On the relationship between structure and function of gene circuits

Study of evolvability and multi-functionality

Alba Jiménez Asins

TESI DOCTORAL UPF / ANY 2016

DIRECTOR DE LA TESI
Dr. Andreea Munteanu i Dr. James Sharpe

Departament
Systems Biology Department, Center for Genomic Regulation

upf. Universitat Pompeu Fabra
Barcelona

A mis padres y a Elena

Acknowledgements

I would like to thank both my tutor Dr. Andreea Munteanu and my supervisor Dr. James Sharpe.

Abstract

In order to study the complex relationship between the structure and function of gene circuits, we focus in the following two properties of circuits : their evolvability and multi-functionality. First, to what extent does the dynamical mechanism producing a specific biological phenotype bias the ability to evolve into new phenotypes? We find that circuits that use alternative mechanisms differ in the likelihood of reaching novel phenotypes through mutation. Second, how can a minimal circuit perform two distinct patterning functions? We discover bi-functional motifs able to perform lateral inhibition or lateral induction depending on the environment they are embedded in. We explore the design properties of multi-functional motifs and discover they can use two distinct types of design –*hybrid* or *emergent*– depending on their ability to be decomposed into distinct sub-circuits, i.e. their modularity.

Resum

Per tal d'estudiar la relació complexa entre l'estructura d'un circuit genètic i la seva funció, ens centrem en estudiar dues proprietas: la seva capacitat evolutiva i la seva multi-functionalitat. Primer, fins a quin punt el mecanisme dinàmic portant a terme una funció fenotípica té un impacte en la capacitat d'evolucionar a nous fenotips? Hem descobert que circuits que utilitzen mecanismes alternatius difereixen en la probabilitat d'assolir noves funcions quan son sotmesos a mutacions. Segon, com pot un circuit mínim assolir dos patrons cel.lulars diferents? Hem descobert motius bi-funcionals capaços d'assolir inhibició lateral o bé inducció lateral depenent del teixit on es trobin. Hem estudiat el diseny d'aquest tipus de circuit i descobert que es poden dividir en dos tipus principals –*hybrids* o *emergents*– que corresponen a la seva capacitat en ser descompostos en differents sub-circuits, és a dir en la seva modularitat.

Contents

List of figures **xiv**

List of tables **xv**

I Global Introduction **1**

0.1 Gene circuits . . . 3
    0.1.1 Gene circuits drive particular biological functions . . . 3
    0.1.2 How to model a gene circuit . . . 5
0.2 Pattern formation in multicellular systems . . . 9
    0.2.1 Hierarchical mechanisms: the gradient-threshold model . 11
    0.2.2 Emergent mechanisms: Turing's reaction-diffusion model and direct cell-cell patterning . . . 13
    0.2.3 Role of noise in pattern formation . . . 16
0.3 Relationship between structure and function of gene circuits . . . 18
    0.3.1 From structure to function . . . 18
    0.3.2 From function to structure . . . 21
    0.3.3 The concept of dynamical mechanism . . . 27
    0.3.4 Dynamical systems . . . 31
0.4 Objectives of this thesis . . . 34

II Dynamical mechanisms of gene circuits shape evolvability **37**

1 INTRODUCTION **39**

1.1 Evolvability from the building of genotype-phenotype maps . . . 39
1.2 Evolvability from the topological features of gene circuits . . . 42
1.3 Mechanism-view on evolvability . . . 42

2 METHODS 45
2.1 The phenotype 45
2.2 The genotype 45
2.3 The gene regulatory model 46
2.4 Exploring the space for solutions 47
2.5 Building mechanism-specific neutral spaces 47
2.6 Evolvability: defined and measured 49
  2.6.1 Choice of mutation process 50
  2.6.2 The mutation process 50
  2.6.3 Functional versus non-functional phenotypes 51
  2.6.4 Classifying functional phenotypes into different patterns . 52

3 RESULTS 55
3.1 Evolutionary potential is mechanism-dependent 55
  3.1.1 Evolutionary potential at a fixed mutation strength 55
  3.1.2 Evolvability profiles obtained as mutation strength is varied 57
  3.1.3 Explaining the differences in evolvability by mechanism . 60
  3.1.4 Oscillatory phenotypes 61
  3.1.5 Influence of the model formalism on the results 61
3.2 Exploring phenotypic transitions 66
  3.2.1 Depicting the phenotypic neighbourhood from a given genotype: phenotype-transition diagrams 66
  3.2.2 Phenotypic transitions are continuous 71
  3.2.3 Phenotypic hubs 72

4 DISCUSSION 77

III Modular properties of multi-functional gene circuits 81

5 INTRODUCTION 83
5.1 Structural modularity 83
  5.1.1 Structural modularity arises as large circuits adapt to perform multiple functions 84
  5.1.2 Structural modularity arises as specific ‘circuit building rules’are considered 86
5.2 Revisiting modularity 87
5.3 Search for compact circuits 87
5.4 Defining a multi-functional circuit 88

6 METHODS **91**
- 6.1 Choice of the model 91
- 6.2 The gene regulatory model 92
  - 6.2.1 Spatial model 92
  - 6.2.2 Flexibility of the regulatory function 95
- 6.3 Tissue specificity 95
- 6.4 Objective functions: lateral induction and lateral inhibition 95

7 RESULTS **99**
- 7.1 Mono-functional circuits 99
- 7.2 Bi-functional circuits 105
  - 7.2.1 Candidates for bi-functionality 105
  - 7.2.2 Search for bi-functional circuits 107
- 7.3 Mechanisms for function switching 110
  - 7.3.1 Underlying mechanisms of hybrid circuits 110
  - 7.3.2 Underlying mechanisms of emergent circuits 112
- 7.4 Continuous versus discontinuous pattern transitions 117
- 7.5 Conservation of design principles in higher-order circuits 121

8 DISCUSSION **125**

IV Global Discussion **129**

List of Figures

0.1 Gene circuits integrate multiple regulatory events 4
0.2 Distinct formalisms to model a gene circuit 6
0.3 Basic developmental mechanisms for pattern formation 10
0.4 The gradient-threshold model 12
0.5 Turing's reaction diffusion model 14
0.6 An illustration of lateral inhibition in proneural clusters in the *Drosophila* neuroectoderm 16
0.7 The C1-FFL motif behaves as a persistence detector when the Z promoter behaves as an AND gate 19
0.8 A gene regulatory motif that generates oscillatory or multiway switch outputs 20
0.9 Neutral spaces of a RNA sequence-to-shape model 22
0.10 Neutral spaces of a boolean genotype-to-phenotype model 25
0.11 Morphogen interpretation 26
0.12 Complexity atlas showing the six fundamental mechanisms for morphogen interpretation 28
0.13 Alternative mechanisms to achieve a single phenotype 30
0.14 Minimal motifs and dynamical mechanisms 31
0.15 Phase portrait of a toggle switch 33

1.1 Features of neutral spaces 40

2.1 Dynamical mechanisms impose a scattered structure of neutral spaces 48
2.2 Measuring evolvability 51
2.3 The stripe and seven novel patterns 54

3.1 Differences in evolvability: schematics of novel phenotypes 56
3.2 Dynamical mechanism influences evolvability 58
3.3 Impact of the geometry of adjacent neutral spaces on evolvability . 59
3.4 Dynamical mechanisms are distinct in their capacity to achieve oscillatory phenotypes under mutation 62

3.5 Complexity atlases drawn for different model formalisms 64
3.6 Evolvability measured under distinct particularities of the model . 65
3.7 Exhaustive exploration of phenotype space from a genotype . . . 67
3.8 Heterogeneous phenotypic neighbourhoods within a single mechanism 68
3.9 Heterogeneous phenotypic neighbourhoods within a single mechanism 69
3.10 Characteristic phenotypic neighbourhoods of two mechanisms . . 70
3.11 Continuous phenotypic transitions 72
3.12 Gene expression patterns of gap gene *giant* in several *Drosophila* mutants 73
3.13 Phenotypic diversity accessible from a genotype: the existence of phenotypic hubs 74
3.14 Distribution of phenotypic diversity per mechanism 75

4.1 Mechanism-based view on evolvability 79

5.1 Structural modularity 85
5.2 Compact encoding of multi-functional circuits 88

6.1 Multi-stable boolean circuits 92
6.2 Defining a bi-functional circuit 94
6.3 Versatility of the regulatory function 96

7.1 Color-coded complexity atlas of mono-functional circuits 100
7.2 Alternative mechanisms to achieve induction and inhibition 102
7.3 Noise can drive lateral inhibition 103
7.4 Distinct mechanisms make use of distinct regulatory logic 104
7.5 Modular candidates for multi-functionality 106
7.6 Complete atlas of multi-functional circuits 108
7.7 Properties of hybrid and emergent multi-functional circuits 109
7.8 Underlying mechanisms of hybrid circuits 112
7.9 Underlying mechanisms of emergent Activation-Inhibition circuits 114
7.10 Underlying mechanism of emergent Pattern Convertor 116
7.11 Continuous pattern transitions 119
7.12 Discontinuous pattern transitions 120
7.13 Universality of multi-functional designs 123

8.1 Impact on evolvability of alternative mechanisms to produce a phenotype 132

List of Tables

7.1 Robustness of multi-functional hybrid and emergent motifs 110

Part I

Global Introduction

One challenge in modern biology is to understand how complex molecular circuits in the cell perform a variety of sophisticated regulatory functions. These biological circuits integrate regulatory events at multiple levels but for simplicity are generally referred to as *gene circuits*. The relationship between the structure of gene circuits and their functionality, i.e. the biological function they perform, lies at the heart of this thesis.

0.1 Gene circuits

0.1.1 Gene circuits drive particular biological functions

Gene expression is a complex process, which can be regulated at multiple levels – RNA transcription, RNA splicing, transport (in eukaryotes), translation, post-translational modifications, phosphorylation of proteins, degradation, extracellular signalling and many others. Gene circuits are abstract models that incorporate all of this biological complexity. The concept of a gene regulatory circuit consists of interactions between a closed set of genes [Kumar and Bentley, 2003] commonly represented as a directed graph. A particular circuit is made of nodes –genes– and edges –interactions– that define how one gene affects the expression of another gene in a positive or negative fashion. In that sense, gene circuits do not aim at explaining a particular step of the regulatory process but instead consider the resulting causal relationship between the gene expression levels of a group of genes.

Most models of gene regulation are concerned with genes that encode for transcription factors, proteins involved in regulating the expression of other genes at the level of RNA transcription [Wolpert et al., 2002]. Indeed, transcription factor gene products act in combination to generate a lot of the molecular circuitry that controls both developmental pattern formation and the differentiation of single cells into a variety of cell types [Davidson and Erwin, 2006].

Gene circuit models have been used to provide intuitive understanding of a variety of biological functions. The term *function* can describe very distinct biological processes. For example, one biologically essential function is to convert a graded stimulus into a binary, all-or-none cellular response [Shah and Sarkar, 2011]. This switch-like behaviour is critical to processes ranging from control of lineage commitment and memory to maintenance of gene expression levels [Alon, 2007]. In other cases, switch-like behaviour can lead to a bi-modal distribution in a cell population, resulting in a mixed phenotype that can better respond to a fluctuating environment [Wolf and Arkin, 2003]. Another interesting biological function is a particular dynamical behaviour observed in many signalling systems called adaptation [Ma et al., 2009]. Biochemical adaptation is a biological function that consists of the ability of a system to respond to a change in input stimulus, and then to reset back to its original prestimulated output level. This sensory function is commonly used to accurately detect changes in the extra-cellular environment and to maintain homeostasis in the presence of perturbations. Examples of adaptation range from the chemotaxis of bacteria [Berg and Brown, 1972], to the response of yeast cells to high external osmolarities –osmo-response– [Muzzey et al., 2009] and calcium homeostasis in mammals [El-Samad et al., 2002].

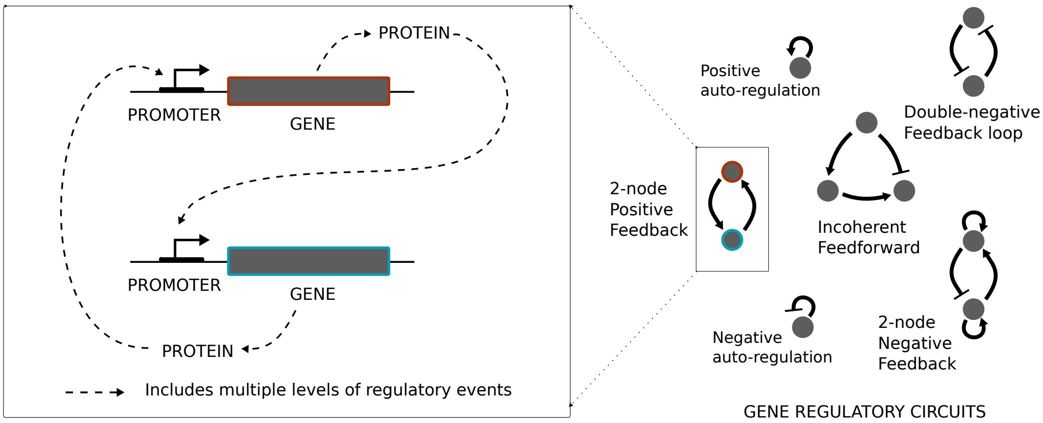

PROTEIN
PROMOTER
GENE
PROMOTER
GENE
PROTEIN
Includes multiple levels of regulatory events
Positive auto-regulation
Double-negative Feedback loop
2-node Positive Feedback
Incoherent Feedforward
Negative auto-regulation
2-node Negative Feedback
GENE REGULATORY CIRCUITS

Figure 0.1: **Gene circuits integrate multiple regulatory events.** Gene circuits are abstract objects that model the complex process of gene regulation. They are represented as directed graphs where nodes represent genes that encode for transcription factors and edges define how the expression of a given node affects that of another node or gene in a positive of negative fashion.

Gene circuits control particular biological functions. For example, both switch-like behaviours and perfect adaptation can be achieved by very simple gene circuits. In the first case, a simple positive autoregulation, i.e. a transcription factor that enhances the transcription of its own gene (see circuit in Figure 0.1), is capable of a switch-like behaviour that can lead to memory effects or to commitment to a particular cell differentiation state [Maeda and Sano, 2006, Becskei et al., 2001, Burrill and Silver, 2010]. In the second case, an incoherent feedforward circuit – a circuit in which the two arms of the feedforward loop act in opposition, one repressing and the other activating (see circuit in Figure 0.1) – can achieve adaptation [Ma et al., 2009].

Notice that these examples concern small gene circuits responsible for distinct biological functions. However, the transcriptome of a cell is composed of thousands of genes interacting in a highly complex fashion. A number of studies support the idea that complex circuits can be divided into simpler circuits that perform particular underlying functions. In order to understand which features of the whole transcriptome are responsible for a given function, it has been proposed that small gene circuits embedded in the entire transcriptome are responsible for distinct cellular functions. Indeed, small circuits such as negative autoregulations and feedforward loops are thought to perform precise functions in the cell and appear in hundreds of gene systems in E.coli and yeast, as well as in other organisms [Shen-Orr et al., 2002, Milo et al., 2002, Alon, 2007]. These findings have raised the hope that the functioning of large networks can be understood in terms of elementary circuits. In Part III of this thesis we will discuss more deeply the idea of the 'deconstruction'of complex circuits into smaller sub-circuits. Whether we agree with the decomposability of complex circuits or not, it is undeniable that the study of small gene circuits provides a minimal representation that allows an intuitive understanding of a particular process [Alon, 2007, Ma et al., 2009, Hornung and Barkai, 2008, Cotterell and Sharpe, 2010, Shah and Sarkar, 2011, Chau et al., 2012, Lim et al., 2013].

0.1.2 How to model a gene circuit

Gene circuits are modeled by simulations of the change of the state of the gene products with time. We will discuss here two prevalent models: the boolean and continuous formalisms.

The boolean formalism relies on the simplification of gene states to a binary variable: on/1 or off /0. For this type of formalism, the response function of genes to inputs must also be idealized to on or off, i.e. either the gene is transcribed or it is not [Thomas, 1973, Kauffman, 1993a]. Genes – nodes – within a boolean circuit respond to inputs using distinct logic functions. Logic functions follow particular combinatorial transcription logic, as they lead to transcription only if a precise combination of requirements is met. Among the most commonly used logic functions are AND and OR gates. A gene responding following an AND gate will be transcribed only if all its inputs are present/on. In contrast, a gene functioning as an OR gate will be transcribed if either of its inputs is present. The output of a particular boolean circuit is represented by a binary vector of 0's and 1 's that represents the transcription state of all genes within the circuit (Fig. 0.2A). Boolean circuits can easily describe the state of a single cell, as a given differentiation state is associated to a particular combination of transcribed and silenced genes [Ciliberti et al., 2007].

The continuous approach is currently the most widely used model for gene circuits

[Kumar and Bentley, 2003]. Indeed, gene expression in real biological systems can take on a wide range of intermediate concentration values. Also, real interactions between genes depend on a variety of factors such as the distinct molecular binding affinity of a transcription factor on a given promoter, or the number of distinct *cis* regulatory regions [Setty et al., 2003, Prud'homme et al., 2007, Carroll, 2008, Ben-Tabou de-Leon and Davidson, 2009]. These features of real biological systems disagree with the all-or-none response inherent in a boolean formalism. Hence, continuous gene circuits consider that both protein concentrations and gene-gene interactions can hold a wide range of real values.

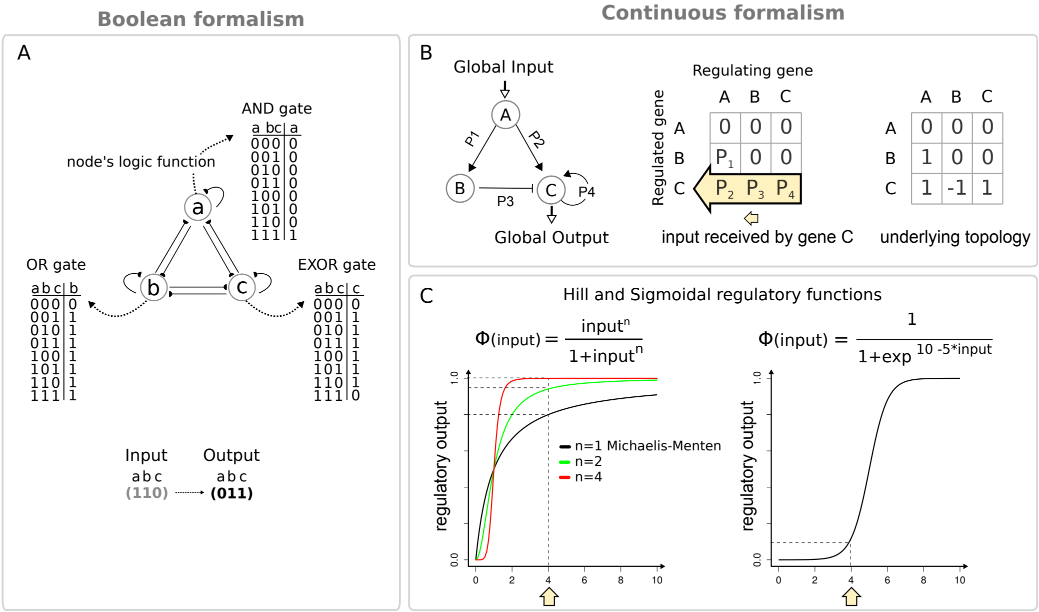

Boolean formalism
Continuous formalism
A
AND gate
a bc | a
000 | 0
001 | 0
010 | 0
011 | 0
100 | 0
101 | 0
110 | 0
111 | 1
node's logic function
a
b
c
OR gate
a b c | b
000 | 0
001 | 1
010 | 1
011 | 1
100 | 1
101 | 1
110 | 1
111 | 1
EXOR gate
a b c | c
000 | 0
001 | 1
010 | 1
011 | 1
100 | 1
101 | 1
110 | 1
111 | 0
Input
a b c
(110)
Output
a b c
(011)
B
Global Input
A
B
C
P1
P2
P3
P4
Global Output
Regulating gene
Regulated gene
A B C
A 0 0 0
B P1 0 0
C P2 P3 P4
input received by gene C
A B C
A 0 0 0
B 1 0 0
C 1 -1 1
underlying topology
C
Hill and Sigmoidal regulatory functions
Φ(input) = input^n / 1+input^n
Φ(input) = 1 / 1+exp 10 -5*input
regulatory output
1.0
0.0
0 2 4 6 8 10
n=1 Michaelis-Menten
n=2
n=4
regulatory output
1.0
0.0
0 2 4 6 8 10

Figure 0.2: **Distinct formalisms to model a gene circuit.** (A) Gene states in a boolean circuit are binary (ON/1 or OFF/0). Each node behaves as a logic gate following a particular combinatorial logic. (B) The total input received by gene C (yellow arrow) can be calculated from the concentration of all genes and the parameters P_2, P_3 and P_4 (C) Regulatory function describing the relationship between the total input into a gene and its output concentration. The regulatory function can take many forms such as Hill and Sigmoidal functions.

The *topology of a given circuit* describes the nature of the interactions between genes and can be represented by a matrix where 1, -1 and 0 represent activation, repression and no interaction respectively (Fig. 0.2B, right table). The *gene circuit itself* can be represented by a similar matrix where the previous values (the 1's, -1's and 0's) are replaced by real parameters that represent particular strengths of gene interactions (Fig. 0.2B, center table).

Given this type of continuous description of a gene circuit, its gene regulatory behaviour (i.e. how the concentrations of the genes vary over time) can be modeled by non-linear ordinary differential equations (ODEs) that describe the rate of production and decay of each gene as a function of the concentrations of all the genes in the circuit. Thus, there is a differential equation for each gene of the circuit. Taking as an example the circuit in Figure 0.2B:

$$\frac{dX}{dt} = \Phi(A, B, C) - \lambda X \tag{1}$$

where $X = \{A, B, C\}$, $\Phi(input)$ is the regulatory function that describes the relationship between the total input into a gene and the rate of change of its output concentration and λ is the decay. In the majority of models, the total input is taken as the sum of all incoming inputs onto the gene which are later 'filtered 'through the regulatory function. However, this 'add and filter 'approach implies that information on how individual genes combine to induce transcription is lost. In order to rescue the combinatorial logic of boolean models, a few studies have proposed distinct strategies where individual contributions are first 'filtered 'through the regulatory function then later added or multiplied –respectively analogous to the functioning of ON and AND logic gates [Uzkudun et al., 2015]. The following three equations describe (2) an 'add and filter 'case, (3) a 'filter and add 'case or (4) a 'filter and multiply'case for gene C in the circuit Figure 0.2B:

$$\Phi(P_2A + P_3B + P_4C) \tag{2}$$

$$\Phi(P_2A) + \Phi(P_3) + \Phi(P_4C) \tag{3}$$

$$\Phi(P_2A) * \Phi(P_3) * \Phi(P_4C) \tag{4}$$

Depending on the biological process being modeled, a regulatory function can take many forms, such as *Michaelis-Menten* or *Sigmoidal* functions (Fig. 0.2C). Each regulatory function holds a precise biological meaning. For example, Michaelis-Menten functions were first used to described basic enzymatic reactions involving a substrate reacting with a given enzyme to form a complex, which in turn is converted into a product [Murray, 1989]. It has been proposed that the transcription process can be considered to be such a reaction, as the substrate is replaced by a transcription factor, the enzyme by the RNA polymerase, and the product by the transcribed protein [Karlebach and Shamir, 2008]. Michaelis-Menten functions are sometimes replaced by Hill functions to account for cooperative phenomena (the $n > 1$ curves in Fig. 0.2C, left plot). Cooperativity describes the presence of more than one binding site in an enzyme, thus the need for various substrate molecules in order to start an enzymatic reaction. The analogy with the transcription process would be when various transcription factor molecules, or possibly distinct occupancies of cis-regulatory regions, are essential to start transcription.

Similar to the cooperativity phenomena, a sigmoidal curve does not lead to a response unless a certain condition is satisfied. A sigmoidal curve only leads to a significant response if the input is above a certain threshold. In that sense, sigmoidal curves sharpen their response to ‘all or nothing ’behaviours that resemble on-off boolean systems.

However, the use of a particular function over another is rather arbitrary. Indeed, as shown in Figure 0.1, gene regulatory circuits incorporate many complex biological processes and it is hard to envisage how they must follow rules from simple enzyme-substrate biochemistry. Regulatory functions serve only to describe in an abstract manner the input-output effective response of a gene. Unless there exist enough experimental data to describe all regulatory details of the system [Schaerli et al., 2014], it is hard to decide on which regulatory function is the most adequate to be used. Thus, most gene regulatory models use either Hill or Sigmoidal functions indiscriminately.

The choice between boolean and continuous formalism depends on the biological process of study. Boolean models show several advantages. As they are computationally less expensive than any other formalism, they are suited to capture features of large circuits. However, the boolean approximation constitutes a disadvantage when comparing to specific real gene circuits for which gene concentrations and gene interactions are inherently continuous. Certain pattern formation mechanisms are impossible to explore using boolean systems. For instance, a spatial gradient cannot be modeled using a basic boolean mode since it requires, by definition, intermediate values of gene concentrations (see section 0.2.1). Furthermore, a Turing reaction-diffusion model (see section 0.2.2) requires particular proportions between distinct diffusion constants, proportions that obviously do not exist in a boolean framework.

Commonly, gene regulatory circuits are used to describe processes occurring inside a cell, such as cascades of interactions within a signalling pathway. However, the concept of a gene regulatory circuit can be extended to understand spatial systems such as the ones mentioned in the previous paragraph. Indeed, multicellular pattern formation has been modeled where the same gene circuit is embedded inside every cell [Salazar-Ciudad et al., 2000, Cotterell and Sharpe, 2010]. The individual cells in a spatial system can communicate in distinct manners. On one hand, they can communicate to one another though diffusible products. In this case, a given product can diffuse and affect the concentration of that same product in neighboiring cells. On the other hand, cells can communicate through direct cell-cell communication – or juxtacrine signalling – where the binding of a particular ligand to its receptor in a neighbouring cell can trigger changes in gene expression of gene products in the cell expressing the receptor.

In this thesis we will explore multicellular pattern formation driven by small gene regulatory circuits. We have chosen to use a continuous formalism due to the following reasons. First, we are interested in the dynamical properties of gene circuits. For this, continuous models are the most appropriate to capture dynamical features. Second, accurately exploring pattern formation necessitates taking into account real aspects of biological circuits, such as the existence of intermediate concentration values or the formation of morphogen graded profiles.

0.2 Pattern formation in multicellular systems

The phenomenon of pattern formation poses a simple but fundamental question: how does a group of originally identical cells develop into a spatially organized set of different cell types? Many different mechanisms of pattern formation have been shown experimentally and can be split into three main categories: cell autonomous mechanisms, morphogenetic mechanisms, and inductive mechanisms [Salazar-Ciudad et al., 2003] (Fig. 0.3).

Cell autonomous patterning mechanisms do not require any mechanical or signalling interactions between cells. Instead, they all involve a particular cellular behaviour: mitosis. The division of a heterogeneous egg provides an example of a cell autonomous mechanism. Indeed, with few exceptions (mammalian and some turbellarian clades), egg cells usually divide in a heterogeneous manner. This occurs due to differences in concentrations of distinct mRNAs and proteins within distinct regions of the cell. Thus, as the egg cell divides, each daughter cell inherits a different complement of RNA and proteins that will determine its cell state in the resulting patterned tissue.

Morphogenetic mechanisms can be defined as those that change the relative arrangement of cells over space without affecting their states. Most morphogenetic mechanisms involve a prior pattern of cells of multiple types. The spatial reorganization of those cells/regions results in a new pattern. One mode by which cells can rearrange their relative positions is by simple migration. For example, the migration of neural crest cells in the mouse appears to be directed towards a chemical gradient of fibroblast growth factors – FGFs – in a migration process called chemotaxis [Kubota and Ito, 2000].

