programmed hierarchical patterning of bacterial populations. *Nature communications*, *9*(1), 776.

Borek, B., Hasty, J., & Tsimring, L. (2016). Turing patterning using gene circuits with gas-induced degradation of quorum sensing molecules. *PloS one*, *11*(5), e0153679.

Burrill, D. R. & Silver, P. A. (2010). Making cellular memories. *Cell*, *140*(1), 13–18.

Cachat, E., Liu, W., Martin, K. C., Yuan, X., Yin, H., Hohenstein, P., & Davies, J. A. (2016). 2-and 3-dimensional synthetic large-scale de novo patterning by mammalian cells through phase separation. *Scientific reports*, *6*, 20664.

Carvalho, A., Menendez, D. B., Senthivel, V. R., Zimmermann, T., Diambra, L., & Isalan, M. (2014). Genetically encoded sender-receiver system in 3d mammalian cell culture. *ACS Synthetic Biology*, *3*(5), 264–272.

Cotterell, J. & Sharpe, J. (2010). An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Molecular systems biology*, *6*(1), 425.

Diambra, L., Senthivel, V. R., Menendez, D. B., & Isalan, M. (2015). Cooperativity to increase turing pattern space for synthetic biology. *ACS synthetic biology*, *4*(2), 177–186.

Diego, X., Marcon, L., Müller, P., & Sharpe, J. (2017). Key features of turing systems are determined purely by network topology. *arXiv preprint arXiv:1708.09645*.

Duran-Nebreda, S. & Solé, R. V. (2016). Toward synthetic spatial patterns in engineered cell populations with chemotaxis. *ACS synthetic biology*, *5*(7), 654–661.

Economou, A. D., Ohazama, A., Porntaveetus, T., Sharpe, P. T., Kondo, S., Basson, M. A., ... Green, J. B. (2012). Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate. *Nature genetics*, *44*(3), 348–351.

Estrada, J., Wong, F., DePace, A., & Gunawardena, J. (2016). Information integration and energy expenditure in gene regulation. *Cell*, *166*(1), 234–244.

Ferrell, J. E. & Machleder, E. M. (1998). The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science*, *280*(5365), 895–898.

Ferrell, J. E., Tsai, T. Y.-C., & Yang, Q. (2011). Modeling the cell cycle: Why do certain circuits oscillate? *Cell*, *144*(6), 874–885.

Gaffney, E. A., Yi, F., & Lee, S. S. (2016). The bifurcation analysis of turing pattern formation induced by delay and diffusion in the Schnakenberg system. *Discrete and Continuous Dynamical System-Series B*, *22*(2), 647–668.

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature*, *403*(6767), 339.

Gierer, A. & Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik*, *12*(1), 30–39.

Green, J. B. A. & Sharpe, J. (2015). Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development*, *142*(7), 1203–1211.

Harrington, H. A., Feliu, E., Wiuf, C, Wiuf, C., Stumpf, M. P. H., & Stumpf, M. P. H. (2013). Cellular compartments cause multistability and allow cells to process more information. *Biophysical journal*, *104*(8), 1824–1831.

Harrington, H. A., Azogui, H. H., Yahalom-Ronen, Y., Plotnikov, A., & Stumpf, M. P. H. (2014). Nuclear to cytoplasmic shuttling of ERK promotes differentiation of muscle stem/progenitor cells. *Development*, *141*(13), 2611–2620.

Howard, J., Grill, S. W., & Bois, J. S. (2011). Turing's next steps: The mechanochemical basis of morphogenesis. *Nature Reviews Molecular Cell Biology*, *12*(6), 392–398.

Ingolia, N. T. & Murray, A. W. (2004). The ups and downs of modeling the cell cycle. *Current Biology*, *14*(18), R771–R777.

Ingram, P. J., Stumpf, M. P., & Stark, J. (2006). Network motifs: Structure does not determine function. *BMC genomics*, *7*(1), 108.

Iron, D., Wei, J., & Winter, M. (2004). Stability analysis of Turing patterns generated by the Schnakenberg model. *Journal of mathematical biology*, *49*(4), 358–390.

Jung, H.-S., Francis-West, P. H., Widelitz, R. B., Jiang, T.-X., Ting-Berreth, S., Tickle, C., ... Chuong, C.-M. (1998). Local inhibitory action of BMPs and their relationships with activators in feather formation: Implications for periodic patterning. *Developmental biology*, *196*(1), 11–23.