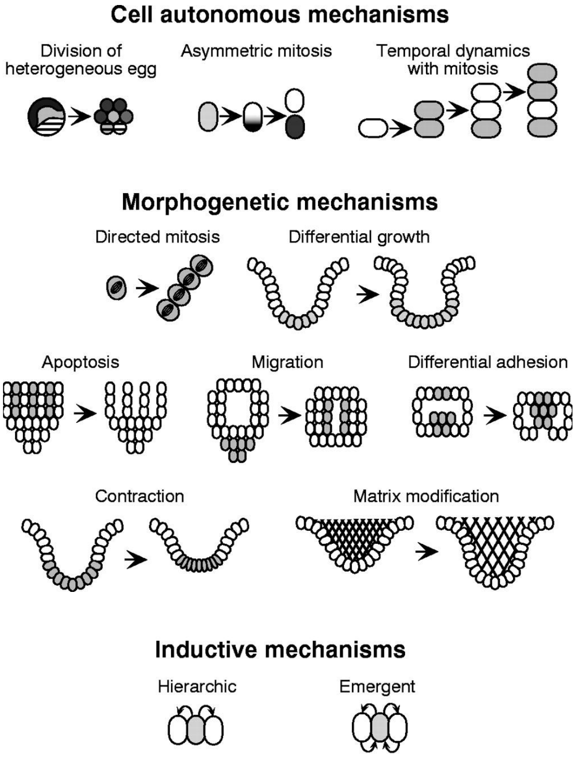

Cell autonomous mechanisms
Division of heterogeneous egg
Asymmetric mitosis
Temporal dynamics with mitosis
Morphogenetic mechanisms
Directed mitosis
Differential growth
Apoptosis
Migration
Differential adhesion
Contraction
Matrix modification
Inductive mechanisms
Hierarchic
Emergent

Figure 0.3: **Basic developmental mechanisms for pattern formation.** *Division of an heterogeneous egg*: different parts of the egg bind different molecules (indicated by different shading) resulting in different blastomere cells. *Asymmetric mitosis*: molecules are differentially transported into different parts of a cell resulting in different daughter cells. *Internal temporal dynamics coupled to mitosis*: cells that have oscillating levels of molecules before their division can produce spatial patterns. *Directed mitosis*: consistently oriented mitotic spindles may direct tissue growth. Differential growth: cells dividing at a higher rate (gray) can alter tissue shape. *Apoptosis*: transformation of an established pattern into another can result from apoptosis affecting specific cells (gray). *Migration*: cells can migrate to a new location. *Adhesion*: a change in pattern can result if a set of cells have differential adhesion properties (strong adhesion among gray cells). *Contraction*: differential contraction of cells can cause buckling of a tissue. *Matrix swelling, deposition, and loss*: matrix swelling can cause budding. *Hierarchic induction*: inducing cell (gray) affects neighbouring cells but the induced cells (white) do not affect the production of the inducing signal. *Emergent induction*: inducing cell affects neighbouring cells, which in turn signal back affecting the production of the inducing signal. Extracted from [Salazar-Ciudad et al., 2003]

Lastly, inductive mechanisms involve one cell influencing the fate of another cell. Cells can affect each other by secreting diffusible molecules, by means of membrane-bound molecules, or by chemical coupling through gap junctions.

Thus, in inductive mechanisms, tissue pattern changes as a direct consequence of changes in cell state. The simpler version of this type of mechanisms is the hierarchical one: where one cell (or tissue) influences the fate of another cell (or tissue) without being affected itself in return. In the more complex emergent form, two or more cells reciprocally influence each-others fate.

Inductive mechanisms for pattern formation have been explored for several decades [Turing, 1952, Wolpert, 1969, Artavanis-Tsakonas et al., 1995, Wolpert et al., 2002]. We will now discuss in more detail three distinct strategies to form a pattern: Wolpert's French flag model, Turing's reaction-diffusion model and lateral inhibition.

0.2.1 Hierarchical mechanisms: the gradient-threshold model

Lewis Wolpert first proposed in 1968 that cells receive 'positional information 'that allows them to develop distinct cell fates depending on their location within a tissue [Wolpert, 1968, Wolpert, 1969]. He used the metaphor of the French flag to illustrate the functioning of his model (Fig. 0.4). The French flag is a simple pattern in which three distinct fields of cells represent three different cell fates. One way in which cells could adopt their identity is by acquiring information about their position in space. Cells would 'compute 'or 'measure 'how far they are from a particular biological boundary, then use this information to adopt different fates (blue, white or red). For this, Wolpert proposed that this positional cue is provided by diffusible molecules, or morphogens. His model assumed that there exist a source and a sink of the morphogen at either end of the tissue, resulting in a morphogen gradient. Hence, morphogens are chemical species whose concentrations vary in space, and can thus control cell fate specification in a spatially dependent manner. Their graded signal acts directly on cells, in a concentration-dependent manner, to specify gene expression changes and cell fate selection [Ashe and Briscoe, 2006, Ibañes and Izpisúa Belmonte, 2008, Kicheva et al., 2012]. The simplest interpretation of the French flag model describes a gradient-threshold mechanisms in which cells could respond in 3 different ways – blue, white and red – depending on the concentration of the gradient – high medium and low concentrations respectively – making use of distinct thresholds. However, real biological systems are far more complex and empirical evidence has typically identified up to seven distinct thresholds [Stathopoulos and Levine, 2002].

Since Wolpert's first proposal, morphogens have been proven to act as graded positional cues that control cell fate specification in many developing tissues [Balaskas et al., 2012, Takei et al., 2004] – from plant roots to the *Drosophila* blas-

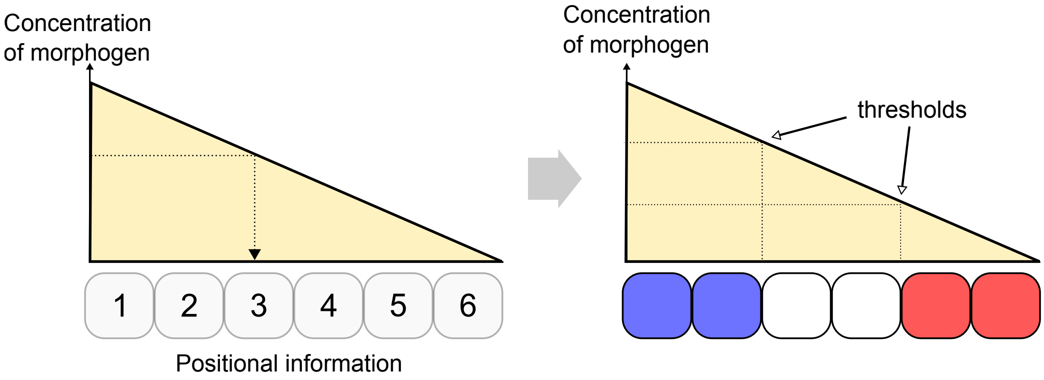


Figure 0.4: **The gradient-threshold model.** Within the French Flag conceptual framework [Wolpert, 1969], a pre-established fixed concentration gradient (input) is interpreted by a one-dimensional row of cells into different cell fates through a threshold-dependent mechanism. The concentration of morphogen is indicated by the height of the orange triangle. At different thresholds of morphogen concentration (2 thresholds are shown by the dashed lines in the right panel) cells take different fates as indicated by the blue, white and red colours.

toderm, imaginal discs, and the vertebrate neural tube. In these different tissues, morphogen gradients can span large spatial regions using different transport mechanisms [Ibañes and Izpisúa Belmonte, 2008]: from simple diffusion arising from the random motion of molecules [Crick, 1970] to active processes such as endocytosis and exocytosis of vesicles carrying the molecule [Tanaka et al., 2005] or cell to cell transport through gap junctions [Esser et al., 2006]. The formation of morphogen gradients has been modeled through imaging of fluorescent protein-morphogen fusions over space. This fluorescent data has shown that a wide variety of gradients – such as maternal Bicoid in the *Drosophila* syncytium or Dpp and Wingless in the fly's wing – show similar approximately exponential shapes. Indeed, morphogens gradients are commonly modeled as exponential profiles with characteristic lengths that depend both on the diffusion and degradation rates of the molecules.

An interesting question not considered in Wolpert's model is how a continuous gradient is transformed into the discrete changes in gene expression that result in distinct cell fates. Is the gradient read out in a defined time-window of development, or do cells continuously integrate the changes in morphogen concentration? [Kicheva et al., 2012]. The second alternative considers that cells interpret the total amount of morphogen they have been exposed to over time [Ahn and Joyner, 2004]. This alternative was explored by distinct theoretical models on the segmentation of *Drosophila* along the anterior-posterior axis, where the graded distributions of maternally provided transcription factors Bicoid and Nanos active the downstream target gap genes [Jaeger et al., 2007, Manu et al., 2009a]. These models showed that the experimentally observed spatio-temporal changes in gap gene expression in the syncytial embryo could occur even when morphogen signalling is maintained constant. Instead, changes in gene expression are driven by the dynamics of the underlying gap gene regulatory circuit. In part II of this thesis we will focus on how the dynamics of gene circuits is required to interpret morphogen gradients.

Although morphogen gradients can show different dynamics with respect to their formation, in this thesis we will not be concerned with how their gradients are formed, but instead with how their steady-states are interpreted. By doing so, we uncouple the specification of positional information from how that information is interpreted.

0.2.2 Emergent mechanisms: Turing's reaction-diffusion model and direct cell-cell patterning

In 1952 Alan Turing proposed a model whereby periodic patterns could form in multicellular systems [Turing, 1952] (Fig. 0.5). The model involves two chemical species that can react with each other and diffuse through the tissue, hence the common name of reaction-diffusion model. The mechanism works by 'diffusion driven instability'whereby amplification of small fluctuations can lead to a spatial pattern of these chemicals. At that time, the model was counter intuitive, since diffusion is usually seen as a force for evening out inequalities rather than as the driving force for the generation of an asymmetry.

The two components of the model are an activator and an inhibitor. While the activator promotes its own expression and that of the inhibitor, the inhibitor represses that of the activator. If the reaction and diffusion constants are set righto appropriate values (both species interact with adequate interaction strengths and the inhibitor diffuses faster than the activator) then an initially homogeneous concentration can spontaneously break the uniform state and form periodic patterns –peaks and valleys of concentration–which in 2D may take the form of spots of stripes.

As interest in Turing's model has increased in the last decade, a number of studies have successfully combined mathematical models with experimental approaches to show how the structural arrangement of many distinct tissues is controlled by reaction-diffusion mechanisms [Sheth et al., 2012]. These patterned tissues usu-

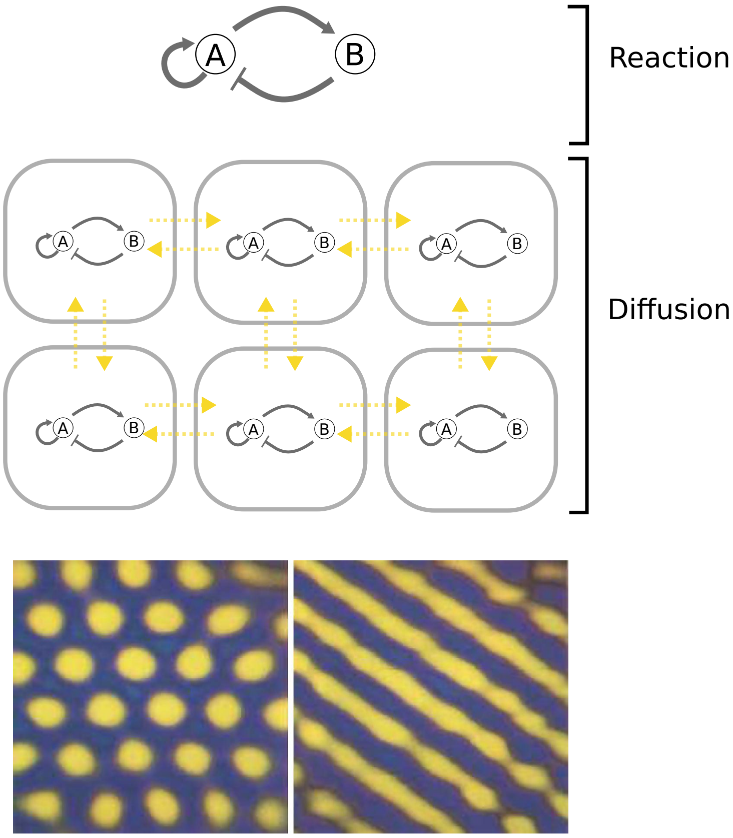


Figure 0.5: **Turing's reaction diffusion model** (Top) Two species; A an auto-activator and B a repressor. (Middle) In a two dimensional spatial context (as illustrated by the circuit in multiple cells) both A and B can diffuse (yellow dashed arrows). (Bottom) With appropriate parameters spots or stripes of A will form (concentration of A indicated by intensity of yellow) Adapted from [Ouyang and Swinney, 1991]

ally involve the positioning of repetitive evenly-spaced structures. As such, the first molecular evidence for a reaction-diffusion driven pattern was found in the patterning of skin appendages [Jung et al., 1998]. Particular molecules – Sonic Hedgehog (SHH) and a member of the Fibroblast Growth Factor family (Fgf4) acted as activators, while Bone Morphogenetic proteins (Bmps) acted as inhibitors – were shown to drive the formation of feather primordials. Recently a Turing-type mechanism was shown to control the patterning of digits during vertebrate limb development, where the authors uncovered the identity of the molecules that produce the periodic pattern of digital and interdigital fates [Raspopovic et al., 2014].

Other organs and tissues in which reaction-diffusion mechanisms control patterning include pigment patterning in the butterfly [Nijhout et al., 2003], the formation of ridges on the roof of the mammalian palate [Economou et al., 2012], and airway branching during lung development [Menshykau et al., 2012].

A distinct example of an emergent pattern is lateral inhibition (Fig. 0.6), which operates by direct cell-cell communication. As previously mentioned, this type of pattern arises as cells regulate each-others'states through membrane-bound receptors and ligands. Lateral inhibition is a fine-grained pattern in which adjacent cells are found in alternating gene expression states.

The notch and delta genes play a key role in lateral inhibition [Artavanis-Tsakonas et al., 1995, Kimble and Simpson, 1997, Lewis, 1998]. Both the ligand Delta and its receptor Notch are transmembrane proteins, and ligand-receptor binding between adjacent cells causes activation and subsequent cleavage of Notch. The released Notch intracellular domain (NICD) acts as a transcriptional regulator, leading to regulation of both Notch and Delta expression. Lateral inhibition models show that NICD downregulates Delta expression. Thus, production of Delta in one cell represses the transcription of Delta in the adjacent cell, leading to the formation of two radically different states Delta-on/Noch-off state and Delta-off/Notch-on state, which lead to different cell fates.

Lateral inhibition drives patterning of a diversity of tissues. For example, these types of mosaic patterns are a ubiquitous feature of epithelia: as bristle precursor cells differentiate from an initially homogeneous tissue of epithelial cells in *Drosophila* [Heitzler and Simpson, 1991]. In vertebrates, lateral inhibition mechanisms also governs cell fate decisions between neurons and non-neuronal cells in the central nervous system [Kawaguchi et al., 2008, Pierfelice et al., 2011], and hair cells and supporting cells in the inner ear [Daudet and Lewis, 2005, Petrovic et al., 2014].

A critical difference between patterns generated from a morphogen gradient compared to emergent patterns is how individual cells process information. Along a morphogen gradient, each position is exposed to a unique morphogen concentration, giving cells precise information about where they are in the field. Thus, cells make appropriate cell fate choices according to this positional information. Instead, as an emergent pattern arises, its particular position in space does not provide a cell with a unique identity. Cells with a maximal 'peak 'concentration have no way to distinguish which peak they are in – they do not have unambiguous positional information. This major difference makes a distinction between positional information and self-organizing patterns [Marcon and Sharpe, 2012]. A possible way to experimentally distinguish between both mechanisms is their scalability in a growing tissue. Most morphogen gradients scale with the size of a tissue. Hence, a stripe generated using a hierarchical mechanism would enlarge proportionally to tissue growth. Instead, as a tissue grows, stripes generated by an emergent mechanism would increase in number while the spacing between them – their wavelength – is conserved.

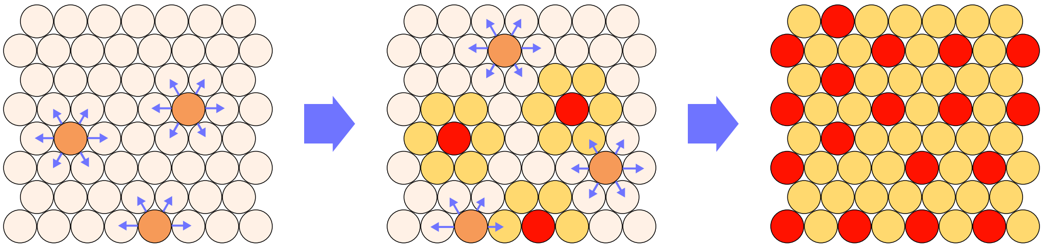


Figure 0.6: **An illustration of lateral inhibition in proneural clusters in the *Drosophila* neuroectoderm.** Neuroectoderm cells are represented by gray circles. Cells of a proneural cluster are shown in pink. Once cell at random will become the neuroblast (darker orange and then red) and inhibit the other cells in the cluster (lighter orange) from taking the same fate using the notch-delta pathway. These other cells will go on to become epidermis like the rest of the neuroectoderm.

While a positional information model – Wolpert 's model – is mostly relevant to regionalization (e.g. positioning the head at one end of the embryo and the tail at the other), self-organizing emergent models – Turing's reaction-diffusion and the lateral inhibition models – are most relevant to repetitive periodic patterns.

In this work we have chosen to focus on both types of inductive patterning strategies, i.e. both positional information and self-organizing patterns. We explore how circuits of genes and their products follow distinct inductive mechanisms to generate pattern formation. In Part II we use a model of positional information where cells communicate through diffusible products to interpret a morphogen gradient into a pattern of distinct cell fates. Instead, in Part III we explore self-organizing patterns that arise as cells regulate each-others'states through membrane-bound receptors and ligands. While the results of Part II have been published under the title 'The dynamics of gene circuits shape evolvability 'in Proceeding of the National Academy of Science, the results of Part III are currently being prepared for publication.

0.2.3 Role of noise in pattern formation

Molecular noise in gene expression is due to the occurrence of biochemical processes in the cell that work at low molecule numbers. These stochastic fluctuations are ubiquitous and can affect several levels of gene regulation [Eldar and Elowitz, 2010, Raj and van Oudenaarden, 2008]. The first concept that we need to consider is that of bursts. Proteins are produced in stochastic bursts because their gene's promoters stochastically switch between on/off states. This results in bursts of messenger RNA production, later amplified to generate corresponding protein bursts. A second important concept is that of propagation. Rates of gene expression are influenced directly by the levels and states of transcription factors and other upstream components that are themselves subject to bursting. As a result, fluctuations in the expression of one gene propagate to generate fluctuations in downstream genes. The gene circuits that regulate biological functions are subject to both noise-generating bursts (intrinsic noise) at the level of their individual components and noise-propagation (extrinsic noise) at the level of gene-gene interactions.

Noise plays two contrasting roles in pattern formation. On one hand, noise can be seen as a nuisance during developmental processes. As such, during the development of an embryo, noise needs to be overcome in order to pattern tissues in a reliable and reproducible manner [Lander, 2013]. From this deterministic point of view, development processes need to be robust to noise.

On the other hand, noise can also have a positive and functional role in pattern formation. It can drive probabilistic differentiation strategies that function across cell populations. In other words, as cells switch between distinct cell fates in a stochastic fashion, a variety of 'salt and pepper' type of arrangements are observed at the tissue level. A number of noise-dependent patterning mechanisms have been observed during development. An example of probabilistic differentiation in stem cells is found in the early mouse embryo, where the inner cell mass gives rise to distinct epiblast and primitive endoderm lineages. The epiblast develops into embryonic tissues and expresses the pluripotency regulator Nanog, whereas the primitive endoderm produces extraembryonic tissue and expresses the transcription factor Gata6. Before any morphological separation between the two fates, individual cells begin to express only one transcription factor or the other, in a heterogeneous 'salt and pepper' fashion. Subsequently, cells of the same type 'sort out' to generate the final spatial arrangement [Dietrich and Hiiragi, 2007, Yamanaka et al., 2010, Morris et al., 2010].

A distinct example in which noise can drive pattern formation is the lateral inhibition described above. The mutual transcriptional repression of Delta between adjacent cells is equivalent to an inter-cellular (double-negative) positive feedback loop. This cell-cell positive feedback loop has the ability to amplify small initial differences between cells. In this manner, noise can cause initial asymmetries to amplify and trigger the formation of a fine-grained pattern [Collier et al., 1996, Barad et al., 2011].

In this thesis we use biologically realistic gene circuits that include noise in gene expression. We explore the two contrasting roles of noise in pattern formation. In part II cells interpret a morphogen gradient to produce developmental patterns robust to noise. In contrast, in part III some of the patterns observed in our direct cell-cell model arise due to noise.

0.3 Relationship between structure and function of gene circuits

The structure of a circuit and the biological function it performs do not hold a one-to-one relationship. By structure we mean the wiring design of a circuit, i.e. its topology. In the first part of this section (0.3.1) we discuss how the structure of a circuit defines or constraints its function. Then, in the second part (0.3.2) we consider how a given function can be achieved by many different circuits. In the third part (0.3.3) we discuss the concept of dynamical mechanisms. Last, the fourth part (0.3.4) introduces dynamical systems theory.

0.3.1 From structure to function

Inferring function from structure

A particular type of study predicts the function a circuit performs from its regulatory structure [Shen-Orr et al., 2002, Milo et al., 2002]. These studies, inspired by Uri Alon's pioneering work [Alon, 2007], are based on detecting network motifs within large transcriptional regulatory networks. Network motifs are defined as patterns of connection that occur much more often than would be expected in random networks. Once these ubiquitous motifs are detected, the cellular function they perform is extrapolated from their topology. A particular family of network motifs detected using this methodology is the feed forward loop (FFL). FFLs are frequently found in transcriptomes of E.coli, yeast and other organisms. This type of motif consists of three genes, where one of the genes is regulated through two different feed forward paths, a direct one and an indirect one.

To take an example to consider in more detail, of the eight possible structural types of FFLs [Alon, 2006], the coherent type-1 (C1-FFL) is the design in which all three regulatory interactions are activations (Fig. 0.7). Alon et al. [Alon, 2007] have assigned a particular biological function to this motif: that of 'persistence

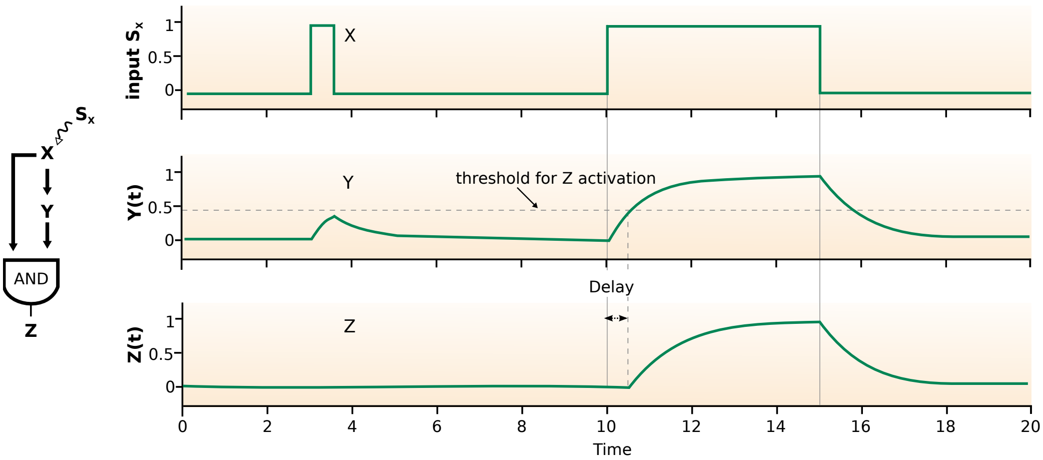


Figure 0.7: **The C1-FFL motif behaves as a persistence detector when the Z promoter behaves as an AND gate.** When the signal S_x appears, X becomes active and rapidly binds its downstream promoters. As a result, Y begins to accumulate. However, owing to the AND input function, Z production starts only when Y concentrations crosses the activation threshold for the Z promoter. This results in a delay of Z expression following the appearance of S_x. In contrast, when the signal is removed, X rapidly becomes inactive. As a result, Z production stops because deactivation of its promoter requires only one arm of the AND gate to be 'shut off '. The delay that is generated by this motif can be useful to filter out brief pulses of signal with no particular meaning for the cell. Adapted from [Alon, 2007]

detector'. The C1-FFL shows a dynamic behaviour in which gene Z shows a delay after stimulation, but no delay when stimulation stops (see details in Fig. 0.7). This way, a signal that appears only briefly does not lead to a response while only persistent signals lead to Z expression. This sensitive-delay serves a protective function. Indeed, as the environment of cells is often highly fluctuating, sometimes stimuli can be present for brief pulses that should not elicit a response. Importantly, this type of behaviour is only observed if gene Z integrates both X and Y inputs using a boolean AND gate. If the Z promoter switches to an OR logic, the same motif (C1-FFL) loses its previous function and adopts a new one. With an OR gate, Z is activated immediately upon an ON step of the signal, because it only takes one input to activate an OR gate. Instead, the motif with an OR gate will show a delay after and OFF step, i.e. when stimulation stops. Indeed, the OR gate C1-FFL can maintain expression of Z even if the input signal is momentarily lost. Thus, the C1-FFL can perform distinct functions –dynamical behaviours– depending on the regulatory logic in use. This is why predictions about the function of motifs must be tested experimentally in the different systems in which the motifs appear. Thus, the structure (i.e. the underlying topology) of a gene circuit alone does not define its function – the details of the circuit parameters are essential to describe the function it performs (see Fig. 0.2B, center table).

A single structure can achieve more than one function

A number of studies have shown that the same regulatory structure (i.e. circuit topology) can produce several biological functions [Panovska-Griffiths et al., 2013, Jaeger and Goodwin, 2002, Verd et al., 2014]. These distinct functions are achieved with distinct relative strengths of gene-gene interactions. Panovska-Griffiths et al. found a particular motif that they termed the AC-DC motif, the topology of which allows either switch-like or oscillatory behaviour, depending on parameter values [Panovska-Griffiths et al., 2013] (Fig. 0.8).

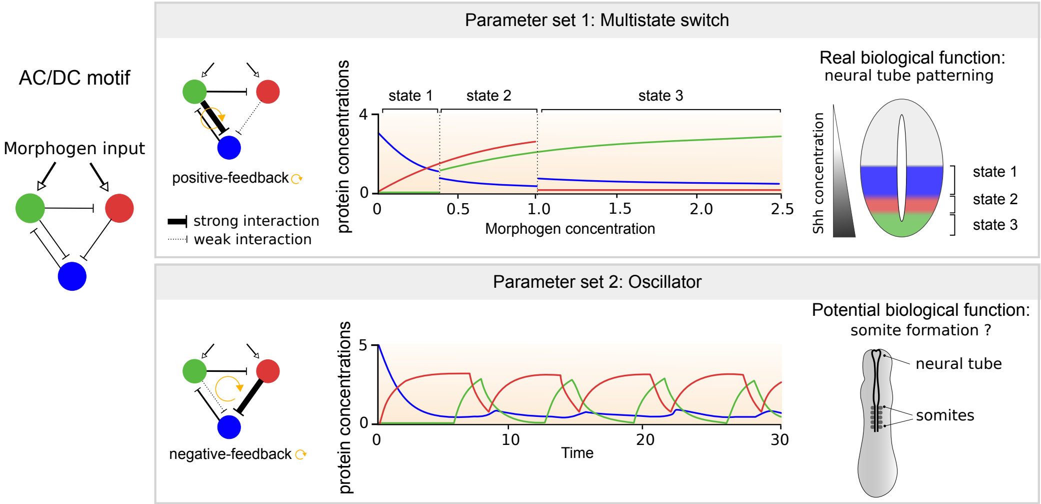

AC/DC motif
Morphogen input
Parameter set 1: Multistate switch
positive-feedback
strong interaction
weak interaction
protein concentrations
state 1
state 2
state 3
4
0
0
0.5
1.0
1.5
2.0
2.5
Morphogen concentration
Real biological function:
neural tube patterning
Shh concentration
state 1
state 2
state 3
Parameter set 2: Oscillator
negative-feedback
protein concentrations
5
0
0
10
20
30
Time
Potential biological function:
somite formation ?
neural tube
somites

Figure 0.8: **A gene regulatory motif that generates oscillatory or multiway switch outputs.** The AC/DC motif consists on repressive interactions between three transcription factors. Protein concentrations of the three genes in the circuit for two distinct parameter sets. Parameter set 1 behaves as a three-state switch that reproduces the formation of the three most ventral differentiation states in the neural tube. Parameter set 2 behaves as an oscillator (numerical profiles obtained at a fixed morphogen concentration) reminiscent of the oscillations observed during somite formation. The transition in dynamical behaviour is governed by the relative strength of the interactions shown as thick repressive arrows. Adapted from [Panovska-Griffiths et al., 2013]

The expression of the three transcription factors (Nkx2.2, Pax6 and Olig2) distinguishes three distinct domains of neuronal subtypes in the vertebrate neural tube. Indeed, in response to a gradient of the secreted protein Sonic Hedgehog (Shh), the vertebrate neural tube is patterned into three most-ventral discrete domains of expression characterized by high expression levels of Nkx2.2, Pax6 and Olig2 respectively. These transcription factors are connected through a set of repressive interactions in a motif termed AC-DC. Panovska-Griffiths et al. [Panovska-Griffiths et al., 2013] modeled the simple three-node AC-DC motif for distinct strengths of the regulatory interactions. Their results show that, for a given set of parameter values, the AC-DC motif is able to show a three-way multistate switch that reproduces the gene expression pattern observed in the neural tube. However, the AC-DC motif is also capable of producing oscillations in the expression profile of the three nodes in the circuit. Indeed, by gradually adjusting the strength of specific repressive interactions within the circuit, they show that it is possible to transition between the multistate switch to an oscillatory behaviour. Thus, for distinct parameter sets, the AC-DC motif shows distinct dynamical behaviours. In Figure 0.8 we show how the transition between behaviours is controlled by the relative strength of two interactions (thicker black interactions): the repression of the blue node by the green node and that of the blue node by the red one. The condition for a multi-state behaviour is that the repression green-to-blue is stronger than that of red-to-blue. In this case, a positive feedback loop forms between and the green and blue genes leading to a switch-like behaviour. Conversely, a strong red-to-blue repression approximates the circuit to a repressilator [Elowitz and Leibler, 2000], i.e. a three-component negative feedback loop exhibiting oscillations. These temporal oscillations are reminiscent of those occurring during the process of somitogenesis, by which somites form along the anterior-posterior axis of the developing embryo. Hence, from an evolutionary point of view, the same motif can be used to interpret a spatial morphogen gradient into either discrete domains of gene expression or to generate oscillatory patterns. In general, motifs such as the AC-DC illustrate that the knowledge of the topology of a circuit is not sufficient to understand the biological function it performs.

0.3.2 From function to structure

Before going into further detail, the relationship between a biological function and the structure of a gene circuit needs to be put into the broader context of genotype-phenotype maps.

Genotype and phenotype are interpreted here as broad concepts holding multiple meanings. The phenotype refers to any observable trait of an organism – from its morphology to its behaviour. We will use it to refer to very distinct concepts, from the secondary structure that an RNA molecule folds into, to any of the biological functions discussed previously. In contrast, the genotype refers to the heritable information encoded in DNA. A genotype can be represented by the direct nucleotide sequence of DNA or RNA or, more abstractly, by gene regulatory circuits that control gene expression.

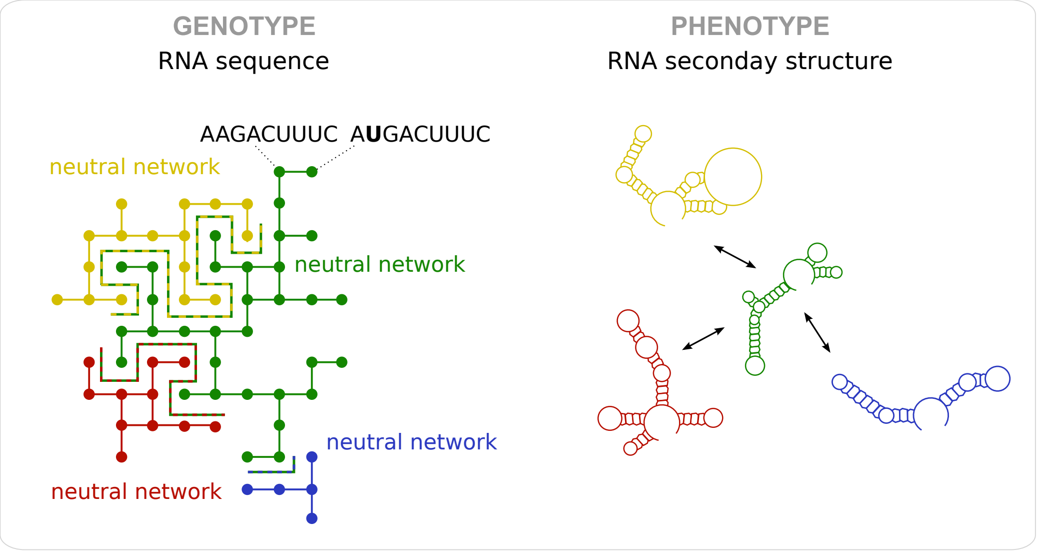


Figure 0.9: **Neutral spaces of a RNA sequence-to-shape model.** RNA combines in a single molecule both its genotype (nucleotide sequence) and phenotype (shape given by its secondary structure). In such a system, a neutral network comprises all sequences achieving a particular secondary structure. Adapted from [Fontana, 2002]

The correspondence between a whole collection of genotypes and a single phenotype was first studied in simple systems such as proteins and RNAs. Indeed, hundreds of animo-acid sequences can fold into the same three-dimensional protein structure [Smith, 1970, Li et al., 1996]. Likewise, hundreds of distinct RNA sequences can fold into the same secondary structure [Schuster et al., 1994]. It was precisely by exploring the relationship between the nucleotide sequence of an RNA molecule and its secondary structure, that Walter Fontana's first discovered the concept of neutral spaces (Fig. 0.9) [Fontana, 2002]. The layout of an RNA neutral network – a set of RNA sequences that fold into the same secondary structure – is the following: it consists on a large connected graph where two nodes – sequences – appear connected if the sequences differ by a single nucleotide substitution. As all connected nodes achieve the same secondary structure, this connectedness implies that, by mutating one nucleotide at a time, one can change the genotype – sequence – while preserving the phenotype – shape. (Notice that certain mutations (those that would cause jumps between the differently coloured neutral networks in Fig. 0.9) lead to the adoption of different phenotypes. We will cover this topic in part II). The connectedness of a neutral network constitutes its main feature: by definition, it forms a neutral space where mutations within it produce no change in the phenotype. Since Fontana's pioneering work, a variety of neutral networks have been explored [Ciliberti et al., 2007, Payne and Wagner, 2014]. Recently, an exhaustive genotype-phenotype map that comprises all DNA sequences encoding binding sites for a particular transcription factor was built [Payne and Wagner, 2014].

Importantly, the concept of the neutral network can be extended to study the relationship between structure and function of gene circuits. Many studies have shown that distinct gene regulatory circuits can produce the same biological function [Ciliberti et al., 2007, Ma et al., 2009, Ma et al., 2006, Cotterell and Sharpe, 2010, Chau et al., 2012, Lim et al., 2013, Munteanu et al., 2014, Schaerli et al., 2014, Cotterell et al., 2015]. This has been proven for a wide range of biological functions: from distinct temporal behaviours to particular patterning functions. We present next three distinct strategies to identify different gene circuits achieving a biological function of interest.

Evolutionary search

A distinct search strategy is based on *in silico* evolution, where a starting set of random circuits is permuted using genetic algorithms [Stich et al., 2007, François and Siggia, 2008]. At each round of evolution, these networks are tested for the target function, and a fraction of the best performing circuits is selected. The selected circuits are then used as the new pool that is subjected to further mutations (where different type of mutations can be explored such as addition/deletion of nodes, addition/deletion of links or subtle changes in parameters). After multiple rounds of mutation and selection (often hundreds of cycles) convergence to particular circuit structures is often observed.

The choice of fitness (more accurately termed a fitness function, since it assigns a number to a given circuit) is essential for this type of search and is defined according to the biological function of study. For example, François and Siggia [François and Siggia, 2008] performed an evolutionary search on gene circuits performing adaptation. As mentioned earlier, adaptation is observed in many sensory systems as the ability to transiently respond after input stimulus then reset back to the original steady-state output level [Ma et al., 2009]. The authors defined a fitness function for adaptation in terms of two functional metrics: the sensitivity – the change in output as a function of the stimulus received – and the precision – the difference between the output steady-state concentrations before and after the stimulus (Fig. 0.14A).

Evolutionary search strategies follow individual paths of evolving circuits. Along these paths either fitness continuously improves, or the path is discarded. This strategy has the advantage of highlighting circuits that can potentially be found by evolution – ones for which there exist a theoretically assessable path in parameter space between a starting circuit (of low fitness) and a final one (with improved fitness). However, evolutionary approaches exhibit a certain bias in their search, as they are more likely than other methods to get trapped at local fitness maxima, missing large areas of genotype space where more fit genotypes exist. In order to minimize this, the simulations described earlier are run many independent times.

Exhaustive enumeration

The last search strategy is based on enumerating all possible architectures within a complete circuit space and evaluating their ability to perform the function. This strategy has been used to reveal the genotype-phenotype maps of a wide variety of biological functions, from temporal behaviours –such as adaptation, sensitivity to noise, or switch-like cellular responses –to spatial patterning functions – such as cell polarization and morphogen interpretation. The advantage of this technique is that it provides unbiased results, including plausible circuit solutions that might be more difficult to reach through random or evolutionary processes. Furthermore, enumeration approaches can depict the entire neutral space for a given biological function. However, this kind of methodology is only possible if the space of possible circuit topologies is computationally feasible to analyze. Indeed, the main disadvantage of exhaustive enumeration approaches is their restriction to explore simple circuits. For this reason all mentioned studies consider only small circuits with two to five genes.

We next discuss two examples of exhaustive searches that built the neutral spaces of two distinct biological functions. Each example uses a distinct model formalism – boolean or continuous. When the model formalism is boolean, enumeration consists in the exhaustive list of all possible topologies. In contrast, if the model formalism is continuous, for each topology many distinct parameters sets are sampled.

The first study explores the neutral space of a particular boolean genotype-to-phenotype model. [Ciliberti et al., 2007] (Fig. 0.10). The phenotype is the steady-state of a small boolean circuit and consists of a binary vector of 1's and 0's that describe the transcription state of every gene in the circuit. A gene can either be transcribed (on/1) or not (off/0). As for the biological meaning, a particular phenotype can represent a given cell differentiation state, i.e. a particular combination of expressed transcription factors. Hence, using this particular genotype-phenotype model, the authors built a neutral space that consists of all boolean topologies

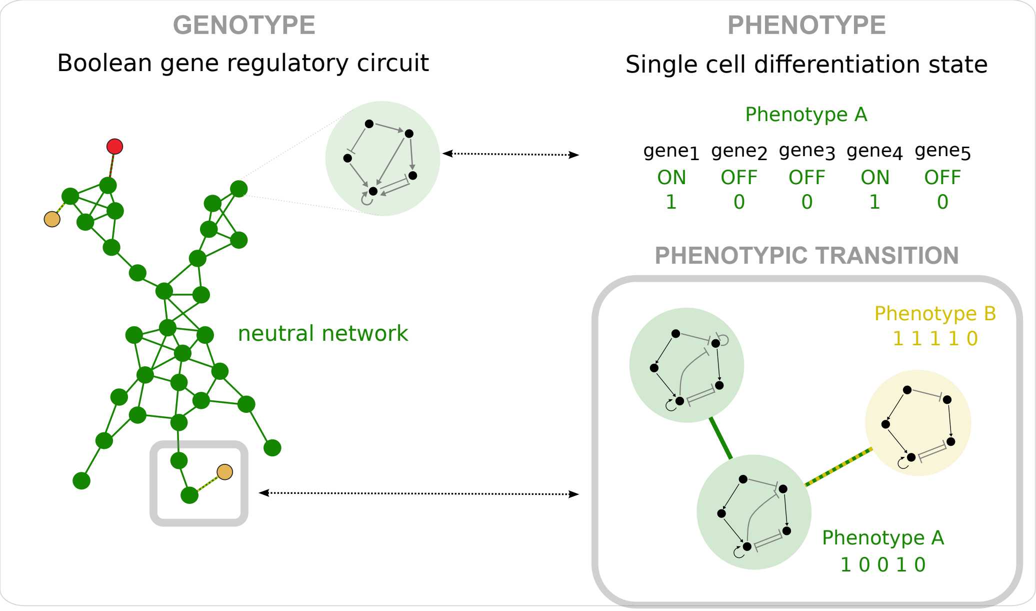

GENOTYPE
Boolean gene regulatory circuit
PHENOTYPE
Single cell differentiation state
Phenotype A
gene1 gene2 gene3 gene4 gene5
ON OFF OFF ON OFF
1 0 0 1 0
PHENOTYPIC TRANSITION
neutral network
Phenotype B
1 1 1 1 0
Phenotype A
1 0 0 1 0

Figure 0.10: **Neutral spaces of a boolean genotype-to-phenotype model.** A neutral space consists of all boolean topologies achieving a particular phenotype, i.e. a specific combination of transcribed (ON or 1) and untranscribed (OFF or 0) transcription factors. Mutations consist in the addition/deletion of single gene interaction. Certain mutations lead to a change in phenotype (i.e. transition into a novel neutral space). Adapted from [Ciliberti et al., 2007]

achieving a particular cell differentiation state (Fig. 0.10). As previously shown in the RNA sequence-to-shape model, the resulting neutral space forms a large connected graph where two gene circuit topologies appear connected if they differ by a single gene interaction, i.e. here mutations consist in the removal or addition of a gene interaction. Analogously, one can travel within a boolean neutral space performing a single structural change at a time while maintaining the phenotype, i.e. the cell differentiation state. Notice that this type of study explores the relationship between structure and function but does not take into account the concept of dynamical mechanism (see section 0.3.4).

The second study is especially relevant to this thesis as Part II is based on its results. Cotterell and Sharpe chose to address a critical patterning event in embryo's morphogenesis [Rogers and Schier, 2011]: the interpretation of a morphogen gradient by a field of cells such that they acquire different cell fates [Cotterell and Sharpe, 2010]. More precisely, the phenotypes of the study are the result of the interpretation of a pre-established, fixed concentration gradient (input) by a one-dimensional row of cells into a stripe-like expression pattern of one of the con-

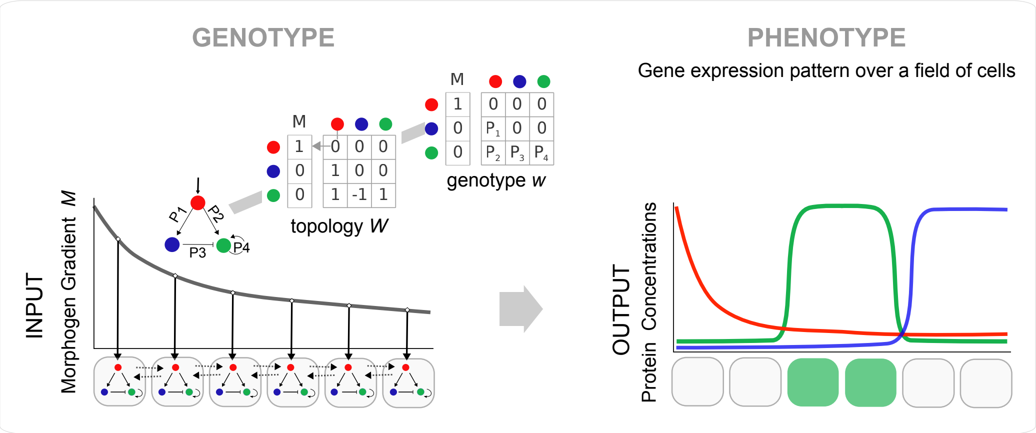


Figure 0.11: **Morphogen interpretation.** A pre-established concentration gradient (i.e. pre-pattern input) is interpreted by a one-dimensional row of cells into different cell fates. Genes regulate each others through a three-noded circuit embedded in each of the cells in the tissue. Additionally, cells communicate to one another through diffusive gene products (dashed arrows).

stituent genes (Fig. 0.11). The genotype is a three-noded circuit embedded in each of the cells in the tissue and consists on a particular topology –wiring design– plus a given set of continuous parameter values for the strengths of gene interactions. After an exhaustive sampling process (see 2.4 for more details), the authors could build a particular type of neutral space termed the ’complexity atlas’ holding all three-noded circuits capable of converting the input morphogen gradient into an output gene expression stripe.

Random sampling

When the number of possible configurations becomes too large to enumerate, several approaches propose to explore genotype space by random sampling. For example, the study by Ciliberti et al. [Ciliberti et al., 2007] (previous section 0.3.2) also explores the properties of neutral spaces for large boolean networks (composed of 20 genes). The authors use a Monte-Carlo algorithm that uniformly samples the space of all networks. The advantage of this methodology is that it provides unbiased results. However, neutral spaces cannot be exhaustively depicted as this approach is incomplete.

A last approach is that of *random boolean networks* [Kauffman, 1969, Kauffman, 1993b, Aldana et al., 2007]. The model consists of a set of N genes or binary variables each acquiring the values 0 or 1 corresponding to the two states of gene expression (off and on respectively as previously discussed). Each gene has exactly

K regulatory inputs randomly chosen from anywhere in the network. In addition, each gene is assigned at random one of the possible boolean logic functions of K inputs. Both the inputs and the boolean rules of each gene are chosen only once and remain fixed throughout the temporal evolution of the given network. Hence, once each gene has been provided with a set of inputs and a boolean rule, the network dynamics are then given by the simultaneous updating of all genes of the network. The number of possible configurations of N genes and K inputs per gene is enormous, even for modest N. Thus, in order to study the generic properties or random boolean networks, this type of approaches are based on random sampling of the possible configurations. Interestingly, random boolean networks are believed to be good candidates for the modeling of real gene regulatory networks. Large numbers of distinct networks converge to a type of cyclic attractors that can be interpreted as alternative cell types (networks with a common cyclic attractor are displayed in fan-like structures, see [Kauffman, 1969, Kauffman, 1993b] for more details). In a nutshell, random boolean network allow to explore particular boolean structures with a common biological function (cyclic attractor).

0.3.3 The concept of dynamical mechanism

As mentioned above, many gene circuits can perform a particular function. From this observation, several questions arise: How do these circuits operate to perform that function? How many of these circuits are interestingly different, in terms of the way they operate? Circuits use a limited number of strategies to produce a given phenotypic function. These few fundamental solutions are referred to as *dynamical mechanisms* [Ma et al., 2009, Ma et al., 2006, Cotterell and Sharpe, 2010, Chau et al., 2012, Lim et al., 2013, Munteanu et al., 2014, Jaeger and Sharpe, 2014]. The concept of dynamical mechanisms is tightly related to that of the *minimal motif*. Let us next briefly discuss what precisely we mean by minimal motif and dynamical mechanism.

Minimal motifs

A number of studies have explored the idea that complex circuits can be deconstructed into simpler circuits that underlie function [Alon, 2007, Lim et al., 2013]. This means that in order to understand which features of the whole circuit are responsible for a given function, one needs to reduce the complexity of a circuit until only the minimal necessary gene interactions remain. Hence, reducing the complexity of a circuit to its minimal configuration(s) allows us to identify the core underlying mechanism(s) for a given biological function.

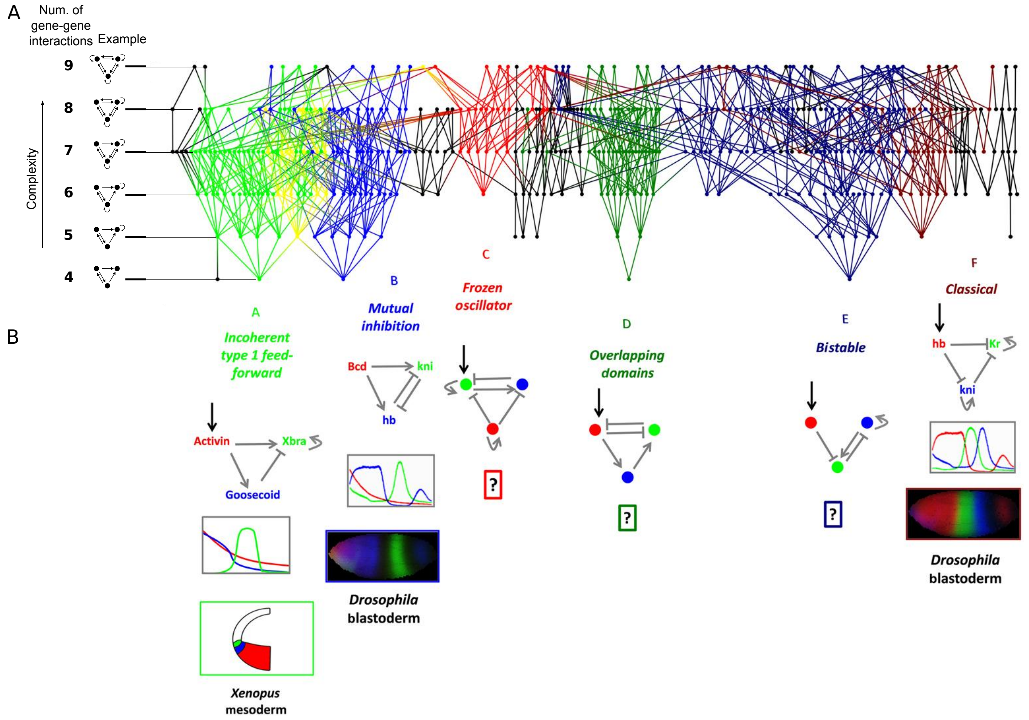

A
Num. of gene-gene interactions
Example
9
8
7
6
5
4
Complexity
B
A
Incoherent type 1 feed-forward
Activin
Xbra
Goosecoid
Xenopus mesoderm
B
Mutual inhibition
Bcd
kni
hb
Drosophila blastoderm
C
Frozen oscillator
?
D
Overlapping domains
?
E
Bistable
?
F
Classical
hb
Kr
kni
Drosophila blastoderm

Figure 0.12: **Complexity atlas showing the six fundamental mechanisms for morphogen interpretation.** (A) The lay-out of the complexity atlas allows to identify six core motifs for stripe formation. The topologies of the complexity atlas are coloured according to the mechanism by which they produce the single stripe of gene expression (see Fig. 0.13). Topologies that are capable of performing multiple mechanisms are shown in yellow. Mechanisms occupy locally connected regions of the atlas, and together cover 78% of the topologies. Hence, circuits can be classified into distinct families of circuits according to their dynamical mechanism. (B) The known biological systems that each mechanism is associated with are shown beneath the circuits. Three of the mechanisms, Mutual Inhibition (*bicoid-hunchback-knirps*), Incoherent Feed-Forward (*caudal-knirps-giant*) and Classical (*hunchback-krüppel-knirps*) are involved in *Drosophila* anterior-posterior patterning [Jaeger, 2011], while Incoherent Feed-Forward controls the mesoderm inducer *Xenopus Brachyury* [Green, 2002]. Extracted from [Cotterell and Sharpe, 2010]

Cotterell and Sharpe used a complexity atlas to extract the minimal core motifs for stripe formation (Fig. 0.12) [Cotterell and Sharpe, 2010]. Indeed, the complexity atlas is a type of neutral network where the layout helps us to extract the minimal core motifs for the given phenotype (in this case, stripe formation). It is built from a topology-focused approach and comprises all topologies with at least one functional stripe-forming genotype. The rules of connectedness are those used in the previously described boolean neutral space: two gene circuit topologies – nodes – appear connected – edges – if they differ by a single gene interaction. In other words, two gene circuits are neighbours if the gain or removal of any one interaction can transform one of the circuit topologies into the other. The layout proposed by the authors orders topologies along the y-axis with respect to their complexity, i.e. number of regulatory links. This layout helped identify the six simplest minimal motifs for morphogen interpretation: Bistable, Incoherent feed-forwards, Mutual Inhibition, Overlapping Domains, Classical, and Frozen Oscillator, which appear at the bottom of the 'stalactites' in the complexity atlas (Fig. 0.12).

Dynamical mechanisms

Cotterell and Sharpe observed that each minimal stripe-forming motif corresponded to a distinct dynamical mechanism [Cotterell and Sharpe, 2010]. By 'mechanism' we mean the causal dynamics responsible for the trajectory of the system (i.e. the spatiotemporal course of gene expression) resulting in the final phenotype. Figure 0.13 shows how each of the six core motifs uses a distinct spatiotemporal course of gene expression in order to produce a stripe of gene expression. The authors could assign each of the genotypes producing a stripe to a given mechanism. They did so by comparing their spatiotemporal dynamics (see section 2.5). Each genotype could be assigned to only one of the six distinct mechanisms, proving that distinct mechanisms are unique and incompatible. The atlas appears colour-coded where each colour represents a distinct dynamical mechanism, and yellow is used for topologies that contain genotypes employing distinct mechanisms, i.e. distinct parameter sets lead to distinct dynamics (Fig. 0.12).

Dynamical mechanisms represent distinct, fundamental ways to achieve a given function (Fig. 0.14). Previous studies have also linked the concepts of minimal motifs and dynamical mechanisms [Ma et al., 2009, Hornung and Barkai, 2008, Shah and Sarkar, 2011, Ma et al., 2006, Chau et al., 2012, Lim et al., 2013, Munteanu et al., 2014]. Using an enumeration approach Ma et al. explored all three-gene circuits performing adaptation citeMa2009. The authors identified two minimal motifs – negative feedback and incoherent feed-forward types of motifs termed NFBLB and IFFLP respectively – that corresponded to two dis-

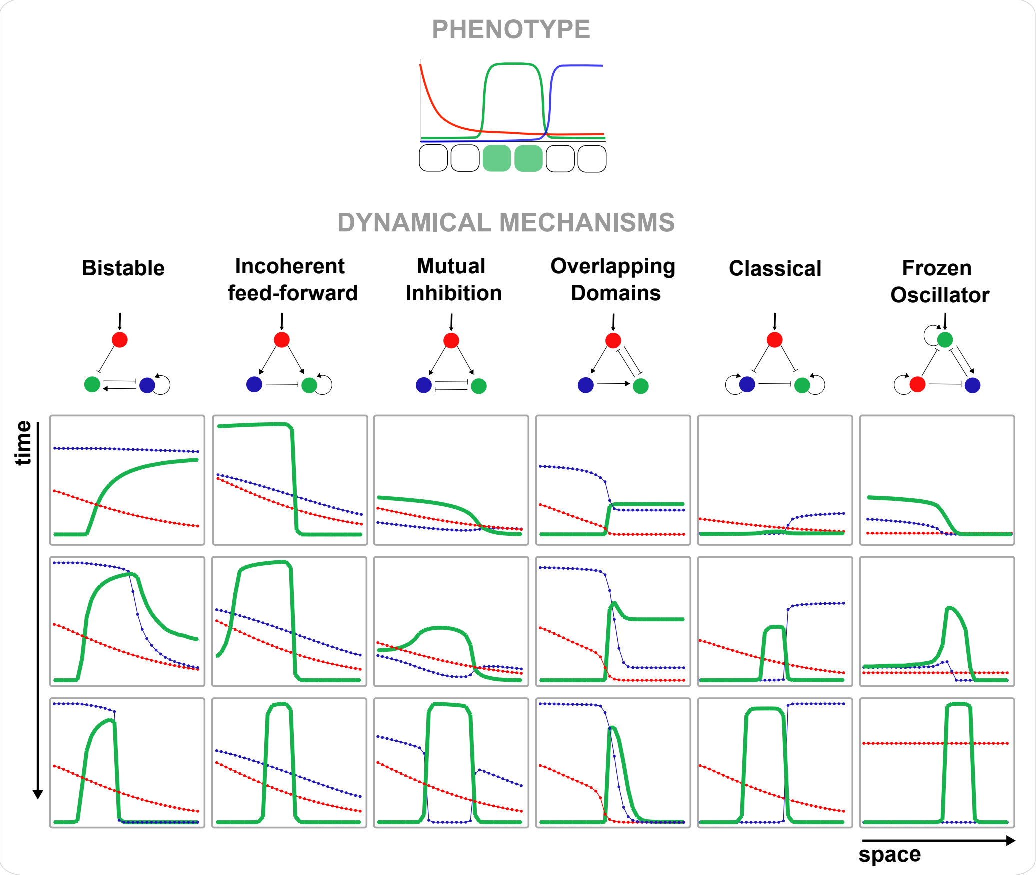

PHENOTYPE
DYNAMICAL MECHANISMS
Bistable
Incoherent feed-forward
Mutual Inhibition
Overlapping Domains
Classical
Frozen Oscillator
time
space

Figure 0.13: **Alternative mechanisms to achieve a single phenotype.** A stripe-forming circuit employs one of six distinct fundamental mechanisms [Cotterell and Sharpe, 2010]. Each mechanism uses a distinct gene expression dynamics in space and time to reach the same phenotype.

tinct dynamical mechanisms. The authors proved that each mechanism holds a distinct class of phase portrait, which results in distinct dynamic trajectories (see next section 0.3.4).