Kirk, P. D. W., Rolando, D. M. Y., Maclean, A. L., & Stumpf, M. P. H. (2015a). Conditional random matrix ensembles and the stability of dynamical systems. *New Journal of Physics*, *17*(8), 083025.

Kirk, P. D. W., Babtie, A. C., & Stumpf, M. P. H. (2015b). Systems biology (un)certainties. *Science*, *350*(6259), 386–388.

Klumpp, S., Zhang, Z., & Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell*, *139*(7), 1366–1375.

Kondo, S. & Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *science*, *329*(5999), 1616–1620.

Liu, P., Shi, J., Wang, Y., & Feng, X. (2013). Bifurcation analysis of reaction-diffusion Schnakenberg model. *Journal of Mathematical Chemistry*, *51*(8), 2001–2019.

Ma, W., Trusina, A., El-Samad, H., Lim, W. A., & Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *138*(4), 760–773.

Maclean, A. L., Kirk, P. D. W., & Stumpf, M. P. H. (2015). Cellular population dynamics control the robustness of the stem cell niche. *Biology Open*, bio.013714.

Maini, P. K., Woolley, T. E., Baker, R. E., Gaffney, E. A., & Lee, S. S. (2012). Turing's model for biological pattern formation and the robustness problem. *Interface Focus*, *2*(4), 487–496.

Marcon, L., Diego, X., Sharpe, J., & Müller, P. (2016). High-throughput mathematical analysis identifies turing networks for patterning with equally diffusing signals. *Elife*, *5*.

Meinhardt, H. (2013). *The Algorithmic Beauty of Sea Shells*. Springer.

Meinhardt, H. & Gierer, A. (2000). Pattern formation by local self-activation and lateral inhibition. *Bioessays*, *22*(8), 753–760.

Moris, N., Pina, C., & Martinez Arias, A. (2016). Transition states and cell fate decisions in epigenetic landscapes. *Nature reviews. Genetics*, *17*(11), 693–703.

Nakamasu, A., Takahashi, G., Kanbe, A., & Kondo, S. (2009). Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. *Proceedings of the National Academy of Sciences*, *106*(21), 8429–8434.

Palmer, A. R. (2004). Symmetry breaking and the evolution of development. *Science*, *306*(5697), 828–833.

Raspopovic, J., Marcon, L., Russo, L., & Sharpe, J. (2014). Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science*, *345*(6196), 566–570.

Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *Journal of molecular biology*, *323*(5), 785–793.

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

Scholes, N. S. & Isalan, M. (2017). A three-step framework for programming pattern formation. *Current Opinion in Chemical Biology*, *40*, 1–7.

Sick, S., Reinker, S., Timmer, J., & Schlake, T. (2006). WNT and DKK Determine Hair Follicle Spacing Through a Reaction-Diffusion Mechanism. *Science*, *314*(5804), 1447–1450.

Smith, S. & Dalchau, N. (2018). Beyond activator-inhibitor networks: The generalised turing mechanism. *arXiv preprint arXiv:1803.07886*.

Tan, Z., Chen, S., Peng, X., Zhang, L., & Gao, C. (2018). Polyamide membranes with nanoscale turing structures for water purification. *Science*, *360*(6388), 518–521.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *237*(641), 37–72.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, *25*(1), 1–47.

Zheng, M. M., Shao, B., & Ouyang, Q. (2016). Identifying network topologies that can generate turing pattern. *Journal of Theoretical Biology*, *408*, 88 –96.

STAR Methods

Contact for reagent and resource sharing

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Michael P.H. Stumpf (m.stumpf@imperial.ac.uk).

Method details

A semi-formal summary of our TP search

To identify potential Turing instabilities, we first consider the stability of the non-spatial system. Suppose we have a system of N interacting molecular species with time-dependent concentrations $x_i(t), i = 1, \ldots, N$. We model the dynamical behavior of such concentrations in terms of a system of ordinary differential equations (ODEs),

$$\frac{d}{dt} x_i(t) = f_i(\mathbf{x}(t)), \quad i = 1, \ldots, N, \tag{1}$$

where $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$, and $x_i(t) \in \mathbb{R}$ is the concentration of the ith species at time t. Equations of the form in (1) cannot typically be solved exactly for nonlinear functions f_i. However, efficient numerical algorithms exist to solve such equations approximately, leading to time trajectories of the system, that is, to solutions $x_i(t)$ (see Figure 2C for examples). $\mathbf{f}(\mathbf{x}(t)) = (f_1(\mathbf{x}(t)), \ldots, f_N(\mathbf{x}(t)))$ in equation (1) encodes the interactions between the different species. If the x_i denote protein concentrations, for example, $\mathbf{f}(\mathbf{x}(t))$ may encode the regulatory mechanisms between the proteins.