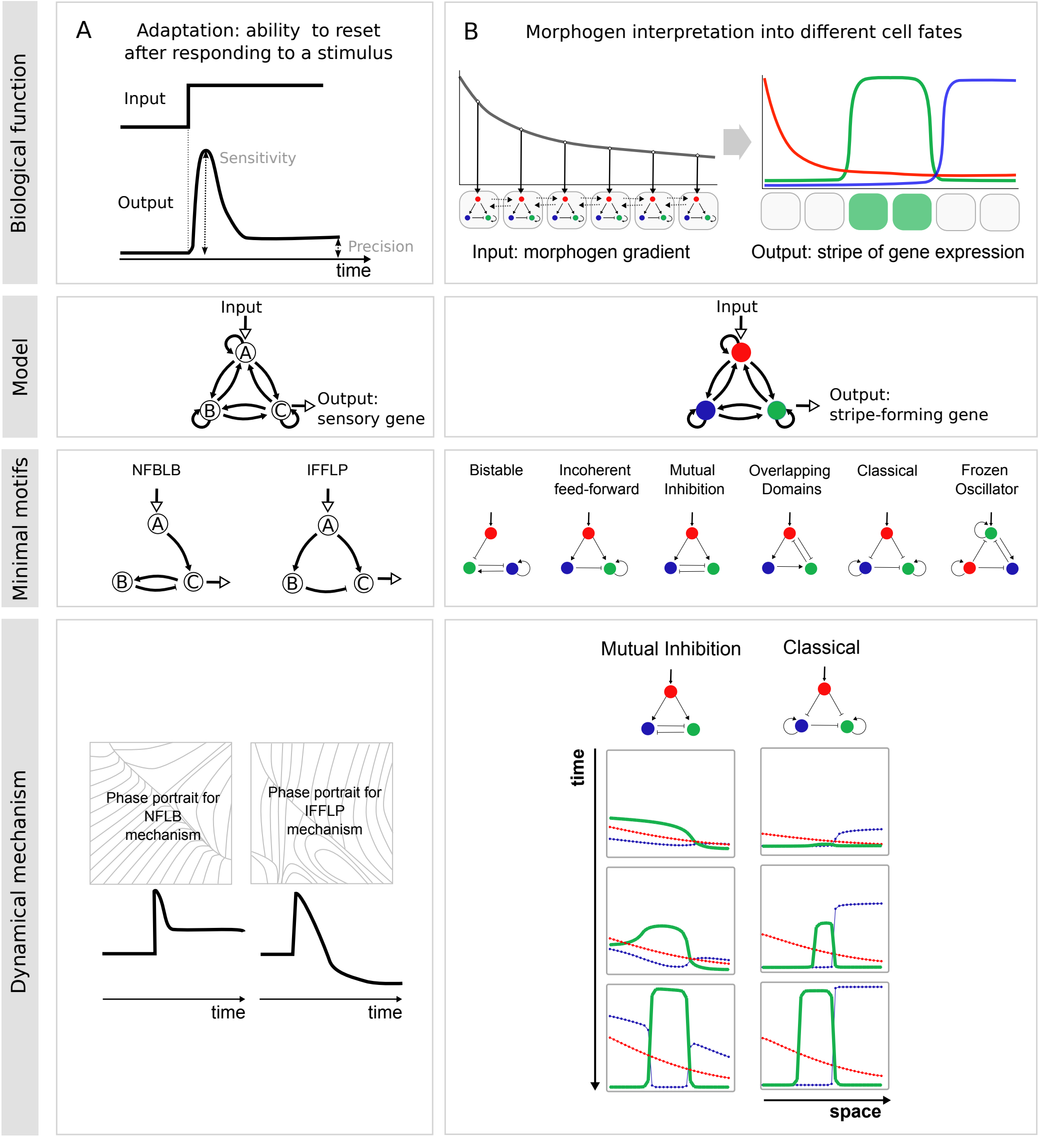

Biological function
Model
Minimal motifs
Dynamical mechanism
A
Adaptation: ability to reset after responding to a stimulus
Input
Output
Sensitivity
Precision
time
B
Morphogen interpretation into different cell fates
Input: morphogen gradient
Output: stripe of gene expression
Input
A
B
C
Output: sensory gene
Input
Output: stripe-forming gene
NFBLB
IFFLP
Bistable
Incoherent feed-forward
Mutual Inhibition
Overlapping Domains
Classical
Frozen Oscillator
Phase portrait for NFBLB mechanism
Phase portrait for IFFLP mechanism
time
time
Mutual Inhibition
Classical
time
space

Figure 0.14: **Minimal motifs and dynamical mechanisms.** (A) Ma et al. [Ma et al., 2009] identified the two minimal motifs for biochemical adaptation. They correspond to distinct dynamical mechanisms each holding a distinct class of phase portrait. (B) Likewise, the six minimal stripe-forming motifs correspond to six distinct dynamical mechanisms [Cotterell and Sharpe, 2010].

0.3.4 Dynamical systems

As illustrated in cases such as the ACDC or the C1-FFL motif (0.3.1), the study of the dynamical behaviour of a gene circuit is essential to understand its function. Dynamical systems theory provides intuitive understanding on the set of behaviours that can be implemented by a gene circuit. In this section we provide a brief introduction to dynamical systems and use as a conceptual model a double negative feedback loop circuit known as the toggle switch [Gardner et al., 2000]. We will introduce the following repertoire of concepts: *phase space, phase portrait, trajectory, steady state, attractor, saddle point, flow, nullcline, basin of attraction*, *separatrice* and *bifurcation* [Strogatz, 2014]. All of these concepts will be in use in Part III of this thesis.

The toggle switch is formed of two genes U and V that mutually inhibit each other (Fig. 0.15A). This circuit's dynamics can be formulated in terms of the following ordinary differential equations:

$$\frac{dU}{dt} = \frac{\alpha_1}{1 + V^{\beta_1}} - U \tag{5}$$

$$\frac{dV}{dt} = \frac{\alpha_2}{1 + U^{\beta_2}} - V \tag{6}$$

$$\tag{7}$$

where the regulatory functions and precise parameters ($\alpha_1 = 4$; $\alpha_2 = 4$;$\beta_1 = 3$; $\beta_2 = 3$) were extracted from Gardner's model [Gardner et al., 2000].

A system's behaviour is defined by its *trajectories*, which represent the change of the state of the system –e.g. consisting of a set of transcription factor concentrations over time. The shape of these trajectories depends on the structure or organization of the underlying regulatory circuit [Goodwin, 1982, Alberch, 1991, Waddington, 1939, Manu et al., 2009b, Verd et al., 2014, François and Siggia, 2012, Jaeger and Crombach, 2012]. Understanding the dynamical behaviour of the circuit comes from the geometrical analysis of phase space, i.e. the analysis of the number, nature and relative arrangement of the steady states of the system.

The *phase space* of a dynamical system is an abstract space, in which each dimension represents the value of a specific state variable. In the case of the toggle switch, the state variables represent the concentrations of U and V. The graphical representation of phase space is called the *phase portrait* and shows the rate of change of the system at any given state. This is known as the *flow* of the system. In the phase portrait of the toggle switch in Figure 0.15B, the flow is indicated by arrows of a given length and direction. If we follow the flow form all possible initial states, we obtain the totality of possible dynamical trajectories.

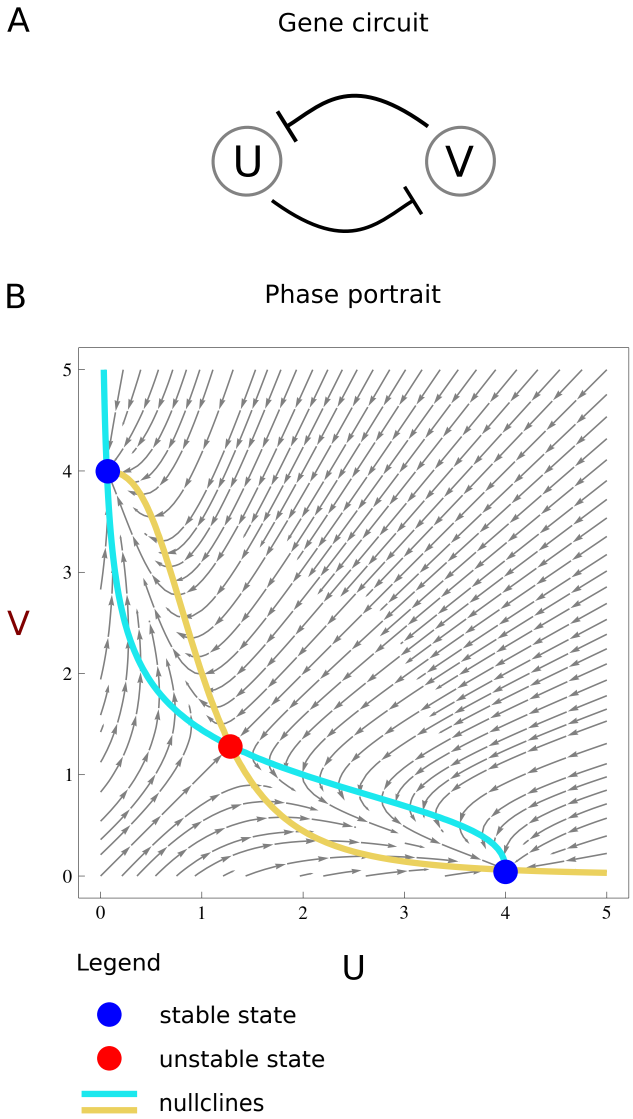


Figure 0.15: **Phase portrait of a toggle switch.** Phase portrait obtained using the equations of Gardner et al. [Gardner et al., 2000] (usage of *Mathematica*). Steady states are found at the crossing of the yellow and light-blue nullclines. The separatrix (not shown) is found in the diagonal of phase portrait and separates the bassins of attraction of the two stable states.

It is clear from the inspection of the flow that trajectories tend to converge to specific points in phase space: the *steady states* of the system. There are two kinds of steady states: stable states (or *attractors*) and unstable states. We plot steady states as points coloured according to their stability: stable states in blue and unstable states in red. These steady states are found where two *nullclines* cross. That is, nullclines are defined as the curves where either $\frac{dU}{dt} = 0$ (yellow nullcline in Figure 0.15B) or $\frac{dV}{dt} = 0$ (light-blue nullcline in Figure 0.15B). Hence, at the crossing of the two nullclines both conditions are true and the system is at steady state.

Attractors, as their name implies, draw trajectories towards them. Furthermore, they have the special property that once a trajectory has reaches an attractor, it will return to it if the system is slightly perturbed. The region of phase space around an attractor, from which all trajectories converge towards it is called its *basin of attraction*. Curves known as *separatrices* set apart the different basins and their attractors.

An example of an unstable steady state is a *saddle point*. Saddle points attract trajectories from some directions, but repel them in others. Usually, the system will move away from a saddle upon perturbation, towards the nearest attractor.

Finally, attractors and saddles, with their associated basins and separatrices, can be created or annihilated through the process of *bifurcation*. Bifurcation represents sudden qualitative changes in the structure of the phase portrait caused by small changes in the values of a given set of control parameters.

0.4 Objectives of this thesis

The relationship between the structure of a gene circuit and its functionality gives rise to the two main questions discussed in this thesis. This work will thus be structured into two parts.

In the first part we explore the following question: *To what extent does the dynamical mechanism producing a specific biological phenotype bias the ability to evolve into new phenotypes?* Indeed, biologists are familiar with the concept of developmental constraints [Maynard Smith et al., 1985]. From a given organism, which undergoes a particular sequence of morphogenetic events, not any new phenotype can evolve. Instead of considering the sequence of particular patterning events, we focus on the specific dynamical mechanisms employed to transition between them. As such, if two organisms both interpret a morphogen gradient into the same stripe of gene expression, but achieve it following distinct dynamics, do they have a different potential to evolve novel functions? Particularly, we assess the ability of the distinct gene circuits that produce a specific biological function to evolve into novel functions. The results show that the dynamical mechanism of a gene circuit impacts on its potential to evolve. We will address concepts such as *evolvability* and *dynamical mechanism*.

In the second part we raise the following question: *What is the minimal circuit design able to perform multiple biological functions?* Within large multi-functional gene circuits, each function can be allocated an independent set of interacting genes, or structural module [Di Ferdinando et al., 2001, Clune et al., 2013, Ellefsen et al., 2015, Kashtan and Alon, 2005]. Modules form independent structures that do not overlap and are thus believed to work quasi-autonomously [Wagner and Altenberg, 1996, Raff and Conway Morris, 1996, Kirschner and Gerhart, 1998a]. However, how does modularity evolve as the size of a circuit diminishes? How can a minimal circuit perform two distinct functions? We explore small minimal circuits capable of encoding two distinct patterning functions without any changes to their topology or modification of their regulatory parameters. Particularly, we discuss their design properties and their ability to be decomposed into distinct sub-circuits. We will address concepts such as *multi-functionality* and *modularity*.

Part II

Dynamical mechanisms of gene circuits shape evolvability

Chapter 1

INTRODUCTION

Evolution occurs through mutations on existing genotypes, potentially transforming the original phenotype or trait into a novel one, with latent beneficial consequences. It is a fundamental problem in biology to understand the relationship between a genotype and the associated phenotypes accessible through mutations, in other words, its *evolvability*. From the many definitions of evolvability [Pigliucci, 2008, Kirschner and Gerhart, 1998b], we refer here to the ability of genotypes to access novel phenotypes, irrespective of the subsequent process of natural selection.

What are the features of a genotype that impact on its capacity to evolve? We introduce three distinct approaches to explore this question. The first approach (1.1) explores evolvability from the perspective of genotype-phenotype maps. The second (1.2) studies how particular topological features of gene circuits influence their evolvability. We propose a third approach (1.3) that focuses on how the dynamical mechanism employed by a gene circuit impacts on its evolvability.

1.1 Evolvability from the building of genotype-phenotype maps

The first approach explores how phenotypic novelty arises in the context of neutral spaces. As previously mentioned, neutral spaces are large connected structures formed by all genotypes with a common phenotype where, by definition, mutations within the neutral space are neutral, i.e. they change the genotype while maintaining the phenotype.

In this context, the tendency of a given genotype to conserve its genotype upon mutation defines its *robustness*. On the contrary, its tendency to adopt a different phenotype upon mutation defines its capacity to evolve or *evolvability*. Phenotypic novelty thus occurs when mutations on a given genotype transform its original phenotype into a novel one. In other words, phenotypic novelty occurs when there is a transition from an original neutral space to a novel one.

The existence of neutral spaces has the following four consequences to the evolutionary process.

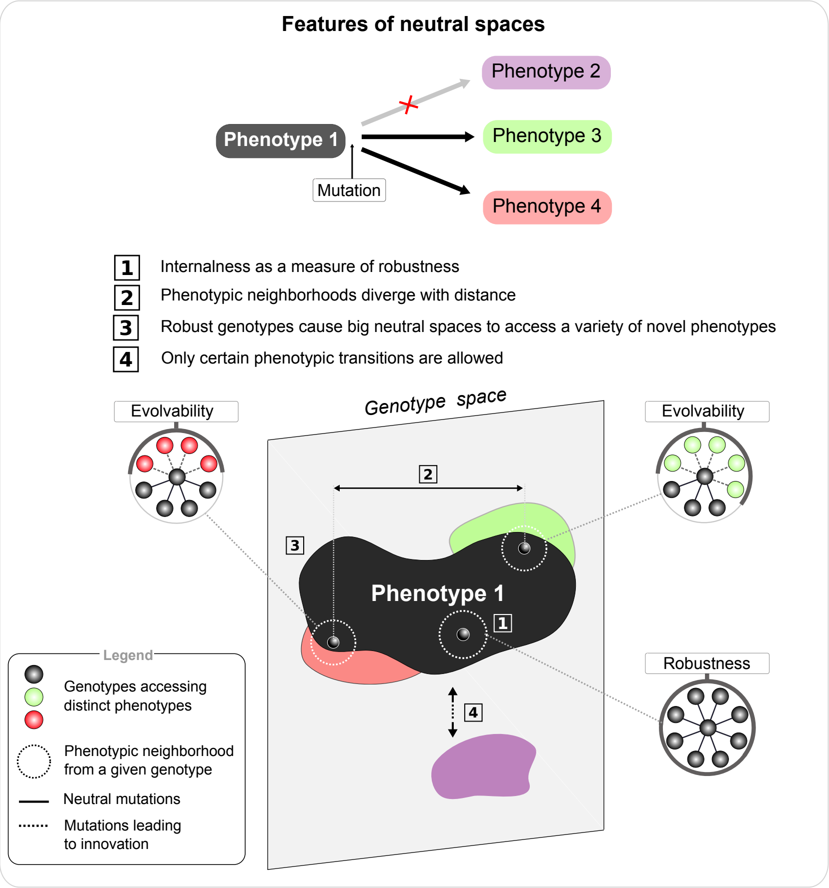

Features of neutral spaces
Phenotype 1
Phenotype 2
Phenotype 3
Phenotype 4
Mutation
1 Internalness as a measure of robustness
2 Phenotypic neighborhoods diverge with distance
3 Robust genotypes cause big neutral spaces to access a variety of novel phenotypes
4 Only certain phenotypic transitions are allowed
Evolvability
Genotype space
Evolvability
Robustness
Legend
Genotypes accessing distinct phenotypes
Phenotypic neighborhood from a given genotype
Neutral mutations
Mutations leading to innovation

Figure 1.1: **Features of neutral spaces.** Evolvability accounts for the accessible novel phenotypes, while developmental constraints imply certain hypothetical forms are not possible: phenotype 2 (purple) is not available by gradual mutation. Innovations accessible from a given genotype constitute its phenotypic neighbourhood. The arrangement and diversity of this neighbourhood is a measure of the genotype's evolvability. Genotype space is high dimensional, but we schematically represent it here in 2D for illustrative purposes.

First, a genotype's position within a neutral space critically determines its potential to evolve. Indeed, neutral spaces often appear as fully connected and dense regions [Ciliberti et al., 2007, Draghi et al., 2010, Payne and Wagner, 2014]. Therefore, although genotypes internal to the neutral space are highly robust to mutations (i.e. not evolvable), only genotypes close to the edge of the neutral space might access novel phenotypes ([1] in Fig. 1.1). This means the 'internalness'of a genotype within genotype space is critical to its ability to access novel phenotypes.

Second, the accessible innovations are critically determined by a genotype's position along the edge of a neutral space [Ciliberti et al., 2007, Draghi et al., 2010, Payne and Wagner, 2014]. These accessible innovations in the vicinity of a given genotype constitute its *phenotypic neighbourhood* [Dichtel-Danjoy and Félix, 2004]. A number of studies have linked the divergence in composition of phenotypic neighbourhoods to distances in genotype space. These results apply whether the genotype corresponds to a boolean gene circuit [Ciliberti et al., 2007], a protein sequence [Ferrada and Wagner, 2010] or an RNA sequence [Huynen, 1996]. This means, the closer two genotypes are, the more similar their phenotypic neighbourhoods. Conversely, the increase in distance between two genotypes correlates with the divergence in the innovations accessible from them ([2] in Fig. 1.1).

Third, robustness facilitates evolvability. This apparently contradictory statement is based in the following observation. While low-robust genotypes lead to small poorly-connected neutral spaces, high robust genotypes ensure the connectedness of a neutral space and thus impact of its size. In other words, high robustness of the individual genotypes that form a neutral space facilitates the formation of large connected structures. Large neutral spaces can then extend throughout genotype space, providing mutational access to a diversity of novel phenotypes from different genotypes ([3] in Fig. 1.1). In this manner, robustness allows for the accumulation of neutral mutations. This so-called neutral drift enables phenotypic innovation by improving the changes of encountering distinct novel phenotypes [Draghi et al., 2010, Hayden et al., 2011, Tóth-Petróczy and Tawfik, 2013, Aguirre et al., 2011, Manrubia and Cuesta, 2015]. As an example, Toth-Petrozy et al. [Tóth-Petróczy and Tawfik, 2013] studied how the accumulation of neutral sequence changes in a given enzyme improved its potential to acquire new protein functions, promoting its evolutionary adaptation. This process was called *cryptic genetic variation* [Hayden et al., 2011].

Fourth, from a particular phenotype, only certain novel phenotypes are accessible ([4] in Fig. 1.1). Already, Fontana and Schuster [Schuster et al., 1994, Fontana, 2002] observed in their RNA sequence-to-shape neutral spaces that it was easier to transition between specific pairs of phenotypes. Indeed, certain phenotypic transitions (indicated by thick lines with alternating colours in Figure 0.9) were more probable than others: for example the transition between green and yellow RNA shapes is highly feasible compared to green to blue transition. Furthermore, some transitions were not allowed – from yellow to blue shapes.

1.2 Evolvability from the topological features of gene circuits

The second approach explores the relationship between a gene circuit's evolvability and its topology [Salazar-Ciudad et al., 2001, Fujimoto et al., 2008]. Salazar et al. [Salazar-Ciudad et al., 2001] classified gene circuits able to produce multicellular patterns into two distinct categories based on their topology. These categories broadly corresponded to circuits containing either feed-forward of feedback motifs. They observed that the capacity of such networks to access novel patterns when subject to mutation depends on their topological features. This type of study aims at predicting the evolvability of genotypes from their particular topological features. We believe these results are not in contradiction with the previous observations. Instead, they provide a distinct and relevant perspective on evovability. However, we will focus on exploring evolvability through the use of neutral spaces.

1.3 Mechanism-view on evolvability

While many features of genotype-phenotype maps have been much studied, none of the existing studies have addressed the impact of the underlying dynamical mechanism of a gene circuit on its evolvability. As pointed out earlier, circuits achieving a biological function can be classified into families of core dynamical mechanisms. Do circuits that function under distinct mechanisms differ in their ability to evolve?

The existence of distinct dynamical mechanisms to achieve a single phenotype suggests the following mechanism-view on evolvability. That is, the discrete nature of mechanisms suggests that neutral spaces do not always form fully connected regions but might instead have sparse structures. Thus, when mechanisms are taken into account, the neutral space of a function breaks up into scattered islands of genotypes characterized by distinct underlying mechanisms. These observations suggest that evolvability may be constrained specifically by the dynamical mechanism of the gene circuit. As neutral spaces can be broken up into a discrete collection of separate islands, the process of neutral drift may be limited to these mechanisms-specific regions.

In order to assess the impact of dynamical mechanism, we take inspiration from the work of Cotterell and Sharpe [Cotterell and Sharpe, 2010] on alternative mechanisms for morphogen interpretation (Fig. 0.11- 0.13). Importantly, the authors had previously shown that mechanisms have a discrete nature, i.e. it was not possible to smoothly and functionally transition from one mechanisms to another.

The choice of a model for morphogen interpretation imposes to study evolvability from the perspective of pattern formation. The evolvability of circuits controlling gene expression patterns is especially relevant for the field of developmental biology. Indeed, the spatial organization of gene expression orchestrates cell differentiation. Their diversification causes evolution of both modest morphological traits, such as novel pigmentation patterns in the wings of a butterfly [Werner et al., 2010], and major evolutionary breakthrough, such as new body structures, for example the transformation of two-winged insects into four-winged insects [Prud'homme et al., 2007, Guerreiro et al., 2013].

For the current study, we analyzed each of the six mechanisms for morphogen interpretation independently and obtained a mechanism-specific measure of evolvability. We found that, indeed, the likelihood of accessing distinct phenotypic innovations is different for each dynamical mechanism, despite the fact that they all produce the same phenotype. Our analysis uncovers key features of the mechanistic neutral spaces and provides useful insight into how phenotypic transitions and thus innovations occur.

Chapter 2

METHODS

2.1 The phenotype

The phenotype of study consists on the interpretation of a pre-established morphogen gradient by a one-dimensional row of cells into different cell fates (Fig. 0.11). The choice on this particular phenotype is based on the knowledge that multiple dynamical mechanisms exist in order to interpret a concentration gradient into a stripe of gene-expression [Cotterell and Sharpe, 2010] (Fig. 0.12).

2.2 The genotype

The genotype is represented by a small three-gene circuit. The same circuit is embedded in each of the cells of the tissue and consists of a particular wiring design –topology– plus a given set of continuous parameter values for the strengths of gene interactions.

As previously defined in section 0.1.2, the topology is represented by a 3x3 matrix W_{ij} where value at the ith row and jth column indicates how gene j regulates gene i. An additional column indicates which gene receives the morphogen input (Fig. 0.11). In the process of parameter sampling (section 2.4), the values of the original topology W_{ij} will be replaced by real parameter values w_{ij} with a [0:10] range for activation and [-10:0] range for repression. Hence, a genotype holds a particular set of precisely 12 parameters: 9 for the strengths of gene interactions (w_{ij}) and 3 for diffusion coefficients of each individual gene.

2.3 The gene regulatory model

For reasons explained in section 0.1.2 of the global introduction, we describe a multi-cellular tissue using a continuous gene regulatory model. The simulation takes places in a one-dimensional row of 32 cells that communicate to one another through diffusive gene products (dashed arrows in Figure 0.11). We use reflective or zero-flux boundary conditions that do not allow any diffusion in or out of the system, therefore modeling the system as isolated from other tissues. Indeed, patterning fields are often boundary restricted zones that are patterned relatively independently.

The pre-pattern takes the form of a gradient described as an exponential profile (see section 0.2.1):

$$M = Id^c \tag{2.1}$$

where I is the morphogen concentration in the left-most cell of the field (set to 1); d is the reduction of morphogen concentration in each subsequent cell (set to 0.93) and c is the cell position. This pre-pattern does not change throughout the simulation and activates only one of the genes. Thus, although morphogen gradients can form according to distinct dynamics, we consider a steady-state gradient.

The model captures the spatiotemporal dynamics of gene patterning and is described by:

$$\frac{\partial g_{ij}}{\partial t} = \chi[\Phi[\sum_{l=1}^{3} w^{il} g_{lj} + M]] + D_i \nabla^2 g_{ij} - \lambda g_{ij} + \eta(t) g_{ij} \tag{2.2}$$

where g_{ij} is the concentration of the ith in the jth cell initially set to 0.1 for every gene in every cell; ϕ is the regulatory function that takes the form of a Michaelis-Menten function (see section 0.1.2); w is the matrix containing the strengths of gene-gene interactions; M is the morphogen input; χ is the Heaviside function to prevent negative gene product production rates; D_i is the diffusion constant for the ith gene and λ is the decay rate equal for all genes.

We chose to include multiplicative noise in the system through the $\eta(t)$ term, which adds uniformly distributed fluctuations ($\pm 1\%$) to the concentration of every gene in every cell at every time step. As mentioned in section 0.2.3, in this part of the thesis we exclusively model development processes that are robust to noise. All patterns studied are robust, i.e. they form in a reliable and reproducible manner in a noisy environment. Hence, noise is included in the model to distinguish between robust and non-robust solutions.

2.4 Exploring the space for solutions

Solutions of our search are genotypes able to interpret the morphogen gradient into a band of gene expression. We search for genotypic solutions in a two-step process. First, we exhaustively enumerate all possible three-gene circuit topologies. Second we sample large numbers of genotypes (i.e. parameter values) for each topology.

Following the method employed by Cotterell and Sharpe [Cotterell and Sharpe, 2010], we generated 9710 unique topologies. Next, we tested 10^6 random parameter sets for each topology. Parameters were generated randomly using uniform distributions of ranges [0:10] for interaction strengths and [0:0.05] for diffusion coefficients. We obtained 103,916 successful genotypes successful at producing a stripe with 702 distinct underlying topologies.

2.5 Building mechanism-specific neutral spaces

As described in section 0.3.3, the original complexity atlas build by Cotterell and Sharpe extracted the minimal core motifs for stripe formation. In the atlas, topologies appear colour-coded according to their mechanism. However, because the atlas is built from a topology-focused perspective, a single topology –node– can hold within it distinct genotypes (yellow topologies in Figure 0.12) making use of distinct dynamics. Because the focus of our study is to explore how distinct mechanisms affect the evolvability of gene circuits, we consider each mechanism separately. We thus propose to display the results through separated complexity atlases each containing all genotypes operating under a given dynamical mechanisms (Fig. 2.1).

For that, we use the same methodology employed in the original study where each genotype –$geno$– in the original atlas is assigned to the mechanism –$mech$– to which it holds the smallest distance $D_{geno-mech}$. We calculate this distance with the following method. On one hand, a particular genotype –$geno$– is characterized by an array Y that contains the values of its spatiotemporal gene expression. On the other hand, a mechanisms –$mech$– is characterized by the spatiotemporal expression arrays Y of all its 'core genotypes', i.e. all genotypes with the underlying core minimal topology of the mechanism. In order to find $D_{geno-mech}$, the spatiotemporal gene expression of a genotype is compared, one by one, to those 'core'profiles. For each genotype-to-genotype comparison, we use a Euclidean distance that measures the differences in time and space between two dynamical

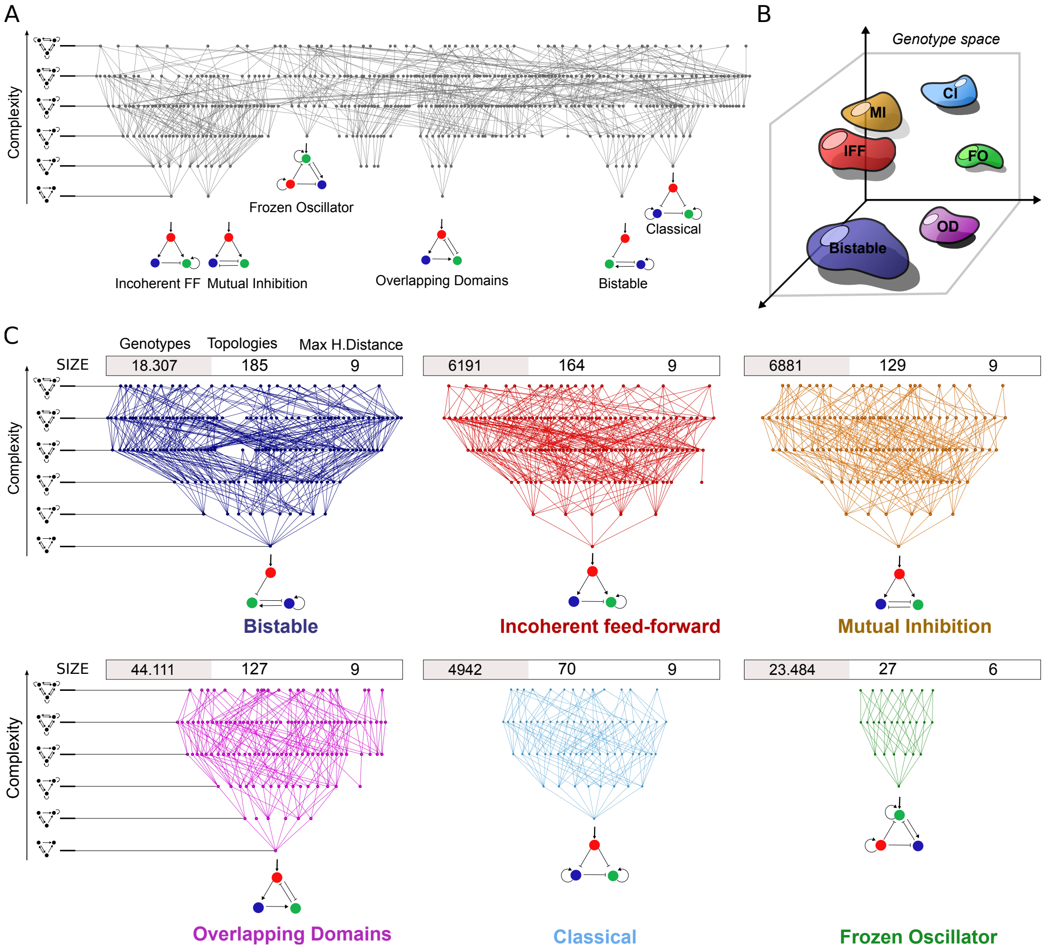

A
Complexity
Frozen Oscillator
Classical
Incoherent FF
Mutual Inhibition
Overlapping Domains
Bistable
B
Genotype space
CI
MI
IFF
FO
OD
Bistable
C
Genotypes
Topologies
Max H.Distance
SIZE
18.307
185
9
6191
164
9
6881
129
9
Complexity
Bistable
Incoherent feed-forward
Mutual Inhibition
SIZE
44.111
127
9
4942
70
9
23.484
27
6
Complexity
Overlapping Domains
Classical
Frozen Oscillator

Figure 2.1: **Dynamical mechanisms impose a scattered structure of neutral spaces.** The neutral space for morphogen interpretation is broken-up into six alternative mechanisms. (A) Original atlas [Cotterell and Sharpe, 2010] where core motifs represent each of the six alternative mechanisms for morphogen interpretation. (B) Mechanisms are found in disconnected regions of genotype space. (C) Building individual complexity atlases for each mechanism. The size of each mechanisms can be measured according to the number of genotypes, of topologies, or using Hamming Distance.

patterns. For example, the distance between two genotypes $geno_a$ and $geno_b$ is:

$$D_{geno_a - geno_b} = \frac{\sum_1^G \sum_1^C \sum_0^T \sqrt{(Y_{a,g,c,t} - Y_{b,g,c,t})^2}}{GCT} \tag{2.3}$$

where the two spatiotemporal arrays Y –that contain the concentrations of every gene g, in every cell c at every time step t– are compared for every gene (G set to 3), every time step (T set to 1000) but for only 3 cells (C set to 3). Indeed, in order to control for differences in the position and size of the stripe, we only compare three cells for each space-time plot –at the center of each of the low-high-low regions of the stripe pattern.

This calculation is repeated for every pair-wise combination between the genotype to assign $geno_a$ and each 'core genotype'of the mechanism:

$$D_{geno_a-mech} = \frac{\sum_{l=1}^{N_{core}} D_{geno_a-geno_l}}{N_{core}} \tag{2.4}$$

where N_{core} is the total number of 'core genotypes'of the mechanism.

Following this method, we obtained six mechanism-specific neutral spaces that represent six disconnected regions of genotype space each making use of a particular dynamics. The size of the distinct neutral spaces can be measured in various manners: according to the number of genotypes, the number of topologies, or the maximal *Hamming Distance* between any two topologies in the dynamical regions (Fig. 2.1).

The Hamming Distance –HD– measures how dissimilar the architectures of two gene circuits are. Its maximal value within a three-node circuit space is 18. The distance between the topology matrices of two gene circuits W and W' can be measured as:

$$HammingDistance(W, W') = \sum_{i,j} \left| sgn(W_{ij}) - sgn(W'_{ij}) \right| \tag{2.5}$$

where i and j represent the position in those matrices. As the matrices are compared in every permutation, the Hamming Distance corresponds to the lowest value obtained. We notice that all mechanism-specific neutral spaces are large as they hold a maximal Hamming Distances from 6 to 9. This means that topologies within a given mechanism can be very dissimilar, suggesting that mechanism-specific neutral spaces percolate through large portions of genotype space.

2.6 Evolvability: defined and measured

The measure of a genotypes' evolvability is tightly linked to the type of mutations the genotype is subject to. Indeed, evolvability can only be assessed if we first define the mutation process we are interested in.

2.6.1 Choice of mutation process

The possible ways in which a genotype can be mutated are defined and constrained by the modeling approach. General considerations of the differences between continuous and discrete models were described earlier in section 0.1.2. Here, we will comment specifically on their impact on the type of mutations they allow.

Boolean circuits use a discretization of gene interactions that implies that only large/dramatic mutations are contemplated. Thus, boolean mutations consist on the complete removal or addition of entire gene interactions. Besides admitting this type of mutations, continuous circuits also allow for smoother mutations. These smoother mutations consist in 'fine-tuning'the values of single interactions and can thus occur without a change in the underlying topology. These mutations are more biologically realistic, as a change in the strength of a given interaction can represent, for example, a slight modification in the binding affinity of a transcription factor (e.g. a mutation in the sequence of a *cis*-regulatory region). Thus, in accordance to the continuous model in use, we thus chose to implement minimal mutations that affect only the strength of single gene interactions, in a process that takes into account real aspects of biological circuits.

Furthermore, as will be shown later, our goal is to map the possible phenotypic transitions within parameter space: which phenotypes directly touch each other and how do these transitions occur? For this purpose, the ability to 'fine-tune 'parameters is important. If instead we were to apply boolean mutations –drastic addition or removal of interactions– on our continuous genotypes, this mutations would easily 'jump over'relevant regions of parameter space.

2.6.2 The mutation process

We explore the genotype space around a given genotype in an exhaustive and systematic manner by mutating all its gene interactions one by one. A mutation affects only one gene interaction, whereas the rest of interactions remain unchanged. This way, a genotype with n gene interactions is subject to 2n independent mutations: each interaction has its original value (P_i) decreased (P_i^-) or increased (P_i^+) by a magnitude referred to as *mutational strength*:

$$P_i^+ = P_i(1 + mutational\,strength) \tag{2.6a}$$

$$P_i^- = P_i(1 - mutational\,strength) \tag{2.6b}$$

We will discuss later how this strength represents distinct mutation intensities but for now we assume it is set fixed at 50%. In the example of Figure 2.2A, a genotype with four interactions undergoes eight distinct mutations. In order to have a mechanism-based measure of evolvability, this same process is repeated for all genotypes that belong to a given mechanisms.

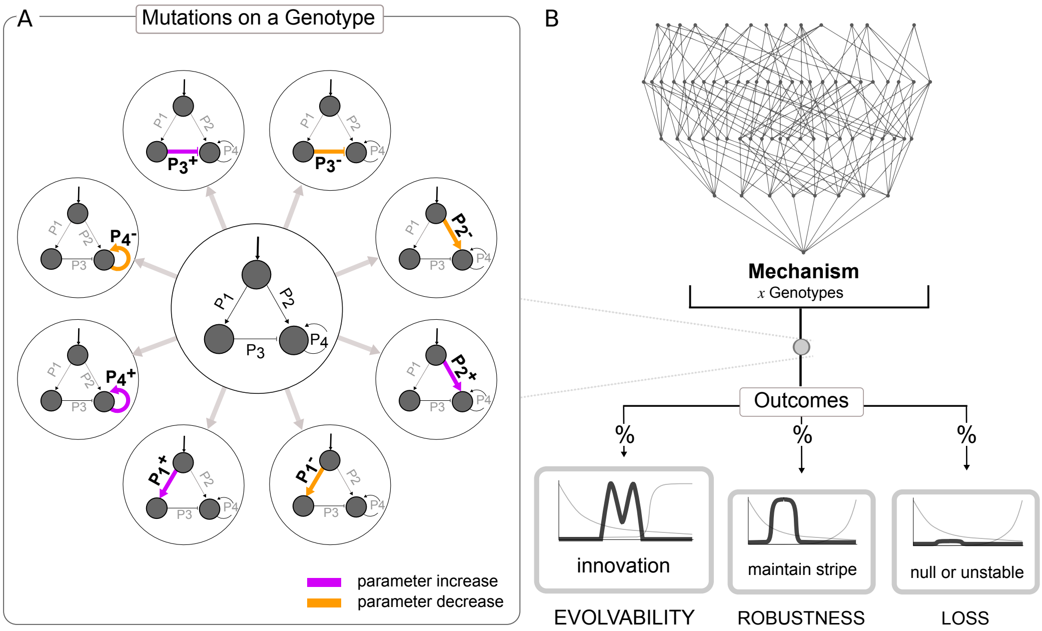

A
Mutations on a Genotype
P1
P2
P3
P4
P3+
P3-
P4-
P2-
P4+
P2+
P1+
P1-
parameter increase
parameter decrease
B
Mechanism
x Genotypes
Outcomes
%
%
%
innovation
maintain stripe
null or unstable
EVOLVABILITY
ROBUSTNESS
LOSS

Figure 2.2: **Measuring evolvability.** (A) Mutating a genotype consists of perturbing all its gene interactions. A mutations affects only one interaction while the rest of interactions remain unchanged. Each interaction has its value increased or decreased by the mutation strength. (B) At a fixed mutation strength, all genotypes belonging to a mechanism undergo the process in (A). Evolvability accounts for the proportion of mutants reaching novel phenotypes.

We measure evolvability as the proportion of genotypes within a mechanism that reach phenotypic innovation when subject to mutation. In other words, evolvability accounts for the likelihood of a genotype with specific dynamics to drift into a novel and stable phenotype. Respectively, robustness is measured as the proportion of genotypes that maintain the original stripe phenotype when undergoing mutation (Fig. 2.2B).

2.6.3 Functional versus non-functional phenotypes

For any given mutation, three phenotypic outcomes are possible: phenotypic innovation –novel pattern– phenotypic maintenance –stripe-pattern– or phenotypic loss –null or oscillatory phenotype (Fig. 2.2B). To assess these three possible outcomes, we established criteria that distinguish functional phenotypes – stripe pattern and novel patterns– from non-functional phenotypes -null or oscillatory phenotypes. Two criteria were chosen: the *pattern criterion* and the *stability criterion*.

The *pattern criterion*, considers that a gene profile conforms a pattern when the level of heterogeneity within its cells (left term of the inequality) is above a certain threshold:

$$\frac{\sum_{cells}\sqrt{(Y_{g,c,t_{eq}}-\sigma)^2}}{cells} > 10 \qquad with \qquad \sigma = \frac{\sum_{cells} Y_{g,c,t_{eq}}}{cells} \tag{2.7}$$

$Y_{g,c,t_{eq}}$ being the gene expression profile of gene g at the time when the pattern has reached equilibrium teq and σ being the average concentration of gene g throughout the tissue.

The *stability criterion* considers that a gene expression profile has reached equilibrium when it remains stable for more than 100 consecutive time steps. Furthermore, stable profiles need to reach equilibrium for each of the four different noise runs.

Functional phenotypes (i.e. stable patterns) fulfill both criteria. Instead, *non-functional phenotypes* dissatisfy at least one of the criterion. A kind of non-functional phenotype is the ‘flat phenotype’that does not fulfill the pattern criterion as it does not reach enough heterogeneity level in its profile. Another kind of non-functional phenotype is an oscillatory phenotype, which does not fulfill the stability criterion, as it does not reach equilibrium. Oscillatory phenotypes consist of all kinds of unstable solutions. Although oscillatory phenotypes are biologically relevant, for the purpose of the main study, we included oscillatory phenotypes as phenotypic loss. However, although it is not the main focus of this thesis, we will briefly discuss in section 3.1.4 the relation between oscillatory phenotypes and stable patterns.

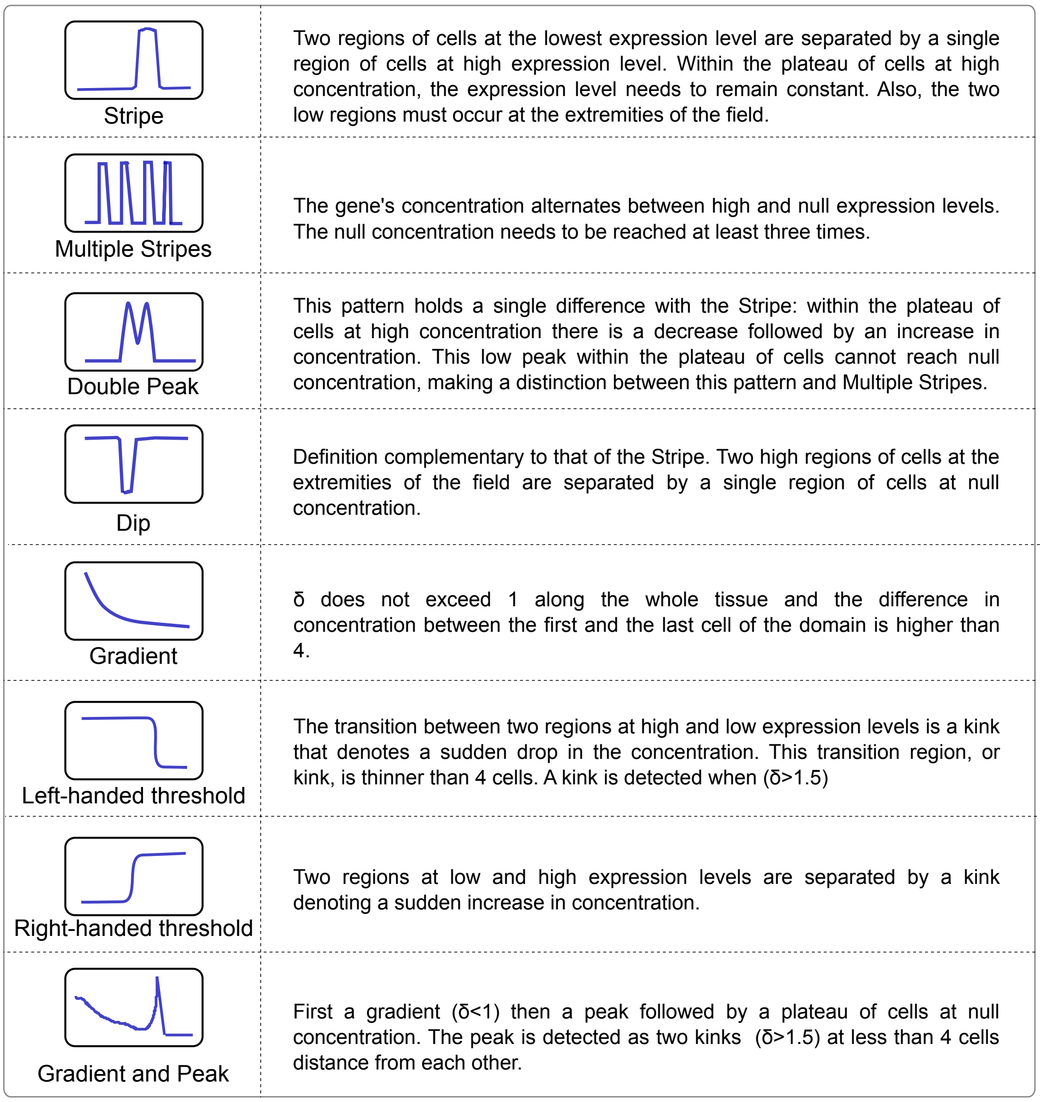
2.6.4 Classifying functional phenotypes into different patterns

Functional phenotypes can be classified into 8 categories: the stripe pattern and 7 novel patterns. The criteria to define each pattern were first defined by J.Cotterell and later refined by the authors of this thesis (Fig. 2.3). In order to characterize particular spatial features of some of the patterns we introduce a measure of change in concentration between two consecutive cells

$$\delta = \sqrt{(g_{i,j})^2 - (g_{i,j+1})^2} \tag{2.8}$$

where g_{ij} is the concentration of the gene of interest i in the cell at position j. This measure is relevant to identify gradients and sudden spatial changes in concentration –peaks and kinks.

The definitions in Figure 2.3 are intentionally loose. For example, whether we consider single or multiple stripes, stripes can be of any width and at any position in the spatial domain. Hence, although the levels of gene expression in individual cells can vary widely between two phenotypes classified as *Stripes* (position and width of the stripe), the above-mentioned behavioural rules are respected. These definitions are also mutually exclusive, in the sense that a pattern cannot belong to two categories at the same time. Following these criteria, we were able to classify 93% of all novel phenotypes into one of the seven novel categories described.



Stripe	Two regions of cells at the lowest expression level are separated by a single region of cells at high expression level. Within the plateau of cells at high concentration, the expression level needs to remain constant. Also, the two low regions must occur at the extremities of the field.
Multiple Stripes	The gene's concentration alternates between high and null expression levels. The null concentration needs to be reached at least three times.
Double Peak	This pattern holds a single difference with the Stripe: within the plateau of cells at high concentration there is a decrease followed by an increase in concentration. This low peak within the plateau of cells cannot reach null concentration, making a distinction between this pattern and Multiple Stripes.
Dip	Definition complementary to that of the Stripe. Two high regions of cells at the extremities of the field are separated by a single region of cells at null concentration.
Gradient	δ does not exceed 1 along the whole tissue and the difference in concentration between the first and the last cell of the domain is higher than 4.
Left-handed threshold	The transition between two regions at high and low expression levels is a kink that denotes a sudden drop in the concentration. This transition region, or kink, is thinner than 4 cells. A kink is detected when (δ>1.5)
Right-handed threshold	Two regions at low and high expression levels are separated by a kink denoting a sudden increase in concentration.
Gradient and Peak	First a gradient (δ<1) then a peak followed by a plateau of cells at null concentration. The peak is detected as two kinks (δ>1.5) at less than 4 cells distance from each other.

Figure 2.3: **The stripe and seven novel patterns.** Criteria for pattern classification.

Chapter 3

RESULTS

3.1 Evolutionary potential is mechanism-dependent

3.1.1 Evolutionary potential at a fixed mutation strength

First, we separately monitor the evolvability of the six distinct mechanisms, i.e. each regulatory parameter was increased or decreased by 50% of its ‘wild-type ’value (50% mutation strength). Although the measure of evolvability depends on the mutation strength considered, we believe that performing the analysis at a given mutation strength will help us get a first intuition on the results.

With this simple initial analysis, we found that circuits using distinct mechanisms for stripe formation differ substantially in the likelihood of reaching novel phenotypes. For each mechanism, Figure 3.1 shows the range of novel phenotypes occurring at a frequency above 1/10,000. In other words, which novel phenotypes are most easily accessible from the original stripe phenotype and how many mutations, out of 10,000, allow access to that phenotype. Here we chose to schematically show novel phenotypes as patterns in a *Drosophila* embryo.

It is clear that some of the mechanisms have a much higher chance of evolving a novel phenotype than others. For example, mutations to *Classical* (top) have 10 times higher chance of producing a novel phenotype than mutations to *Overlapping Domains* (bottom). Additionally, the actual novel phenotypes that can be easily reached are different for each of the mechanisms. While all mechanisms are able to access *Left* and *Right-handed thresholds* easily, the access to *Gradient*, *Double Peak*, and *Multiple Stripes* is mechanism-dependent. Indeed, *Frozen Oscillator* is the only mechanism from with *Multiple Stripes* is easily accessed while *Incoherent feed-forward* and *Mutual Inhibition* are the two mechanisms from which the *Cradient* is most likely to be reached (Fig. 3.1).

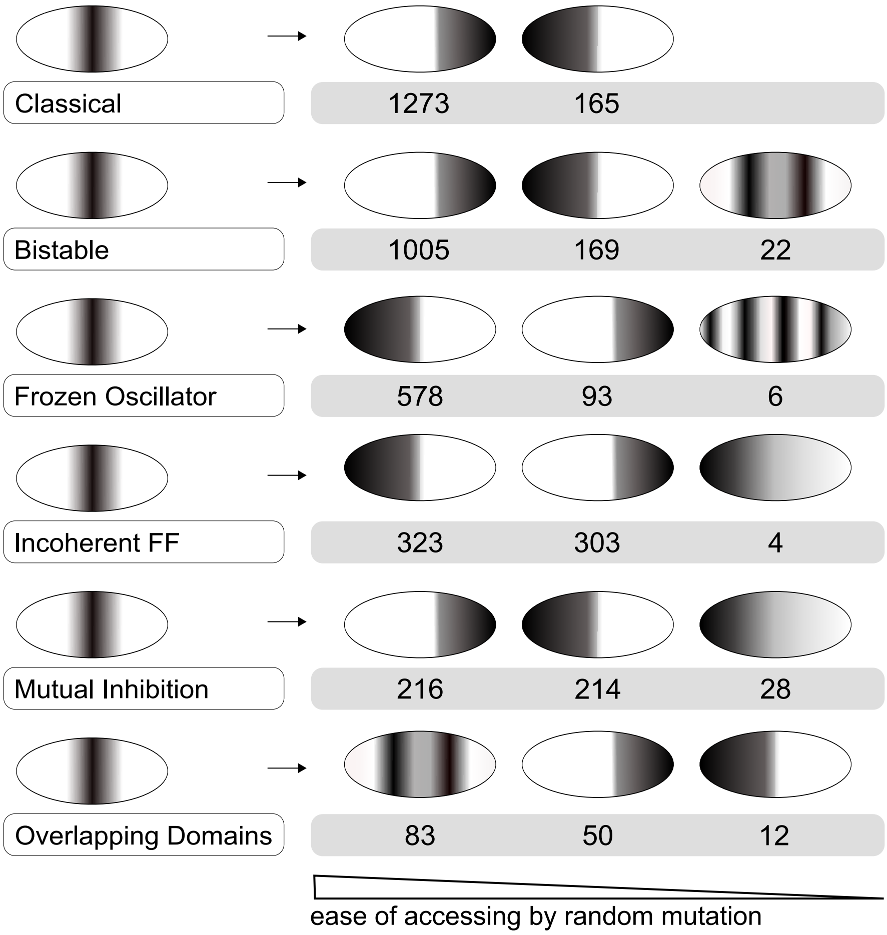

Classical
1273
165
Bistable
1005
169
22
Frozen Oscillator
578
93
6
Incoherent FF
323
303
4
Mutual Inhibition
216
214
28
Overlapping Domains
83
50
12
ease of accessing by random mutation

Figure 3.1: **Differences in evolvability: schematics of novel phenotypes.** Evolutionary potential is shown as the number of mutations out of 10,000 that permit access to distinct novel phenotypes. Values correspond to 50% mutation strength. Phenotypes accessible with a frequency below 1/10,000 (0.01% chance to be accessed through mutation) are not considered.

Novel patterns can be classified as major or minor according to how common they are, i.e. their accessibility by random mutation (see classification Fig. 3.2A). We remark that major patterns –*Left* and *Right-handed* thresholds– are highly accessible from all mechanisms, whereas minor patterns are more restricted to specific mechanisms (also observed in later results Fig. 3.2B). Understanding the stripe-forming process of each mechanism can offer an intuitive explanation of these differences in accessibility. Indeed, the stripe phenotype is created through the establishment of a left and a right expression boundary, being the superposition of these two patterns. Any mutation that disrupts the establishment of one boundary leads to the complementary phenotype. This explains why the major patterns of right and left-handed thresholds are easily accessible, as they constitute a functional loss of one of the two modules. By contrast, a single mutation that simultaneously disrupts the establishment of both boundaries, leading for example to the gradient phenotype, is less likely to occur.

3.1.2 Evolvability profiles obtained as mutation strength is varied

Next, we explored how the evolvability of each mechanism changes as the mutation strength is varied. For that, we repeat the previous analysis as the mutation strength ranges from $\pm 1\%$ to $\pm 99\%$. Figure 3.2 shows how evolvability varies as the intensity of the mutational environment increases. For each mechanism, the black line depicts the whole evolvability profile, i.e. the likelihood of a genotype to drift into any novel phenotype. Evolvability is then subdivided into the seven colour-coded novel categories (Fig. 3.2A). The specific fractions of novel phenotypes appear in stacked area charts.

The six mechanisms clearly have different evolvability profiles. For all mechanisms, the percentage of novel phenotypes initially increases as the mutational strength is raised. However, beyond a mutational strength of 20%, two mechanisms start to decrease again, *Incoherent feed-forward* and *Mutual Inhibition*, whereas other mechanisms continue a slower increase until saturating around 50% (*Classical* and *Bistable*). We suggest that these differences can be considered a reflection of the geometry of local phenotype space, i.e. the distance, shape and extension of novel neutral spaces adjacent to the original mechanism.

To explore this issue, we use a toy model that simplifies the high dimensional geometries of neutral spaces (holding genotypes with up to 9 gene-interactions and thus 9 distinct dimensions in genotype space) into simple 2D cartoons (Fig. 3.3). In this simplified 2D space, evolvability at a particular mutation strength would correspond to the intersection between the perturbed –enlarged– mechanism and the novel neutral space. Intuitively, as mutation strength increases, we further explore the adjacent novel neutral space. In this 2D hypothetical space, we propose four distinct geometries for adjacent neutral spaces. They consist on four possible simple shapes represented by cartoons. As illustrated in Figure 3.3, the four distinct geometries can cause (a) a continuous increase in evolvability (b) an increase followed by a stabilization (c) an increase followed by a decrease (d) an increase then a sudden drop. The difference between a decrease or a drop depends on how smooth or sudden are changes in the geometry of the novel space.