Figure 1A shows a network representation of the Gierer-Meinhardt model, and its governing ODEs (Gierer & Meinhardt, 1972). The nodes indicate the interacting species and arrows the direction of interaction. We distinguish between two possible types of interactions: activating (blue arrows) and inhibiting (red arrows), which we encode in the corresponding components of $\mathbf{f}(\mathbf{x}(t))$, whose functional form is not specified by the graph (we employ Hill functions in the reaction equations, see section **System of ODEs** below for details).

We next include the spatial diffusion of molecules and extend the model in Equation (1) to a spatial setting. In this case the molecule concentrations $x_i(t)$ become space-dependent concentration fields $x_i(\mathbf{r}, t)$, where $\mathbf{r} = (r_1, \ldots, r_d)$ and r_i denotes the spatial location in the ith spatial direction and d is the spatial dimension of the system. These fields satisfy the set of coupled partial differential equations (PDEs)

$$\frac{\partial}{\partial t} x_i(\mathbf{r}, t) = D_i \nabla^2 x_i(\mathbf{r}, t) + f_i(\mathbf{x}(\mathbf{r}, t)), \quad i = 1, \ldots, n, \tag{2}$$

which is obtained from Equation (1) by adding the diffusion term $D_i \nabla^2 x_i(\mathbf{r}, t)$, where D_i is the diffusion constant of the ith species and $\nabla^2 = \partial^2/\partial r_1^2 + \ldots + \partial^2/\partial r_d^2$ the Laplace operator with respect to position $\mathbf{r}$, and concentrations are now functions of space and time, $x_i(\mathbf{r}, t)$.

The next step is to screen for the formation of patterns through diffusion-driven instabilities (Maini et al., 2012). In this case, for any small spatial fluctuations one may expect the concentrations $x_i(\mathbf{r}, t)$ to become spatially constant again for large times. For many systems, this is indeed the case, but for some systems the interplay of diffusion and reactions can lead to the molecules' concentrations forming spatial patterns with certain wavelengths that are stable and reproducible in time.

In the Turing pattern framework we start with $\mathbf{x}^*$, which is the stable steady state concentration of the non-spatial system in (1), i.e. $f_i(\mathbf{x}^*) = 0, i = 1, \ldots, n$: if the system is in state $\mathbf{x}^*$, it remains there for all times, and if the system is close to $\mathbf{x}^*$ it will converge towards $\mathbf{x}^*$. If spatial diffusion of molecules is included into the model, as described in Equation (2), it is possible that deviations from the steady state of certain length scales do not decay towards the homogeneous steady state, but are instead amplified. This is then the diffusion-driven or Turing instability. If a finite wave length exists that experiences the strongest amplification, we speak of a Turing I instability. In this case, a system typically forms a pattern of the wavelength for which the amplification is maximal (see Figure 1C).

Mathematically, the stability is determined by the dependence of the real part of the largest eigenvalue of the Jacobian matrix of the system on the wavenumber q (see below for more details). We also call this dependence the "dispersion relation" (see Figure 1C, 2C and 4D for examples). For $q = 0$ (corresponding to a spatially homogeneous perturbation) the real part of the largest eigenvalue is negative, if evaluated at a stable steady state. If, however, the dispersion relation becomes positive for some wavenumber q, these wavenumbers become amplified and the steady state becomes unstable. For cases where the dispersion relation has a maximum for a finite value of q, a Turing I instability is present and the pattern is formed for this maximum wavenumber q_{max}. If the dispersion relation remains positive and becomes maximal for large wavenumbers, we speak of a Turing II instability (see bottom right panel in Figure 2C and Figure 4C Type IIa and IIb). In this case deviations on arbitrarily small length scales become amplified, and no stable pattern is formed. We thus only search for Turing I instabilities in the following.

To find Turing instabilities and determine the instability type, we first need to identify the stable steady states of a given system, and subsequently study their dispersion relation. Figure 2C summarizes the computational procedure.

For each candidate network we perform this procedure for a wide range of kinetic and diffusion parameters. We therefore have to specify the intracellular parameters determining the regulatory functions f_i, as well as the diffusion constants. The former consist of the parameters k (dissociation rates), V (scaling factors), b (basal production rates) and the degradation rate μ (see Equations in Figure 2B), which we vary across biologically relevant values between 0.1 and 100 (0.01-1 for μ). We vary them on a regular grid in logarithmic scale with 3 values per parameter amounting to 3^P intracellular parameter combinations per network, where P is the number of parameters. The extracellular/diffusion parameters are varied over a range between 10^{-3} and 10^3. Due to computational cost, we vary the parameters V and b only for the 2-node networks, but fix them to 100 and 0.1, respectively, for 3-node networks. In total, we screened more than 3×10^{11} network-parameter combinations for TP formation, which we believe is the largest study of its kind to date.