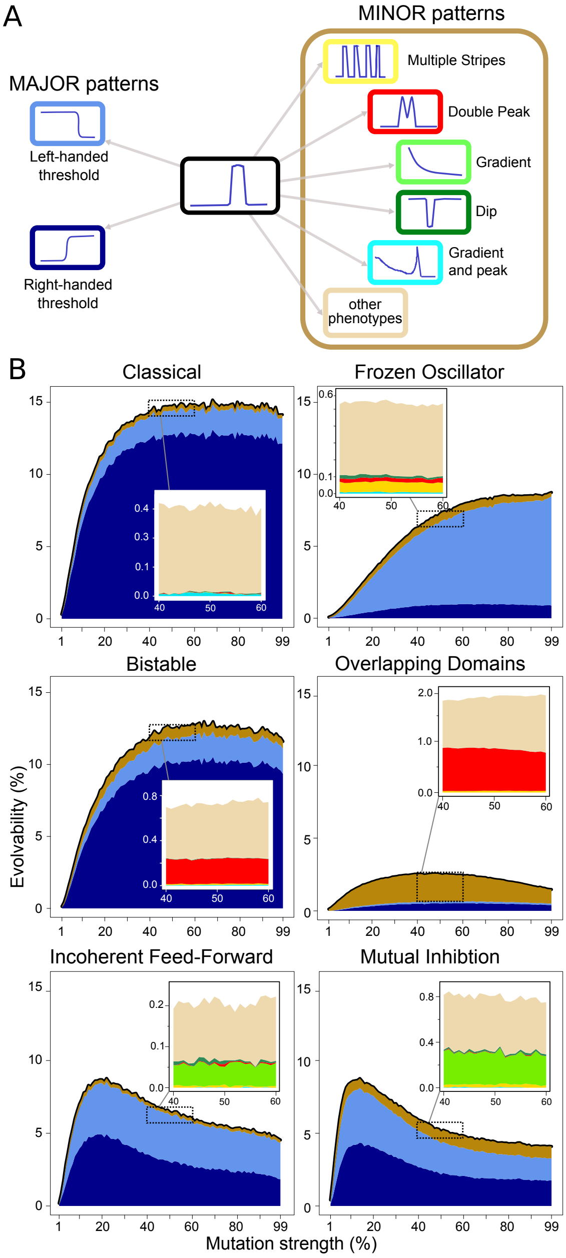
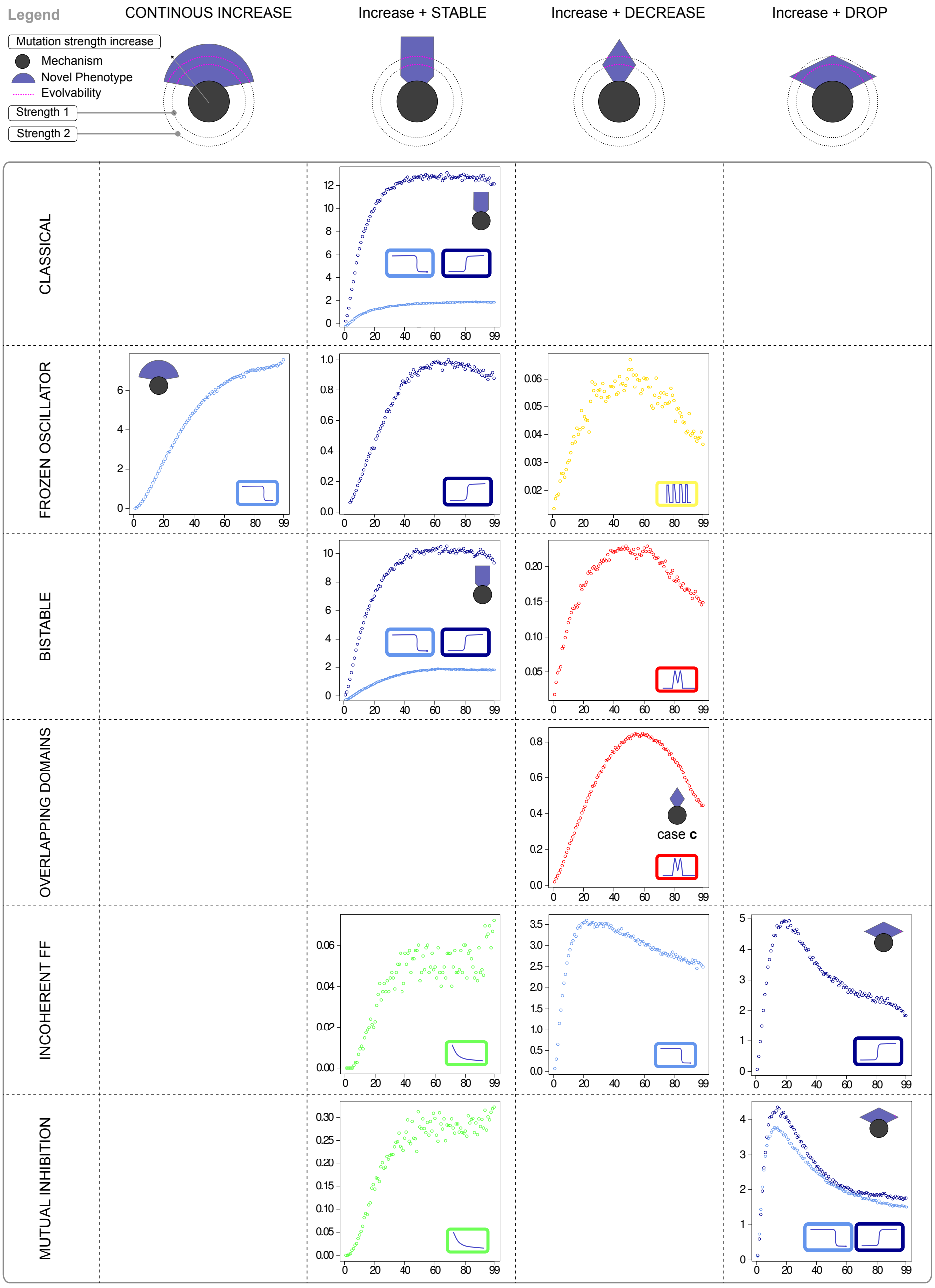


Figure 3.2: **Dynamical mechanism influences evolvability** (A) Phenotype space is segmented into eight phenotypes: the stripe expression and seven novel gene expression patterns. They cover 93% of observed phenotypes accessible from the stripe phenotype. The remaining unclassified 10% largely represent rare hybrids of the seven groups mentioned (e.g., gradient plus threshold). (B) Evolvability as a function of mutation strength. Black line depicts whole evolvability profile, whereas the specific fractions of novel phenotypes appear in stacked area charts (zooms of the minor patterns in upper windows).



Plots of evolvability per novel phenotype (%) as a function of mutation strength

Figure 3.3: **Impact of the geometry of adjacent neutral spaces on evolvability.** (A) Four proposed distinct geometries of novel neutral spaces have distinct impacts on evolvability. (B) Evolvability subdivided into access to particular novel phenotypes. Only phenotypes that have more than a 0.01% chance be accessed (1 mutation out of 10,000) are included.

For each mechanism, we separately measure the probability to access each novel phenotype as mutation strength increases (Fig. 3.3). According to the profiles obtained, we can allocate distinct geometries to each novel neutral space. For example, we observe an ‘increase and drop’in the evolvability profile of *Right-handed threshold* pattern for *Incoherent feed-forward* mechanism. Particularly, this profile is the one that most influences the global evolvability profile –all patterns considered– observed for *Incoherent feed-forward* mechanism (Fig. 3.2B). Likewise, the slow increase -from 50% mutation strength on- in the global evolvability profile of *Frozen Oscillator* (Fig. 3.2B) can be explained by the individual *Left-handed threshold* profile. In this manner, certain adjacent geometries –marked with a small cartoon– have a greater impact on the global evolvability profiles of Figure 3.2B. Although this type of toy model has strong limitations, it can help us acquire some intuition on the impact of the geometry of adjacent neutral spaces on evolvability.

3.1.3 Explaining the differences in evolvability by mechanism

Here we aim at understanding why minor patterns are easier to access from specific mechanisms. For example, the modular design of *Frozen Oscillator* explains why it is the mechanism from which the *Multiple stripes* pattern is most readily accessible. Indeed, *Frozen Oscillator* consists of two modules: the negative feedback loop formed by the green and blue genes (Fig. 0.13) forms an oscillator that is represented by the red gene acting to freeze these oscillations. A stripe corresponds to a single frozen oscillation. Hence, mutations modifying the frequency of oscillations can produce more than a single oscillation and therefore lead to a periodic patterns such as multiple stripes.

Likewise, *Incoherent Feed-Forward* and *Mutual Inhibition* patterns easily access the *Gradient* phenotype. This means that single mutations on those mechanisms are able to completely disrupt the establishment of both boundaries of the stripe, suggesting that the stripe is formed in an integrated and non-modular way.

In brief, in these first part of the results we have observed that, consistent with our hypothesis on a mechanism-based view, the likelihood of accessing distinct innovations (i.e. the evolvability) varies for each mechanism. Furthermore, we provide intuitive understanding of why particular phenotypic innovations are more probable from certain mechanisms.

3.1.4 Oscillatory phenotypes

The relation between oscillatory phenotypes and stable patterns is an interesting question, as the oscillatory behaviour is of interest in several developmental stages, such as somitogenesis, which produces a spatial pattern as a result of a temporal oscillation. Indeed, later on, some of the examples of phenotypic neighbourhoods (Fig. 3.9) will show that it is possible to transition in and out of oscillatory regions.

We explore the capacity of distinct mechanisms to access non-stable patterns (Fig. 3.4). For that, we classify the phenotypic outcomes into four classes - innovation, phenotypic maintenance, null phenotype and oscillatory phenotypes- that sum 100% of the possible outcomes. Again, we consider the whole range of mutation strengths and monitor how each of the four phenotypic outcomes evolves as the mutation strength increases. We observe that mechanisms not only differ in their evolvability for stable patterns but also in their capacity to achieve non-stable patterns. Of these, *Overlapping Domains* has the highest capacity to achieve oscillatory phenotypes.

3.1.5 Influence of the model formalism on the results

In order to strengthen the results obtained, we aim at showing that the former conclusions do not depend on the particularities of our model. Indeed, an important question in any theoretical study is to what extent the main conclusions depend on the assumptions of the model. Next we show that the results obtained are consistent when we vary some of the assumptions of the model, such as the regulatory function or the size of the field of cells.

For each change on the model, we re-built a complete atlas (Fig. 3.5B-D) and obtained evolvability profiles for each mechanism (Fig. 3.6B-D) to further compare them to the original results (Fig. 3.5A-Fig. 3.6A). Precisely, these changes concern:

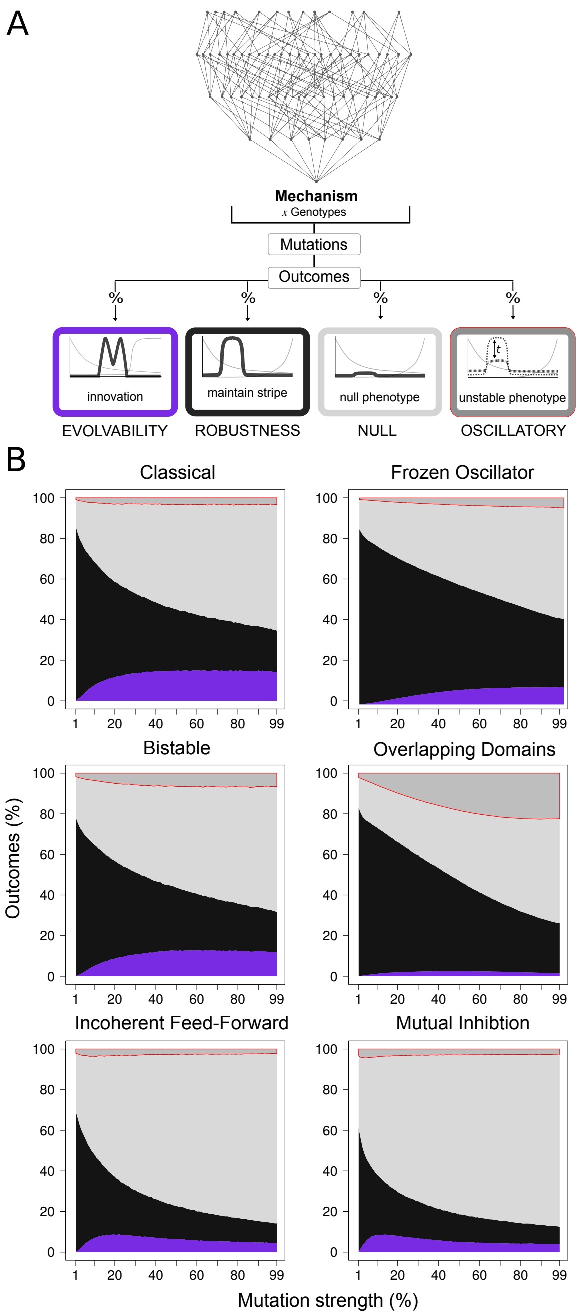


Figure 3.4: **Dynamical mechanisms are distinct in their capacity to achieve oscillatory phenotypes under mutation.** (A) Four phenotypic categories sum 100% of all possible outcomes. Examples of transitions from a stripe to temporal oscillations can be found in Figures 3.8 and 3.9. (B) Mechanism Overlapping Domains has a higher capacity to achieve oscillatory phenotypes.

Influence of the size of the field of cells

We change the number of cells in the tissue from 32 to 64 cells. The complexity atlas and the evolvability profiles obtained under this assumption (Fig. 3.5B-Fig. 3.6B) show only minor changes with respect to the original result.

Influence of diffusion

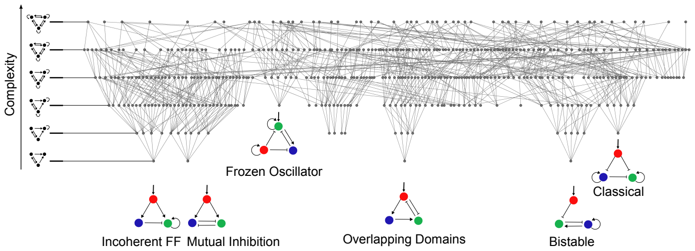
We switched diffusion off for all genes' products. Only one of the mechanisms is significantly affected: the core circuit of *Overlapping Domains* is not able to form a stripe. Nevertheless, additional links onto the core circuit can compensate for the lack of diffusion, as seen in the complexity atlas from Figure 3.5C. Genotypes of higher complexity than the 4-link level (1,2,3 in Figure 3.5C) are a subset of the original genotypes working under the dynamics of *Overlapping Domains* mechanism. However, as the 4-link core circuit does not form a stripe in the absence of diffusion, we consider that *Overlapping Domains* does not operate as a valid mechanism when diffusion is switched off. Therefore, we include the results for that mechanism (Fig. 3.6C) under a shaded mask, and we will not use them for comparison. Among the other mechanisms, we remark that *Frozen Oscillator* and *Bistable* need diffusion to access minor patterns *Multiple Stripes* and *Double peak*.

Influence of the regulatory function

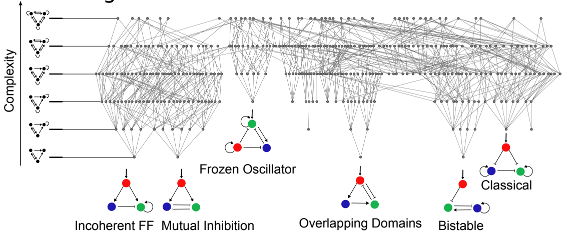
We employed the Sigmoidal regulatory function instead of the Michaelis-Menten one. The total number of mechanisms is now 5 (Fig. 3.5D). While *Mutual Inhibition* and *Incoherent feed-forward* fuse into Incoherent feed-forward type 1, *Overlapping Domains* mechanism is able to create a stripe with one regulatory interaction less, becoming Incoherent feed-forward 3 (under the nomenclature of U.Alon [Alon, 2006, **?**]). Hence, we remark that the Sigmoidal function facilitates the development of the stripe phenotype using circuits of lower complexity. Additionally, the *Incoherent feed-forward* mechanism is not able to achieve the *Gradient* phenotype under mutation (Fig. 3.6D). This could be a direct effect of the higher steepness of the Sigmoidal regulatory function compared to the Michaelis-Menten one, translating into the difficulty of reproducing the smoothness of the input gradient.

In a nutshell, under distinct assumptions, the key original results are conserved: the evolvability of a circuit varies according to its underlying mechanism, as distinct mechanisms show specific likelihoods to access distinct novel phenotypes. Moreover, we remark the homogeneity of the evolvability profiles in spite of the variety of theses assumptions.

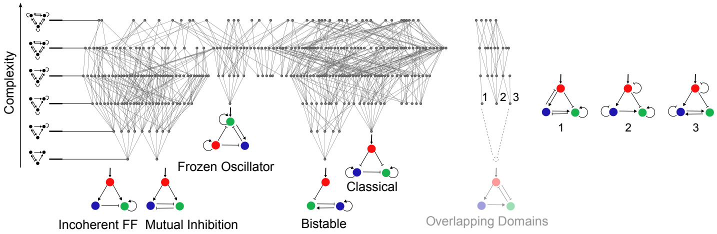
A Original model formalism


Complexity
Frozen Oscillator
Classical
Incoherent FF
Mutual Inhibition
Overlapping Domains
Bistable

B Change in tissue size


Complexity
Frozen Oscillator
Classical
Incoherent FF
Mutual Inhibition
Overlapping Domains
Bistable

C Absence of diffusion


Complexity
1
2
3
1
2
3
Frozen Oscillator
Classical
Incoherent FF
Mutual Inhibition
Bistable
Overlapping Domains

D Change in regulatory function

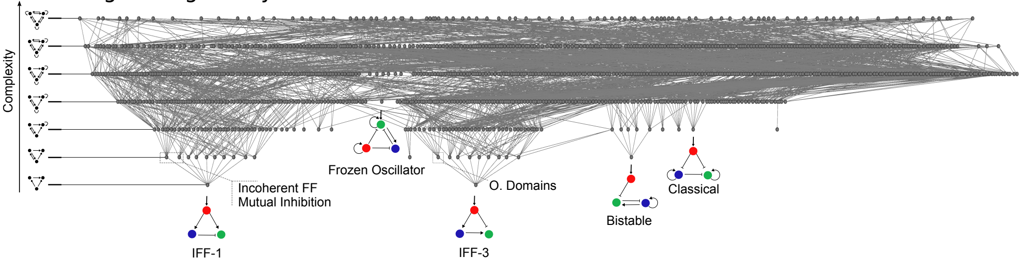

Complexity
Frozen Oscillator
Incoherent FF
Mutual Inhibition
O. Domains
Classical
Bistable
IFF-1
IFF-3

Figure 3.5: **Complexity atlases drawn for different model formalisms.** (A) Original model formalism. (B) Larger tissue (64 instead of 32 cells). (C) Absence of diffusion. (D) Use of Sigmoidal regulatory function

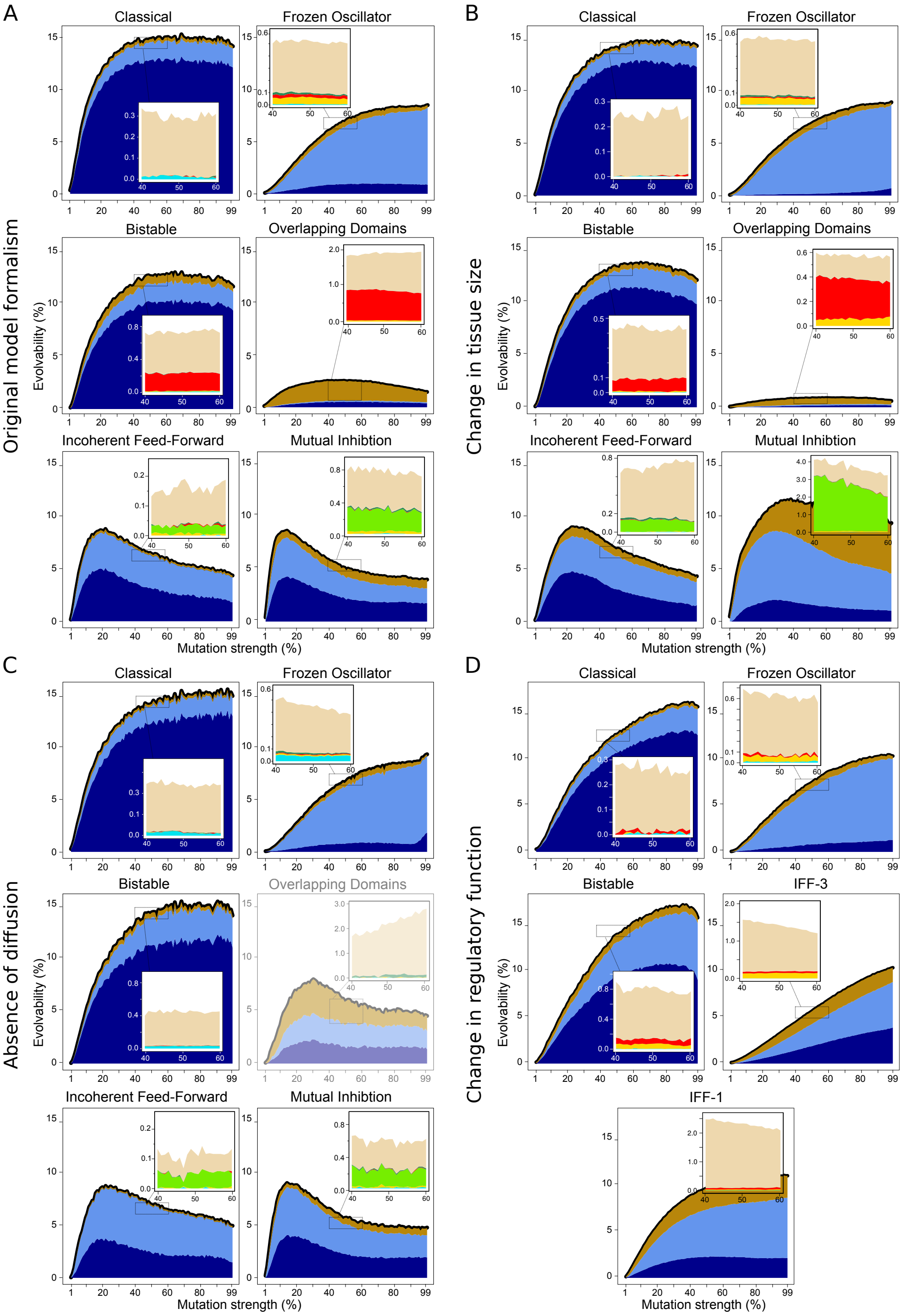


Figure 3.6: **Evolvability measured under distinct particularities of the model.** Labelling corresponds to that of Fig. 3.5. Results sampling 500.000 parameter sets instead of the 10^6 sampled in Fig. 3.2.

3.2 Exploring phenotypic transitions

3.2.1 Depicting the phenotypic neighbourhood from a given genotype: phenotype-transition diagrams

To illustrate how the original phenotype is transformed into a novel phenotype, we use *phenotype-transition diagrams* to visualize transitions between patterns. The objective of these phenotype-transitions diagrams is to exhaustively depict the phenotypic neighbourhood from a given genotype and thus explore the nature of transitions between phenotypes. Hence, for this part of the study, we switch from a mechanisms perspective to a single-genotype perspective.

Because individual genotypes reside in a high-dimensional genotype space (i.e the number of gene interactions determines the dimension), we construct all possible 2D sections through this genotype space. A section is created by choosing two gene interactions and allowing their value to vary while the rest of interactions are fixed. For each couple of interaction values, we assess the resulting phenotype and colour-code the corresponding 'pixel' accordingly. Because the value of interactions is varied at very small intervals, these diagrams resemble high-resolution images made of coloured pixels. As such, the phenotypic neighbourhood from a given genotype is composed of neutral spaces depicted as continuous coloured zones.

As an illustrative example, we include all phenotype-transition diagrams for an original genotype producing the *Stripe* phenotype with underlying *Bistable* mechanism (Fig. 3.7). The original genotype (pink box) is projected into six distinct sections that help us visualize the surrounding phenotypic neighbourhood.

First, by predicting the phenotypic outcome of a circuit subject to perturbation, these diagrams point out which specific gene interaction needs to be mutated, and with which intensity, in order for the circuit to produce a given novel phenotype.

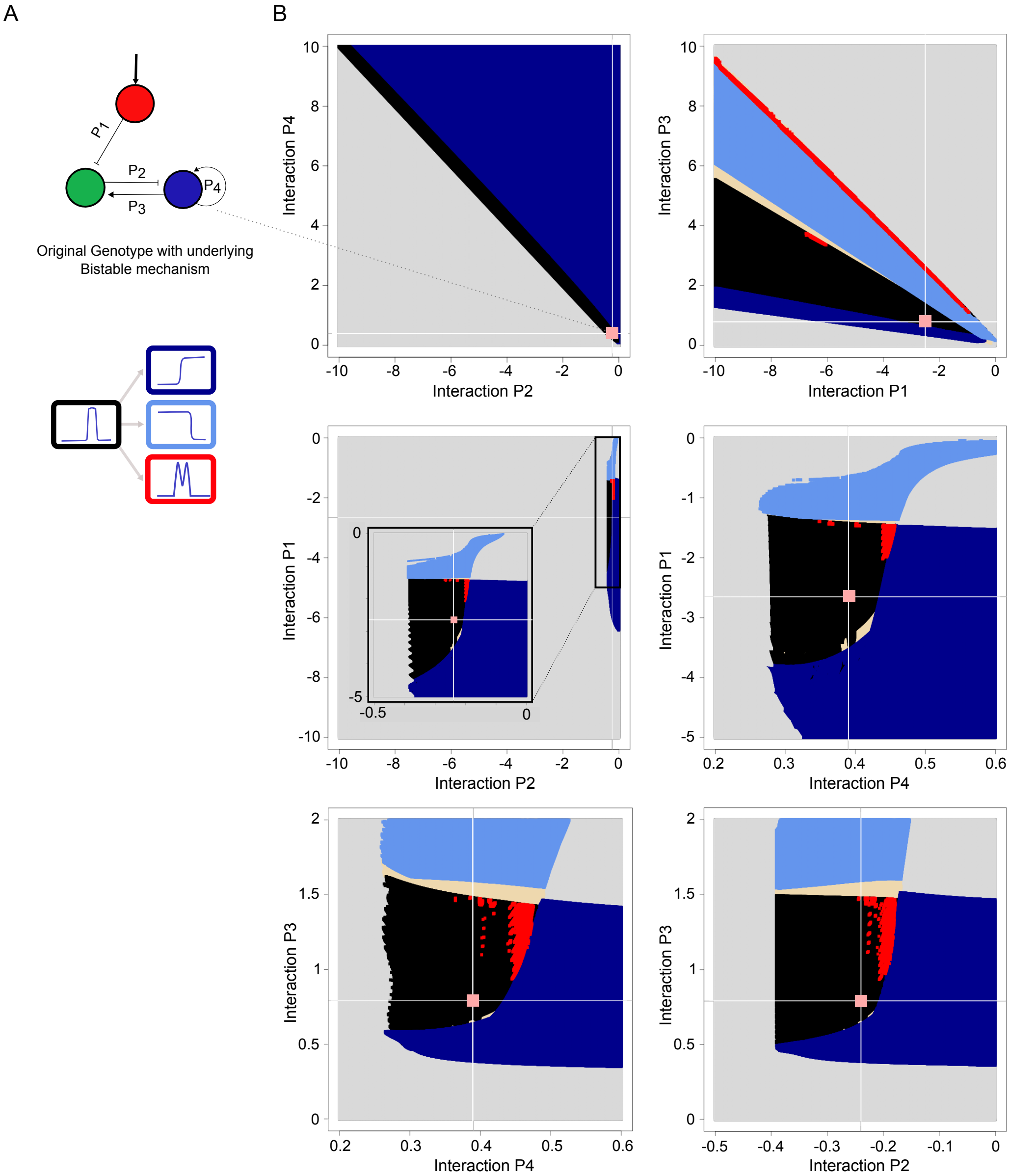


Figure 3.7: **Exhaustive exploration of phenotype space from a genotype.** (A) Original genotype producing the stripe phenotype with underlying Bistable mechanism and interaction strengths P_1-P_4={-2.65; -0.24; 0.79; 0.39}. (B) Phenotype-transition diagrams. The phenotypic neighbourhood comprises phenotypes Right-handed threshold, Left-handed threshold and Double Peak. Two combinations of gene-gene interactions (P_2-P_4 and P_1-P_3) hold a linear relationship. This relationship involves similarities between the rest of phenotype-transition diagrams.

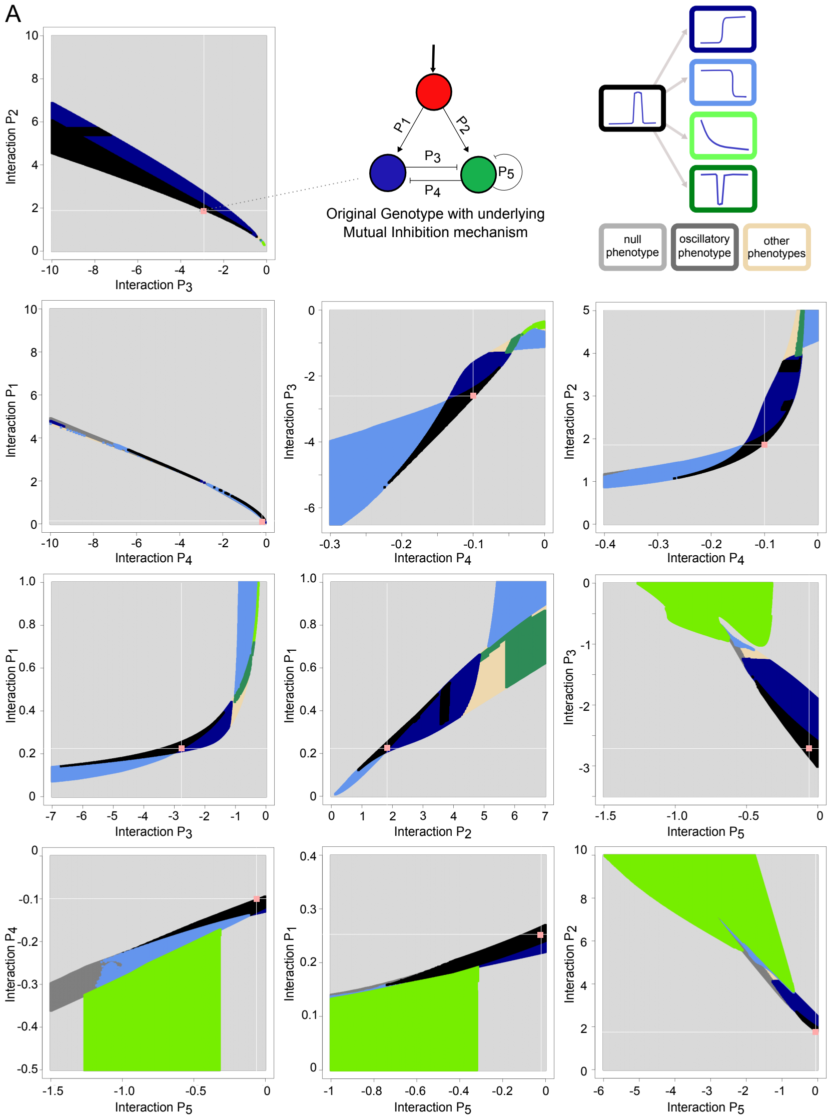


Figure 3.8: **Heterogeneous phenotypic neighbourhoods within a single mechanism.** Phenotype-transition diagrams from an original genotype with underlying Mutual Inhibition mechanism and interaction strengths P_1-P_5={0.25; 1.88; -2.7; -0.1; -0.07}.

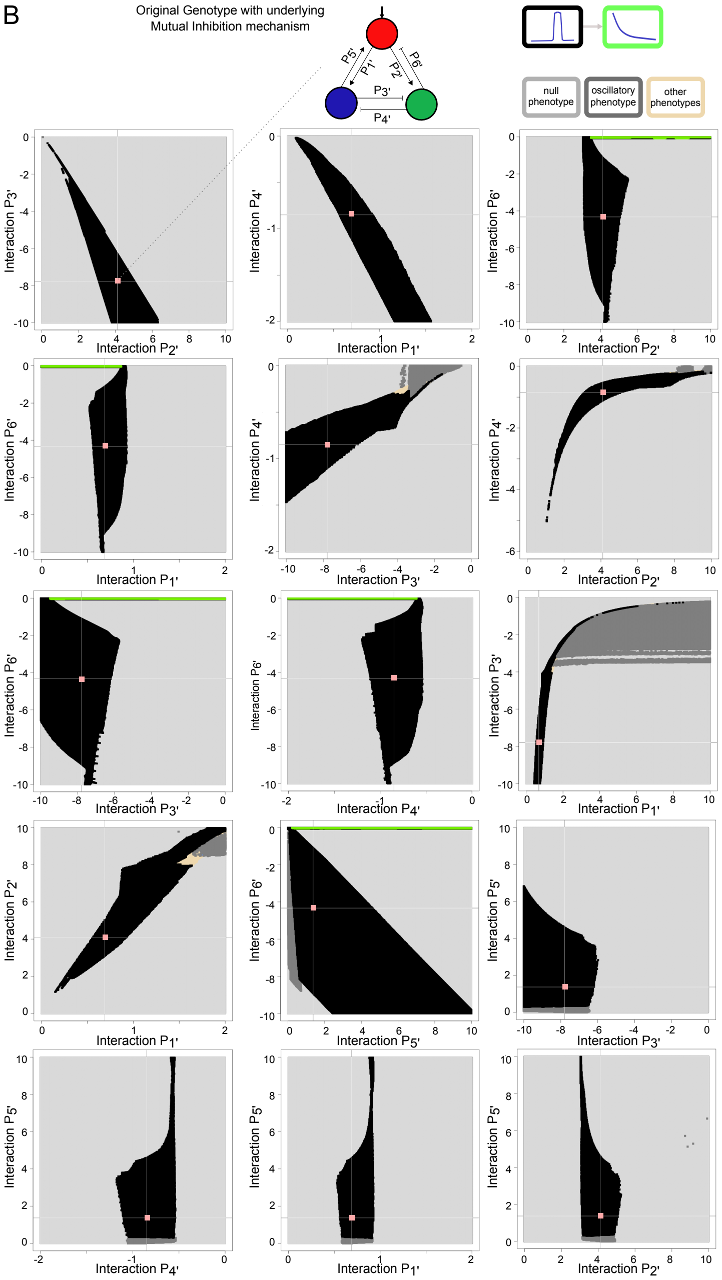


Figure 3.9: **Heterogeneous phenotypic neighbourhoods within a single mechanism.** Phenotype-transition diagrams from an original genotype producing the stripe phenotype with underlying Mutual Inhibition mechanism and interaction strengths P'_1-P'_6={0.7; 4.1; -7.8; -0.85; 1.37;-4.33}

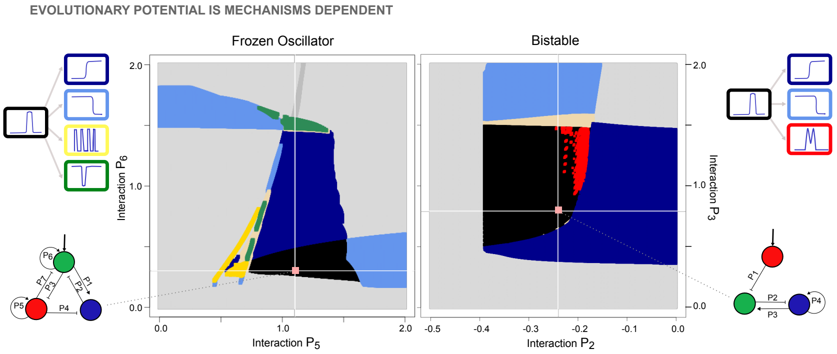


Figure 3.10: **Characteristic phenotypic neighbourhoods of two mechanisms.** Phenotype-transition diagrams from two original genotypes with underlying Frozen Oscillator and Bistable mechanisms and interaction strengths P_1-P_7={4; -1; -0.08; -8; 1.1; 0.3; -0.25} and P_1-P_4={-2.65; -0.24;0.79; 0.39} respectively. Each point in the section corresponds to a particular genotype that is assigned a colour corresponding to its resulting phenotype. Understanding particular concepts that are recurrent in studies of robustness and evolvability is intuitive using this type of visualization. The parameter robustness of an interaction is visually inferred and corresponds to the width of the black zone along the corresponding axis, for example [0.7-1.5] range for interaction P_3 of the genotype with Bistable mechanism. In addition, the distance along the axis from the original genotype to the nearest novel phenotype provides a measure of the mutation strength needed to switch phenotype.

Second, these diagrams make it clear and understandable why for certain gene circuits, many parameter changes lead to novel phenotypes (Fig. 3.8), whereas for others, very few parameter changes are innovative (Fig. 3.9). Indeed, even within a single mechanism, phenotypic neighbourhoods can be highly heterogeneous. We studied two genotypes with underlying *Mutual Inhibition* mechanism, which topologies are at a Hamming Distance 3 (i.e. three interactions away) (Fig. 3.8-3.9). These genotypes present very distinct neighbourhoods: while the first genotype can access four distinct patterns (Fig. 3.8), the second can only transition to a single novel phenotype (Fig. 3.9). This is an expected result: as we pointed out in the introduction, distances in genotype space lead to the divergence in composition of phenotypic neighbourhoods. We expect this heterogeneity to be a feature of all mechanisms, as their neutral spaces are large -large Hamming Distances (Fig. 2.1)- and can thus extend throughout genotype space, providing mutational access to a diversity of novel phenotypes from different genotypes.

If we could simultaneously visualize all phenotypic neighbourhoods around a given mechanism, we would see that the proportions of coloured zones representative of novel neutral spaces would correspond to the distinct proportions observed within the evolvability profiles shown in Figure 3.2. That is why, as a case study, Figure 3.10 shows two characteristic sections from genotypes belonging to *Frozen Oscillator* and *Bistable mechanisms*.

3.2.2 Phenotypic transitions are continuous

As we already mentioned, phenotype-transition diagrams point out how given mutations lead to specific phenotypes. However, how does the drifting between phenotypes occur? We observe that a transition from a phenotype to another does not require a transient loss of phenotype. In Figure 3.11, we illustrate the continuity of phenotypic transitions by smoothly changing particular interactions in a genotype. Some of the transitions reveal the role of particular gene interactions in pattern formation. For example, the strength of interaction P_6 controls the shift of the stripe's right boundary (Fig. 3.11, frames 2-7), whereas the strength of both P_1 and P_5 interactions controls the shift of the stripe's left boundary (Fig. 3.11). We observe that P_5 allows a switch from the *Stripe* phenotype to *Left-handed threshold* and *Multiple stripes* patterns, whereas P_6 allows access to the *Right-handed threshold* phenotype and then to the *Dip* phenotype. These particular interactions, which can be tuned to access more than one novel phenotype, behave as *sensitive interactions*.

Although a direct equivalent of these conclusions cannot (as yet) be supported from studies of real species, we do mention a related example: a well-studied stripe domain (*giant*) can indeed shift, through mutation, from a *Stripe* to a *Right-handed threshold* (Fig. 3.12). During early embryogenesis, *Drosophila's* gap gene *giant* (*gt*) protein is expressed in two broad anterior and posterior regions. These two domains of *gt* expression can be treated as two separate entities as they are under the separate control of different maternal morphogens: while Bicoid defines the anterior *gt* domain, Caudal activates the posterior *gt* domain. We thus focus on the generation of the *gt* posterior domain under control of maternal morphogen Caudal. In a series of gap gene mutants, Kraut and Levine [Kraut and Levine, 1991] observed how the phenotype could shift from a *Stripe* to a *Right-handed threshold*. Particularly, the position of the left boundary of the stripe slides to distinct positions in different mutants, until it completely disappears in the Torso mutant (Fig. 3.12-[5]). It is important to keep in mind that this type of mutations are distinct from our model formalism in which mutations are done in particular gene-gene interactions at fixed mutations strengths. Indeed, this series of phenotypes is not from an allelic series of a given mutation, neither do we know the regulatory mechanism of *gt* in detail.

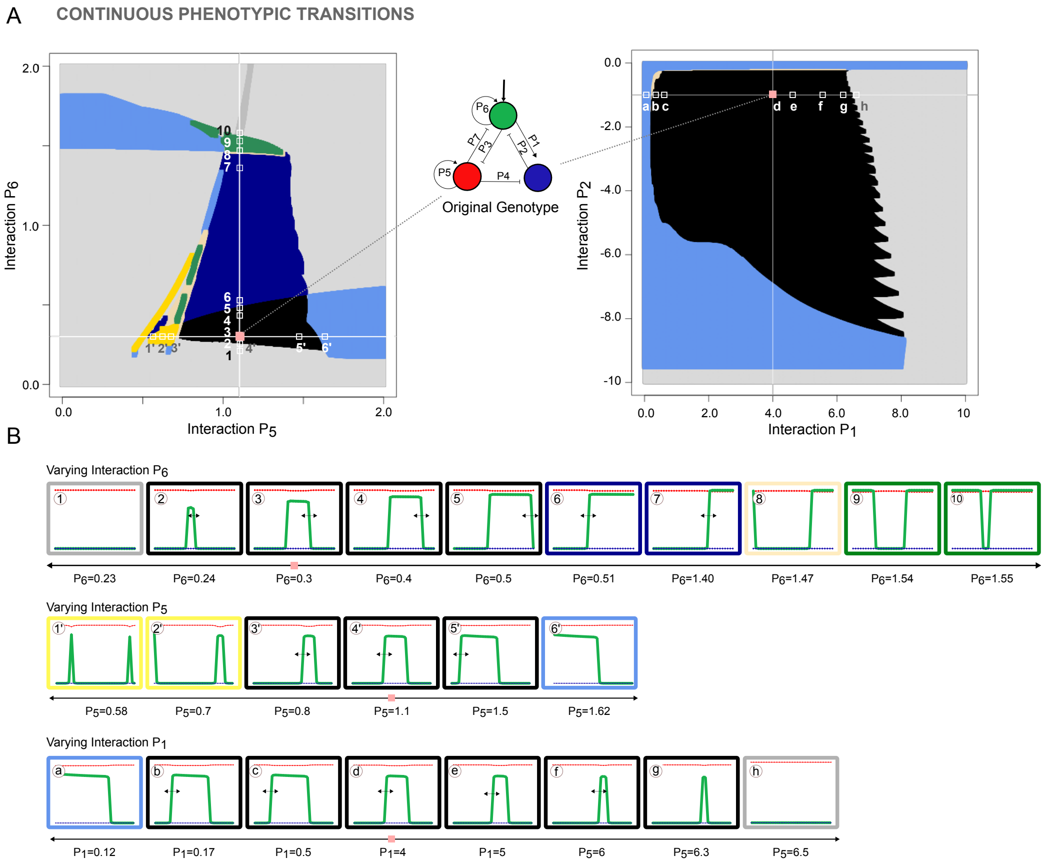


Figure 3.11: **Continuous phenotypic transitions.** (A) Various phenotypic-diagrams from the Frozen Oscillator genotype of figure 3.10 (B) Evaluating the resulting phenotype while varying interactions 6, 1 and 5 shows transitions between phenotypes are continuous. Particularly, frames 2 to 7 suggest interaction 6 governs the position of the right boundary of the stripe (black arrows). Frames 3' to 5' and b to f suggest both interactions 1 and 5 control the position of the stripe's left boundary (black arrows)

3.2.3 Phenotypic hubs

Phenotype-transition diagrams reveal the richness of phenotypes around specific regions of genotype space –such as the previously shown neighbourhood in Figure 3.8. Indeed, certain genotypes can access a high diversity of innovations when different gene interactions are mutated (Fig. 3.13).

Here we propose an alternative measure of evolvability: instead of measuring the likelihood to transition into novel phenotypes like we have done in Figure 3.2, we now chose to measure evolvability as the number of accessible distinct phenotypes per genotype: *phenotypic diversity*. We observe that its spectrum is quite

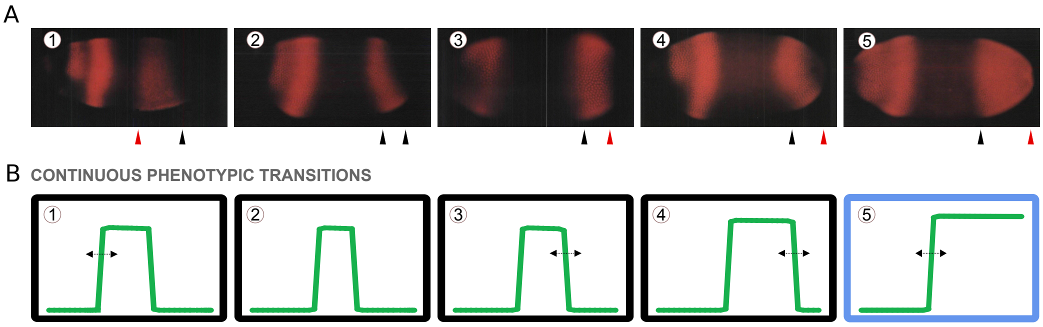


Figure 3.12: **Gene expression patterns of gap gene *giant* in several *Drosophila* mutants.** (A) Gene expression patterns of *giant* at cleavage cycle 14. Data obtained from Kraut and Levine [Kraut and Levine, 1991]. Anterior end to the left and dorsal side up. Red arrows indicate the shifts in the stripe boundaries of *giant* posterior domain (stripe number 4). 1) *Krüppel* mutant. Stripe 4 expands anteriorly to mid-embryo. 2) Wild-type expression. 3) *Hunchback* mutant. Stripe 4 expands posteriorly. 4) *Tailless* mutant. Stripe 4 retracts only partially from the posterior tip, it expands more posteriorly compared to *Hunchback* mutant. 5) Embryo lacking maternal product *Torso*. Posterior expansion, extending all the way to posterior pole, showing a left-handed threshold (note that the Caudal gradient is inverted compared to the one in our model (Fig. 0.11)). (B) Some of the phenotypic changes discussed in the paper are at least biologically plausible (Fig. 3.11).

wide (0-6), as certain genotypes can access no novel phenotype, whereas others connect the stripe phenotype to more than one novel phenotype (Fig. 3.14). We refer to the genotypes that can reach a high diversity of distinct innovations as *phenotypic hubs* (phenotypic diversity ≥ 3). In this alternative measure, mechanisms show distinct proportions of phenotypic hubs, with certain mechanisms showing a sevenfold higher proportion of hubs than others (*Frozen Oscillator* compared to *Classical*).

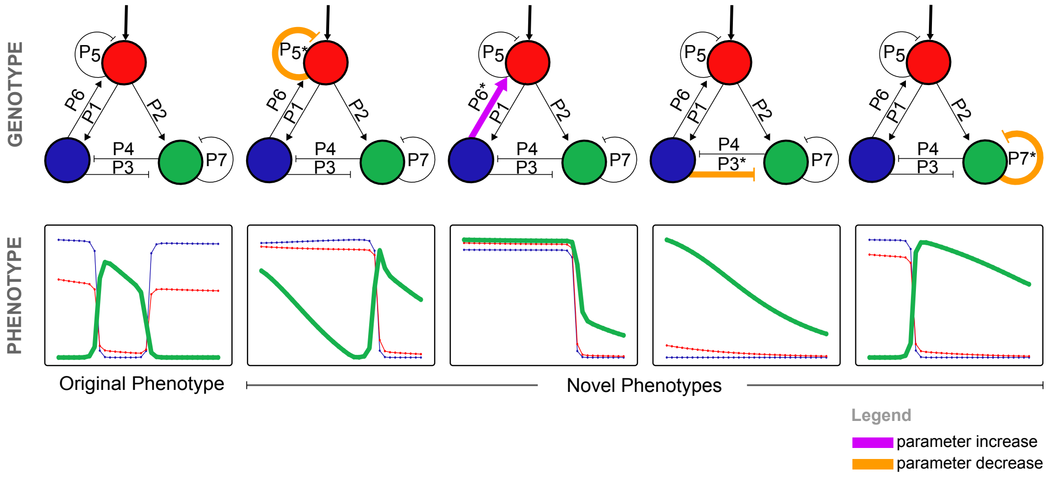

GENOTYPE
PHENOTYPE
P5
P6
P1
P2
P4
P3
P7
P5*
P6*
P3*
P7*
Original Phenotype
Novel Phenotypes
Legend
parameter increase
parameter decrease

Figure 3.13: **Phenotypic diversity accessible from a genotype: the existence of phenotypic hubs.** From an original genotype producing the stripe phenotype with underlying Mutual Inhibition mechanism and interaction strengthsP_1-P_7={0.35; 5.84; -5.35; -0.08; -0.48; 0.33; -0.25}, 4 novel phenotypes are accessible. Single mutations lead to Gradient and peak(P_5^*=-0.15), Left-handed threshold (P_6^*=0.9),Gradient (P_3^*=-3) and Right-handed threshold (P_7^*=-0.05).

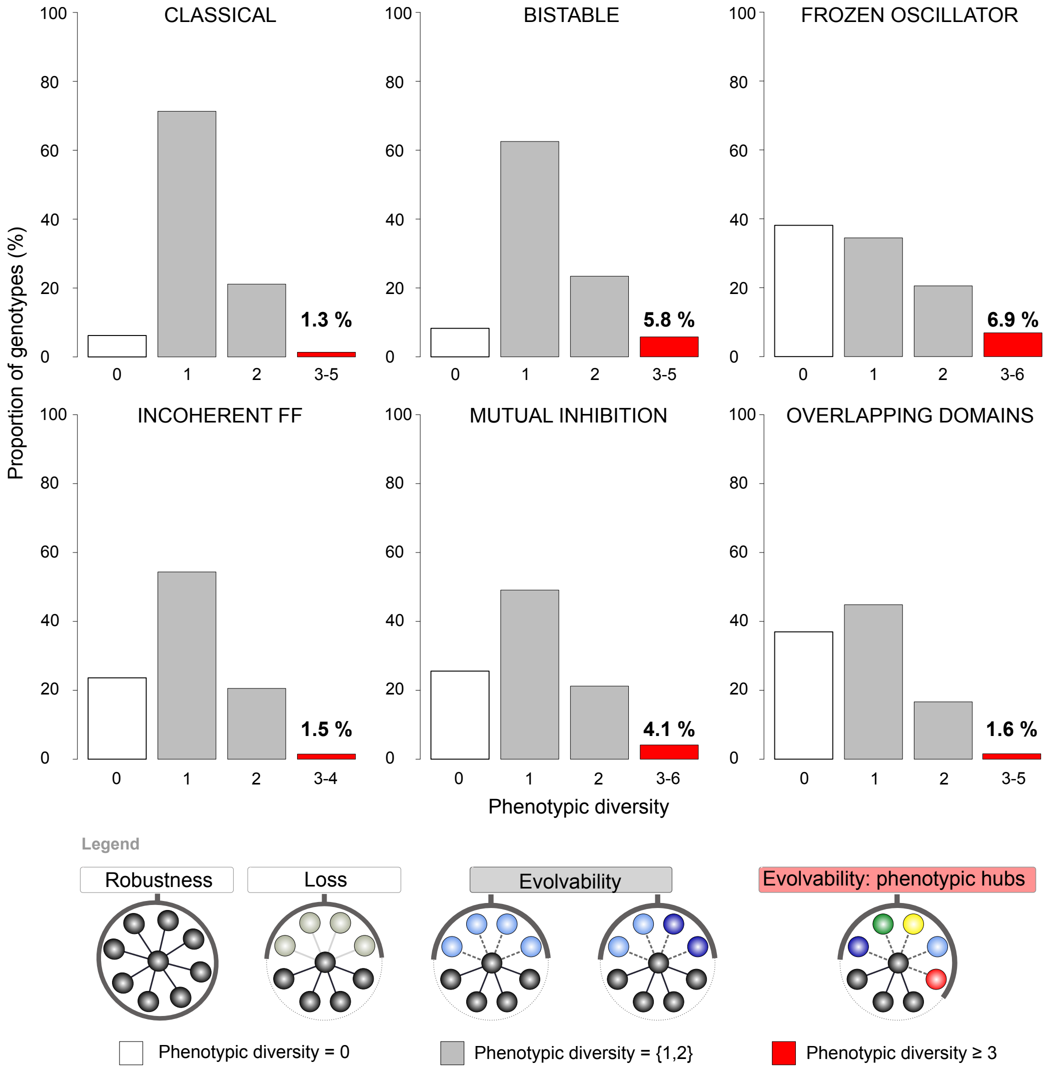


Figure 3.14: **Distribution of phenotypic diversity per mechanism.** Proportion of genotypes with different phenotypic diversity within their neighbourhoods. Phenotypic diversity measures the number of distinct phenotypes accessible in the neighbourhood of a genotype for the whole range of mutation strengths. Percentages over red bars indicate the proportion of genotypes behaving as phenotypic hubs (phenotypic diversity ≥ 3). The existence of phenotypic hubs is a characteristic of all mechanisms but their proportion varies.

Chapter 4

DISCUSSION

To summarize, we measured the evolvability of different gene circuits that use different dynamical mechanisms to achieve the same function: to convert a morphogen gradient into a single stripe of gene expression. For each mechanism, we explored the landscape of accessible novel phenotypes and found that the evolutionary potential clearly varies between alternative mechanisms.

Many previous studies emphasized that the majority of successful genotypes are grouped together into a large and connected neutral region of genotype space (Fig. 1.1). Our mechanisms-based view makes an important distinction. By using a biologically realistic model, with continuous variables, parameters, and regulatory functions, our neutral region within genotype space is broken up into separate islands, which represent distinct dynamical mechanisms (Fig. 2.1). As a consequence, genotypes can neutrally drift only within the region of their dynamical mechanism and therefore access only the phenotypes adjacent to this region (Fig. 4.1). In other words, neutral drift, which is considered to have a key role in evolution, is strongly restricted by genotype's dynamics.

A distinctive feature of the current study stands out in the general context of genotype-phenotype studies. Because our emphasis is on evolution of novel phenotypes, we chose to make our phenotypic categories represent qualitatively distinct patterns. A small qualitative shift in an expression pattern (e.g. 2% wider stripe) could theoretically have functional (selectable) consequences, but we do not consider it a novel phenotype for this study. This choice differs from the approaches taken in previous studies. For example, in studies of the relationship between RNA sequence and secondary structure, the phenotype space for an RNA molecule of N bases is considered to have approximately 1.8^N distinct phenotypes citeHuynen1996, even though many of these are only tiny alterations in secondary structure. This approach has led to the finding of *perpetual innovation*: the concept that the number of accessible novel structures is roughly constant irrespective of the molecule's position in its neutral space [Huynen, 1996]. Hence, in contrast to the vastness of RNA phenotype space, our definition of phenotypes reduces the diversity of the phenotype space itself and therefore impacts on the measure of evolvability. We consider this definition to be more appropriate for the question of phenotypic novelty.

From the broad perspective of developmental biology, the concept of evolvability focuses on the generation of novel phenotypes (i.e. a potential), whereas the concept of developmental constraints refers to restrictions on the production of certain phenotypes (i.e. a limitation). Independent of the context, both evolvability and developmental constraints describe the available novelties. Our results show that the dynamics of gene circuits itself constitutes a developmental constraint, as initially envisioned by Maynard Smith in the 1980s who defined the constraints as the 'biases on the production of variant phenotypes...caused by the structure, character, composition, or dynamics of the developmental system'[Maynard Smith et al., 1985].