Definition of networks and ODEs

Networks

To generate all possible N-node networks we first compute all possible $(N \times N)-$matrices with elements 0, 1 and -1, where a 0 $(1/-1)$ represents the absence (presence) of an activating or inhibiting interaction/edge. The number of matrices is reduced by considering only connected networks and accounting for symmetries. We also remove matrices that correspond to networks including nodes without any incoming or outgoing edges. Each remaining matrix M serves as an adjacency matrix for a network, where the element M_{ij} being 1 (-1) represents a positive (negative) edge from node i to node j.

System of ODEs

For each adjacency matrix we then construct the corresponding set of ODEs. Each non-zero entry in the adjacency matrix corresponds to a Hill-type term which are combined in either a non-competitive or competitive case. If S_i^+ (S_i^-) denotes the set of positive (negative) edges ending in node i, then in the non-competitive case, the different regulators act (and saturate) independently of each other, and we have,

$$\begin{aligned} f_i(x_1, ..., x_N) = & V_i \cdot \prod_{j \in S_i^+} \left(\frac{1}{1 + \left(\frac{k_{ij}}{x_j}\right)^{n_{ij}}} \right) \\ & \times \prod_{j \in S_i^-} \left(\frac{1}{1 + \left(\frac{x_j}{k_{ij}}\right)^{n_{ij}}} \right) + b_i - \mu_i x_i. \end{aligned} \tag{3}$$

Here, V_i is the maximal induced production rate, b_i the basal production rate, μ_i the degradation rate, k_{ij} the concentration value at which the regulation of the ith species by the jth species is half its maximal value, and n_{ij} is the corresponding Hill coefficient determining the steepness of the response.

In the competitive case the different regulators compete for the binding site which leads to an additive combination of terms:

$$\begin{aligned} f_i(x_1, ..., x_N) = & \, b_i - \mu_i x_i \\ & + V_i \cdot \frac{\sum_{j \,\in\, S^+} \left(\frac{x_j}{k_{ij}}\right)^{n_j}}{1 + \sum_{j \,\in\, S^+} \left(\frac{x_j}{k_{ij}}\right)^{n_j} + \sum_{j \,\in\, S^-} \left(\frac{x_j}{k_{ij}}\right)^{n_j}}. \end{aligned} \tag{4}$$

Numerical analysis

Steady state estimation

We generated a customized Matlab (R2016a) script to find steady states numerically. For a given system the parameters are chosen from a logarithmic grid and for each set of parameters the system of ODEs is solved numerically until time $t = 1000$ using the Matlab ODE solver ode15s. Whenever the algorithm encounters numerical problems, the (slower but more robust) algorithm ode23s is invoked. Using such a trajectory of the system, we define the system's dynamical behaviour. We excluded oscillations/limit cycles from the subsequent analysis as these can not give rise to Turing patterns in our definition. In contrast, damped oscillations can do so. We found that in this case the ODE simulations can result in small but significant errors that are propagated through the subsequent analysis leading to inaccurate results. To avoid this we use the Matlab function `fsolve` following the ODE simulations to obtain a more accurate estimate of the steady state for such cases.

To account for the possibility of multiple stable steady states, we solved the set of ODEs on a grid of initial conditions of three values per species. The endpoints of the resulting trajectories are clustered using k-means clustering. Using the cluster centroids as final initial condition, the system of ODEs is solved again to verify that the system has converged sufficiently. For steady states that are approached via damped oscillations, we again do not solve the ODEs numerically but instead search for a fixed point of the ODEs directly using the Matlab function `fsolve`.

Stability Analysis

The stability of a steady state $\mathbf{x}^*$ of Equation (1) is assessed by a linear stability analysis. We add a small perturbation around $\mathbf{x}^*$

$$\mathbf{x}(t) = \mathbf{x}^* + \delta \tilde{\mathbf{x}}(t), \tag{5}$$

with a small constant $\delta \in \mathbb{R}$, and then linearise Equation (1), to obtain

$$\frac{\partial}{\partial t} \tilde{\mathbf{x}}(t) = J \tilde{\mathbf{x}}(t) + O(\delta). \tag{6}$$

The Jacobian, J, is defined as

$$J_{i,j} = \frac{\partial f_i(x_1, \ldots, x_N)}{\partial x_j}, \quad i, j = 1, \ldots, N. \tag{7}$$

Equation (6) constitutes a linear dynamical system with steady state $\tilde{\mathbf{x}} = 0$. This steady state is asymptotically stable if and only if the real parts of all eigenvalues of the matrix J are negative. This in turn means that $\mathbf{x}^*$ is a locally asymptotically stable steady state, in the sense that there exists a neighborhood around $\mathbf{x}^*$ such that any solution of the ODEs starting from this neighborhood asymptotically converges to $\mathbf{x}^*$. Accordingly, it is sufficient to compute the eigenvalues of the Jacobian defined in Equation (7) to assess the local stability of a steady state of Equation (1).

To assess if such a stable steady state can exhibit a diffusion-driven instability, we need to analyze Equation (2). Similarly to the case without diffusion, we perturb the system around $\mathbf{x}^*$ to perform a linear stability analysis of this steady state, but this time with a harmonic wave with wavenumber q:

$$\mathbf{x}(r, t) = \mathbf{x}^* + \delta \tilde{\mathbf{x}}(t) e^{i\mathbf{qr}}, \tag{8}$$

where $\delta \in \mathbb{R}$ is a small constant, $\mathbf{q} = (q_1, \ldots q_d)$, and $\mathbf{qr}$ denotes the scalar product of the vectors $\mathbf{q}$ and $\mathbf{r}$. We further define $q = |\mathbf{q}|$. Inserting this into Equation (2) and expanding to first order in δ, one obtains a linear dynamical system similar to the one in (6), but with a modified Jacobian $\tilde{J}$ given by:

$$\tilde{J} = J - q^2 D, \tag{9}$$

where $D = \mathrm{diag}(D_1, \ldots, D_n)$ is a diagonal matrix with the diffusion constants D_i on the diagonal. For differential diffusion of the molecular species, the Jacobian in Equation (9) can have eigenvalues with a positive real part for finite wavemumber q; when this occurs the stable steady state of the system in Equation (1) becomes unstable with diffusion, and we speak of an diffusion-driven Turing instability. Depending on the behavior of the largest eigenvalue of $\tilde{J}$ for large q we distinguish different types of Turing instabilities, see Figure 4D.

Note that since we consider spatially homogeneous systems with rotational symmetry, the modified Jacobian in (9) does only depend on the norm $q = |\mathbf{q}|$ of the wavevector $\mathbf{q}$. This implies that higher spatial dimensions with more components of the wavevector $\mathbf{q}$ do not add any degrees of freedom in terms of the eigenvalues of the Jacobian (they do in terms of resulting pattern variety though of course). The dispersion relations in different spatial dimensions are hence equivalent. Consequently, if a Turing instability of a certain type exists in some dimension, it also exists in other spatial dimensions.

PDE simulation

Wherever necessary we simulate the spatial behaviour of the system both in one and two dimensions by numerically solving the corresponding PDEs. We used the Mathematica funcion NDSolve for this task. The resulting patterns are for example used to analyze the patterns' phasing by analyzing the position of the maxima and minima of each species with respect to each other.

Workflow

We implement the estimation of stable steady states of the non-spatial system and identification of Turing instabilities into an automated workflow. Overall, the computational analysis consists of the steps:

i) Define a network and its governing ODE equations.

ii) Define a grid that covers the full parameter space of interest.

iii) For each set of parameters:

   a) Perform numerical simulation of ODE system for different initial conditions*

   b) Cluster endpoints using k-means clustering to find set of steady states.

   c) Confirm steady state guesses by repeated simulation from each cluster center or by using `fsolve` for damped oscillations.

iv) Perform stability analysis for all steady states and parameter combinations:

   a) Calculate Jacobian matrix (without diffusion) and test if all real parts of the eigenvalues are negative to verify stability.

   b) Expand Jacobian matrix by diffusion term and calculate eigenvalues for a set of diffusion values and wavenumbers q.

   c) Classify possible Turing instabilities as Type Ia, Ib, or IIa or IIb (see Figure 2C).

* The initial simulation is used to determine if oscillations are present (these are excluded) or if damped oscillations are found (requires additional step of `fsolve` rather than numerical simulation).

Data and software availability

We provide code for an automated steady state estimation an stability analysis together with documentation in SI document 8. The code can analyze systems with arbitrary regulatory functions and node numbers. Average running times for all systems are shown in SI document 7.