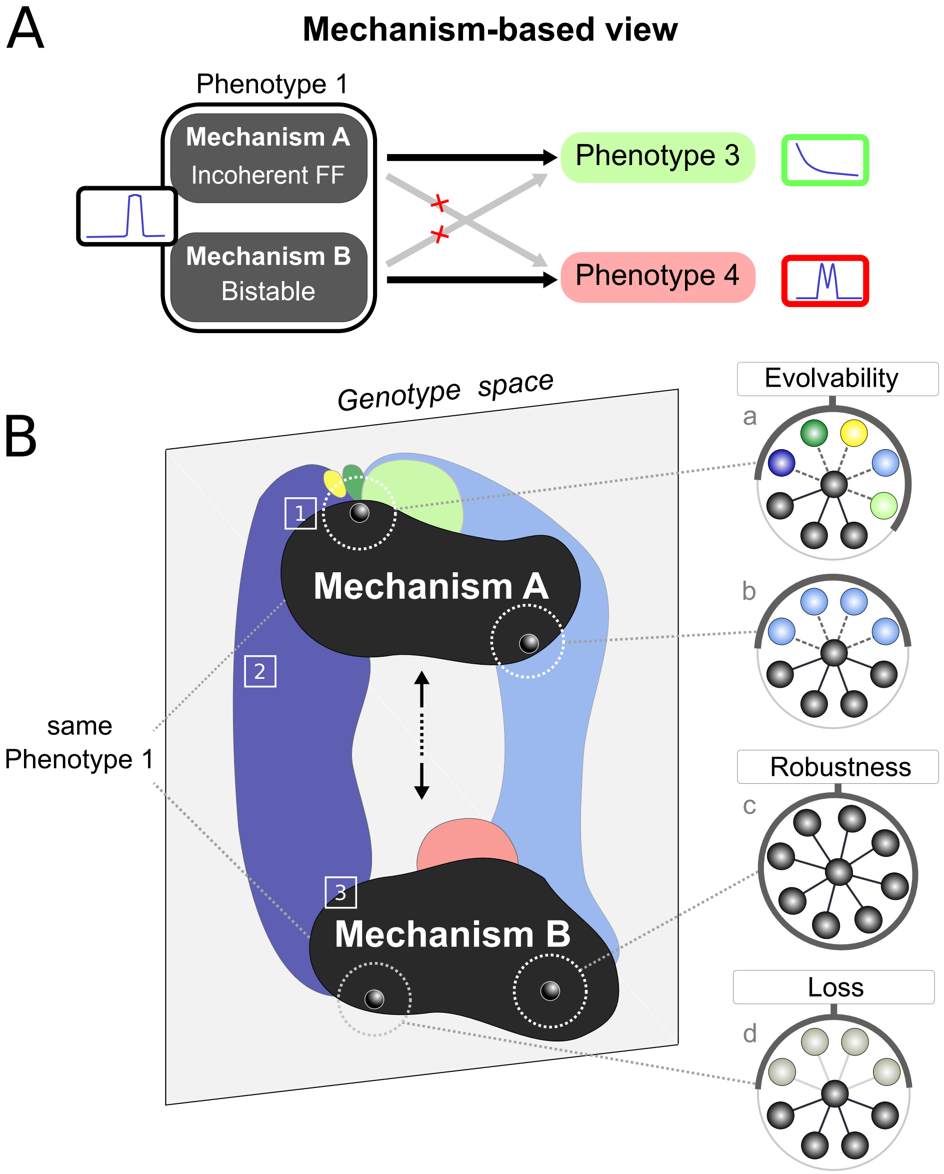

A
Mechanism-based view
Phenotype 1
Mechanism A
Incoherent FF
Mechanism B
Bistable
Phenotype 3
Phenotype 4
B
Genotype space
1
2
3
Mechanism A
Mechanism B
same
Phenotype 1
Evolvability
a
b
Robustness
c
Loss
d

Figure 4.1: **Mechanism-based view on evolvability.** (A) Distinct mechanisms to achieve a common phenotype have access to distinct phenotypes: Gradient phenotype appears exclusively accessible from *Mechanism A*– Bistable or Overlapping Domains–as is Double Peak from *Mechanism B* – Incoherent Feed-Forward or Mutual Inhibition. (B) Main features of genotype space: existence of phenotypic hubs (1), ubiquitously accessible phenotypes (2) and continuous phenotypic transitions (3). Four phenotypic neighbourhoods stand out as representative cases of either potentially evolvable genotypes – found close to the edge of the neutral space (panels a,b and d) and completely robust – internal – genotypes (panel c).

Part III

Modular properties of multi-functional gene circuits

Chapter 5

INTRODUCTION

The genome encodes a variety of different biological functions through the use of gene regulatory circuits. For example, the transcriptome of a cell can be pictured as a large circuit of interacting genes responsible for the many functions a cell performs. From a wider perspective, gene circuits can drive multiple functions at distinct levels: from differentiation of single cells into distinct cell types to the formation of a variety of multi-cellular patterns and organs. Which are the design properties of circuits with the capability to perform distinct functions –i.e. multi-functional circuits?

5.1 Structural modularity

In order to understand their capability to perform distinct functions, it has been proposed that biological circuits are modular, i.e. composed of distinct modules each performing a particular biological process. Although a module is essentially defined by its function –one module/one function– [Kholodenko et al., 2002, Eisen et al., 1998, Segal et al., 2003, Ten Tusscher and Hogeweg, 2011, Irons and Monk, 2007, Alexander et al., 2009], it is most frequently detected according to its structure. Indeed, from the many meanings of modularity, the most popular is *structural modularity*. Structural –or architectural– modularity considers modules as highly interconnected sets of genes. Structural modules are composed of genes that share more regulatory interactions among them than with genes outside the module [Hartwell et al., 1999, Wagner et al., 2007]. In this sense, structural modularity allows a large circuit to be decomposed into separate groups of genes –each group performing a distinct function. This type of arrangement is believed to confer several advantages such as high robustness and evolvability: the change or failure of one function does not necessarily lead to the malfunctioning of the rest, and each module/function can thus evolve relatively independently [Wagner and Altenberg, 1996, Raff and Conway Morris, 1996, Kirschner and Gerhart, 1998a, Brandon, 1999, von Dassow and Munro, 1999, Raff and Sly, 2000, Schlosser and Wagner, 2004]. In addition to these evolutionary advantages, structural modularity is also popular due to its reductionist nature –decomposing a large circuit into quasi-independent structures, and separably associating them to distinct biological functions, provides an intuitive understanding on how each biological process occurs. Because of the advantages it confers –high robustness and evolvability, intuitive understanding on independent biological processes – structural modularity is proposed as a suitable arrangement for circuits to perform multiple functions [Wagner and Altenberg, 1996, Raff and Conway Morris, 1996, Hartwell et al., 1999, Wagner et al., 2001, Wagner et al., 2007, Gerhart and Kirschner, 2007, Shubin et al., 2009, Espinosa-Soto and Wagner, 2010, Davidson, 2010]. Next we introduce two distinct classes of studies that promote structural modularity through distinct modelling criteria.

5.1.1 Structural modularity arises as large circuits adapt to perform multiple functions

A first class of studies evolves large circuits to achieve multiple goals or tasks. When large circuits are evolved to perform multiple biological functions, they tend to allocate distinct interconnected sets of genes –or modules- to each function [Di Ferdinando et al., 2001, Kashtan and Alon, 2005, Kashtan et al., 2009, Clune et al., 2013, Ellefsen et al., 2015, Espinosa-Soto and Wagner, 2010, Ten Tusscher and Hogeweg, 2011, Solé and Valverde, 2008].

In these studies, the tasks –or biological functions– that circuits must achieve can be very distinct. For example, a neuronal circuit is asked to visually recognize whether an object appears in the left and right side of the retina (Fig. 5.1A). A particular object, or image, is divided into two distinct groups of pixels for the left and right sides of the image. The left and right pixels respectively feed into two separate groups of nodes in the circuit. As the neural circuit is trained to recognize a number of objects, i.e. input patterns, it evolves into the following structure: the large circuit separates into two modules, each detecting a different side of the retina. Likewise, other studies consider similar neuronal circuits able to learn distinct independent tasks [Di Ferdinando et al., 2001, Clune et al., 2013, Ellefsen et al., 2015]. In this case, in order to perform task A while retaining the ability to perform task B, the circuit evolves to allocate a different set of neurons to each of the tasks (Fig. 5.1B). Both cases consider circuits that are bi-functional as they can simultaneously perform two distinct functions: monitoring both sides of the retina or achieving two distinct tasks.

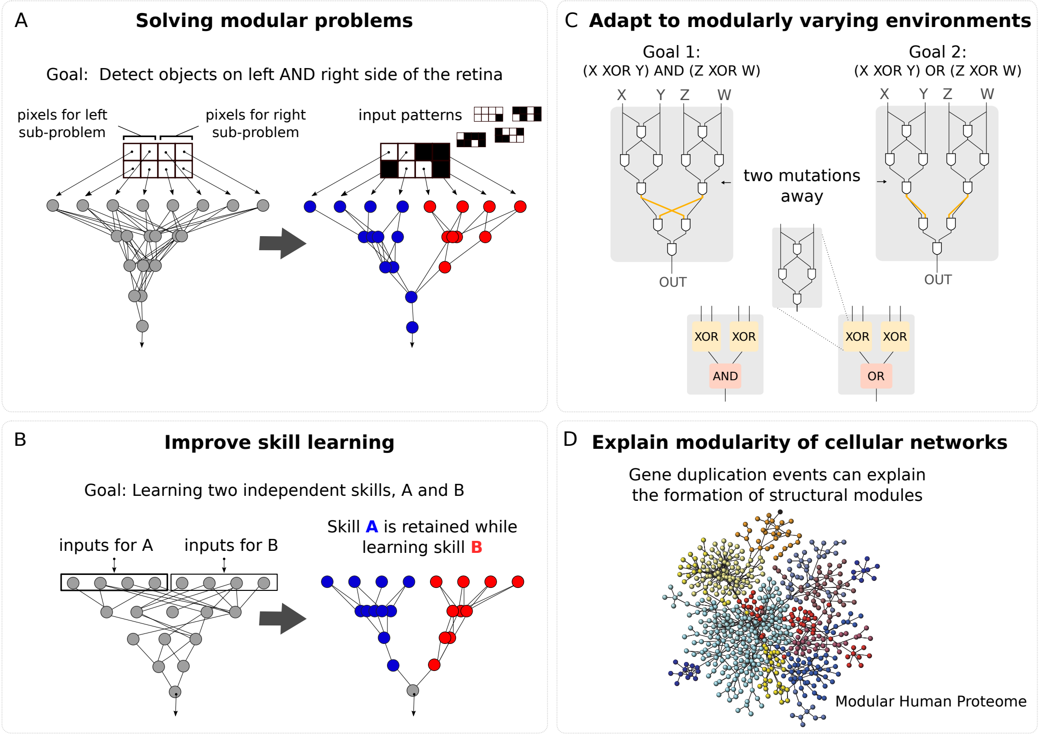

A
Solving modular problems
Goal: Detect objects on left AND right side of the retina
pixels for left sub-problem
pixels for right sub-problem
input patterns
C
Adapt to modularly varying environments
Goal 1:
(X XOR Y) AND (Z XOR W)
Goal 2:
(X XOR Y) OR (Z XOR W)
X
Y
Z
W
two mutations away
OUT
XOR
AND
OR
B
Improve skill learning
Goal: Learning two independent skills, A and B
inputs for A
inputs for B
Skill A is retained while learning skill B
D
Explain modularity of cellular networks
Gene duplication events can explain the formation of structural modules
Modular Human Proteome

Figure 5.1: **Structural modularity.** (A) and (B) display boolean neuronal circuits evolved to solve distinct modular problems. Each panel is adapted from [Clune et al., 2013] and [Ellefsen et al., 2015] respectively. (C) Structural modularity observed under modularly varying environments. Adapted from [Kashtan et al., 2009]. (D) A modular network is illustrated by means of the human proteome, i.e. a web of interactions where nodes are proteins and links indicate their physical (protein-protein) interaction. Gene duplication events can explain the structural modularity observed in such cellular networks. Adapted from [Solé and Valverde, 2008]

A similar approach reveals that modularity arises when circuits adapt under changing environments [Kashtan and Alon, 2005, Kashtan et al., 2009] (Fig. 5.1C). Large circuits are evolved to adapt to 'modularily varying goals'–MVG, i.e. the environment periodically switches between two distinct objective functions. As goals switch periodically, the circuit is left to modify its interactions –mutate– in order to adapt between them. For example, a first goal is to perform a boolean calculation goal 1: (X XOR Y) AND (Z XOR W) while a second goal consists in a different computation goal 2: (X XOR Y) OR (Z XOR W). Both goals can be divided into sub-goals –(X XOR Y) and (Z XOR W)– that are shared between both tasks. Intuitively, the circuit evolves to form two distinct structural modules for each of the sub-goals. Indeed, structural modularity is the most efficient arrangement in order to perform periodically changing tasks, as the circuit can switch between functions solely by changing a few interactions. Any other more complex wiring design would adapt slower between goals –more mutations needed. Although this type of circuit shows structural modularity it is not multi-functional. Unlike the previous neuronal circuits, this circuit cannot perform two functions simultaneously. Instead, as the circuit constantly modifies its interactions, two distinct effective circuits perform each of the two functions.

As we analyse the previous models, two critical issues arise. First, in all these cases structural modularity is clearly promoted/forced due to the decomposable nature of the functions to perform. Every problem to solve is decomposable into distinct sub-problems: two sides of the retina, two independent tasks, two sub-goals within boolean calculations. Hence, the resulting structural modularity appears as rather predictable. Would a circuit separate into distinct structural modules in a randomly varying environment? Indeed, as the same authors replace an MVG environment by a freely changing one where random goals, i.e. goals non decomposable into modular sub-goals, are changed periodically, complex wired solutions appear with no clear separable modules [Kashtan and Alon, 2005, Kashtan et al., 2009].

The second critical issue in these studies is circuit size (Fig. 5.1A-C). If a circuit is large enough to contain two non-overlapping modules, then it can split into two mono-functional circuits. Would instead a small circuit be able to separate into distinct modular components? Hence, it is not clear how a more limited amount of resources –fewer genes or neurons- would impact on structural modularity.

5.1.2 Structural modularity arises as specific ‘circuit building rules’are considered

A second class of studies aims at explaining the structural modularity observed in real biological circuits. These studies promote structural modularity imposing criteria such as neural connection costs or gene duplication events [Solé and Valverde, 2008, Ellefsen et al., 2015] (Fig. 5.1D). These criteria impact on the architecture of the circuit. For example, the rules on gene duplication events –as a gene (node) is duplicated, the new node inherits all original interactions with the rest of nodes– impose non-random patterns of connections that favour the formation of highly interconnected structures –structural modules. Interestingly, the authors believe this type of model can explain the structural modularity observed in cellular networks such as the human proteome.

5.2 Revisiting modularity

In all studies previously mentioned, there exists a bias towards picturing networks as composed of quasi-autonomous structures, i.e. structural modules. Structural modularity is promoted through distinct model formalisms –either imposing the solving of a modular problem, or considering distinct rules on gene interactions such as connection costs or gene duplication events. In contrast, non-modular solutions (i.e. multi-functional circuits that perform two functions in a non-decomposable manner) are rarely considered. However, integrated complex solutions have been proven to be as efficient in a few approaches [Bullinaria, 2007, Tosh and McNally, 2015, Ellefsen et al., 2015].

A particular observation in developmental an evolutionary biology question the separability of circuits into distinct functional groups of genes. Indeed, during the development of an individual organism, pleiotropy is the norm: the same sets of genes are essential for the morphogenesis of many different organs [Carroll et al., 2013]. This is exemplified by mutations in a single gene having phenotypic effects in distinct organs. Not only the same genes but also often the same entire signalling pathway plays multiple patterning roles. As such, just a handful of cell-cell signalling pathways (Hedgehog (Hh), wingless related (Wnt), transforming growth factor-B (TGF-B), receptor-tyrosine kinase (RTK), Notch, Janus kinase signal transducer and activator of transcription (JAK/STAT) and nuclear hormone pathways) are used repeatedly throughout development to drive patterning of many different organs [Pires-daSilva and Sommer, 2003]. In other words, the same sets of genes contribute to a variety of biological functions, and therefore are shared between functions. Hence, the phenomenon of pleiotropy challenges that of structural modularity as it suggest that the overlapping of modules within circuits may be common.

5.3 Search for compact circuits

In order to understand the multi-functional properties of circuits two contrasting views arise. On one hand, structural modularity suggests that large circuits can be broken up into distinct structural modules each performing a distinct function. On the other hand, observations such as pleitropy suggest that multiple functions might be encoded into the same small collection of genes.

We chose to explore this alternative scenario and search for small circuits that have the capability to perform multiple functions. Precisely, we explore how two functions can be encoded into the minimal collection of genes. This corresponds to

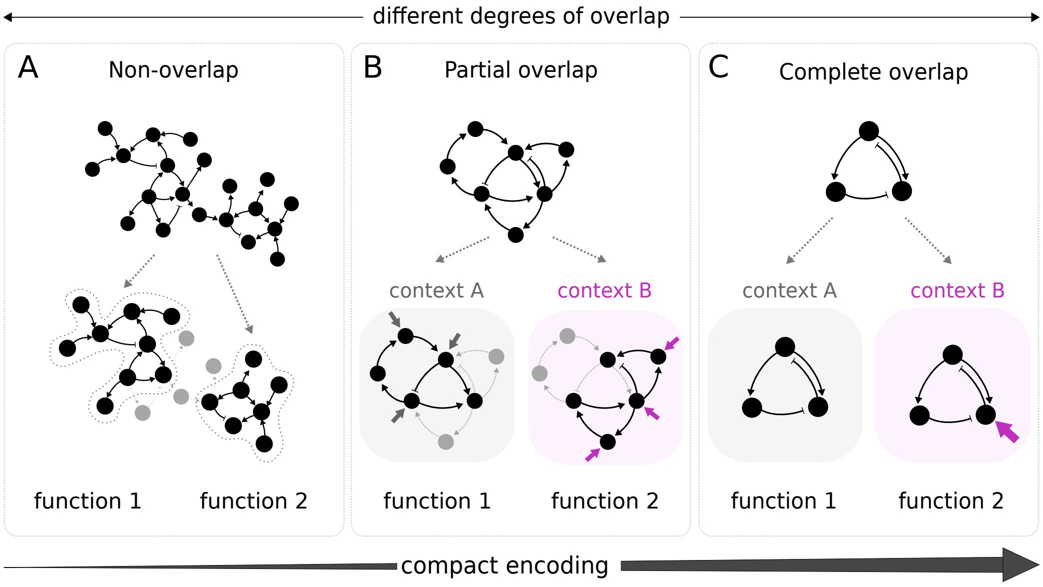

different degrees of overlap
A
Non-overlap
B
Partial overlap
C
Complete overlap
context A
context B
context A
context B
function 1
function 2
function 1
function 2
function 1
function 2
compact encoding

Figure 5.2: **Compact encoding of multi-functional circuits.** (A) In large multi-functional circuits, each function can be allocated an independent module [Di Ferdinando et al., 2001, Kashtan and Alon, 2005, Kashtan et al., 2009, Ten Tusscher and Hogeweg, 2011, Clune et al., 2013, Ellefsen et al., 2015]. (B, C) As circuits decrease in size there is a progressive overlap of functional modules. (B) Pink arrows represent the selective regulation of numerous specific genes in a context dependent manner. This selective regulation has been observed within single cells that adapt to different cell cycle states or stress conditions [Luscombe et al., 2004, Bandyopadhyay et al., 2010]. (C) Could a minimal circuit achieve multiple functions using the same collection of interacting genes? Genes would interact identically in distinct contexts but the production rate of a single gene (pink arrow) changes in a context-specific manner, allowing a function-switch. These small multi-functional circuit confer strong information compression.

the strongest case of information compression –more functions encoded by fewer genes. The focus of this part of the thesis is thus to search for compact circuits that compress two different functions into the smallest collection of interacting genes; in other words, the search for minimal multi-functional circuits (Fig. 5.2).

5.4 Defining a multi-functional circuit

n order to precisely define a multi-functional circuit we need to make the following distinction. We have reviewed in the introduction a particular motif –the AC/DC circuit– able to perform two distinct functions –a switch-like behaviour and an oscillatory one– depending on the strengths of gene interactions [Panovska-Griffiths et al., 2013] (Fig. 0.8). Indeed, adjusting the strength of specific repressive interactions within the circuit causes a switch from one function to another. The AC/DC circuit describes a case of a particular topology achieving multiple functions for distinct parameter sets. While a given topology can display multiple behaviours or functions, mutations are necessary to evolve between the two distinct functions. We do not consider this type of circuit as multi-functional. Instead, we define a multi-functional circuit as a set of genes interacting with a fixed set of parameters –interaction strengths– able to achieve multiple functions. A multi-functional circuit can be assimilated to single signalling pathway, where genes that interact in the same manner are able to perform distinct patterning functions in a single embryo.

We take inspiration from the Notch-Delta signalling pathway, which exhibits two mutually exclusive, and qualitatively distinct behaviours: lateral induction or lateral inhibition. Lateral induction is the process by which a cell signals to its neighbours to adopt the same gene expression state. It causes a dynamic, progressive spreading of this state across the tissue, finally resulting in continuous domains of cells expressing the same genes. In contrast, during lateral inhibition, a cell inhibits its neighbours from adopting its own fate, leading to a 'salt and pepper'pattern of cells in alternating differentiation states. Within a single organism, Notch can drive induction in one tissue (e.g. patches of pro-sensory cells in the chick inner ear [Formosa-Jordan et al., 2012]) and inhibition in another tissue (e.g. the mosaic of neurogenic versus non-neurogenic cells in the chick's retina [Petrovic et al., 2014, Daudet and Lewis, 2005]). Many studies have modelled the molecular details of how Notch and Delta interact in real systems [Collier et al., 1996, de Celis and Bray, 1997, Panin et al., 1997, Lewis, 1996, Huppert et al., 1997, Horikawa et al., 2006, Formosa-Jordan and Ibanes, 2009]. By contrast, here we use this cell-signalling paradigm as the inspiration for a purely conceptual model, to answer a broader question on the possible design features of multi-functional circuits.

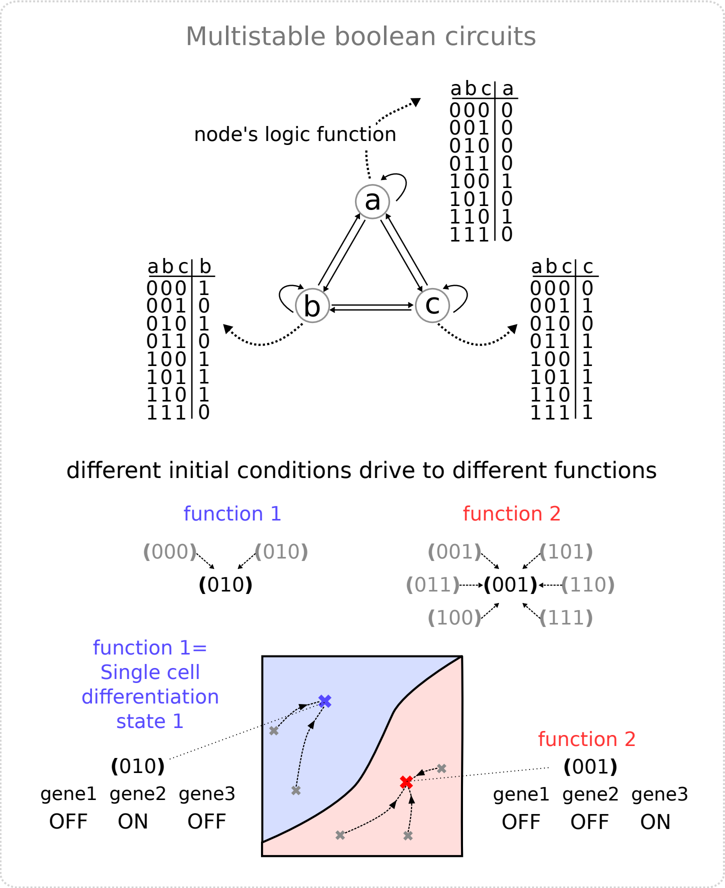
We find that indeed bi-functional circuits exist, which can switch between two qualitatively different patterning functions without any changes to their topology or regulatory parameters, but simply by changing the basal expression level of one gene in the circuit (which could be controlled by a tissue-specific transcription factor). Furthermore, our analysis uncovers two types of circuit designs: firstly, those which are composed in an intuitive way from two simpler mono-functional circuits, and secondly, emergent designs which cannot be easily decomposed into their component sub-circuits. These findings illustrate the potential complexity and efficiency regarding how phenotypic information is encoded in the genome. Importantly, both genes in these circuits are essential for both functions, and this has important consequences for our understanding of modularity.

Chapter 6

METHODS

6.1 Choice of the model

We aim at exploring the design features of a minimal circuit with the capability to perform distinct patterning functions. The search for multi-functional circuits has been explored from a distinct perspective: that of multi-stable boolean circuit [Martin and Wagner, 2008, Payne and Wagner, 2013] (Fig. 6.1). Indeed, certain boolean circuits can show multi-stability, which is used as a proxy for multi-functionality. As previously described, a given boolean circuit can achieve a particular phenotype, or function, that corresponds to a binary vector describing the transcription state of a groups of genes –either transcribed on/1 or not off/0. Martin et al. [Martin and Wagner, 2008] have shown that, as a boolean circuit is initialized from different initial states, it is able to reach distinct functions, i.e. single-cell differentiation states. Thus, each of the attractors of the system corresponds to a particular abstract biological function (Fig. 6.1). Instead, we chose a model of tissue patterning where the functions we explore are specific biological patterning functions. Although there exist similarities between the more abstract boolean model and our model, we investigate innovative aspects such as the formation of real and concrete multi-cellular patterns and, more importantly, the modular properties of multi-functional circuits.


Multistable boolean circuits
node's logic function
a
b
c
different initial conditions drive to different functions
function 1
(000)
(010)
(010)
function 2
(001)
(101)
(011)
(001)
(110)
(100)
(111)
function 1=
Single cell
differentiation
state 1
(010)
gene1 gene2 gene3
OFF ON OFF
function 2
(001)
gene1 gene2 gene3
OFF OFF ON

abc	a
000	0
001	0
010	0
011	0
100	1
101	0
110	1
111	0

abc	b
000	1
001	0
010	1
011	0
100	1
101	1
110	1
111	0

abc	c
000	0
001	1
010	0
011	1
100	1
101	1
110	1
111	1

Figure 6.1: **Multi-stable boolean circuits.** Boolean circuit where each node behaves as a logic gate (see Fig. 0.2). Different initial conditions –inputs, lead to distinct attractors – outputs. Each attractor can represent a distinct biological function, where a given function corresponds to a particular single cell differentiation state. Adapted from [Payne and Wagner, 2013]

6.2 The gene regulatory model

6.2.1 Spatial model

To explore how paracrine-signalling leads to distinct patterns, we developed a simple model of direct cell-cell communication (similar to [Salazar-Ciudad et al., 2000, Plahte, 2001]) which considers both *intra-cellular* and *inter-cellular* gene regulation in a one-dimensional spatial system comprising 33 cells (Fig. 6.2). To search for minimal multi-functional circuits, we consider two-gene circuits with only one gene (the black node in Figure 6.2A) able to regulate expression of either gene in neighbouring cells.

The molecular details that characterize paracrine-signalling are implicit within the model. For example, an inter-cellular auto-inhibition of the signalling gene could encapsulate distinct molecular regulatory steps in a real biological system: delta binds to notch in a neighbouring cell, triggers cleavage of notch intra-cellular domain and inhibits its own expression.

Analogous to the model previously described in section 2.3, a circuit holds an underlying topology and a specific set of parameters. Notice that in this part of the thesis we have replaced the terminology of *genotype* for that of *circuit*. Here the topology is composed of a 2x2 intra-cellular matrix W_{intra} that describes how genes interact among them within the cell, and a 2x1 inter-cellular matrix W_{inter} specifying how the signalling black gene regulates genes in neighbouring cells. Within those matrices, the values 1, -1 and 0 represent activation, repression and no interaction, respectively.

A circuit holds a particular set of precisely 9 parameters: 6 for the strengths of gene interactions w_intra and w_inter, 2 to control gene-specific regulatory functions (1 parameter α per gene), and 1 controlling the steepness of the regulatory function β. In the process of parameter sampling (see section 7.1), parameters were generated randomly within the following ranges: regulation [0:10] range for activation and [-10:0] range for repression, α [-60;60] and $\beta 5, 10$.

The model captures the spatiotemporal dynamics of gene patterning and is described by:

$$\frac{\partial g_{ij}}{\partial t} = \chi[\Phi[\sum_{l=1}^{2} w_{intra}^{li}\, g_{lj} + w_{inter}^{li}\, g_{lj} + +S + C]] - \lambda g_{ij} + \eta(t) g_{ij} \qquad (6.1)$$

where g_{ij} is the concentration of the ith gene in the jth cell initially set to 0.1 for every gene in every cell; ϕ is the sigmoidal regulatory function–see details 6.2.2; w_{intra} and w_{inter} are matrixes containing the strengths of gene-gene interactions; we use two types of inputs kept constant throughout the simulation: S is the central signal, or trigger, received by one of the genes in the central cell (set to 1) and C is the context signal received by one of the genes in all cells throughout the tissue (set to 1 or 0 depending on the tissue-type); χ is the Heaviside function to prevent negative gene product production rates; λ is the decay rate (set to 0.05) and $\eta(t)$ is a noise term, which adds uniformly distributed fluctuations ($\pm 1\%$) to the concentration of every gene in every cell at every time step. Zero-flux boundary conditions are used throughout this work.

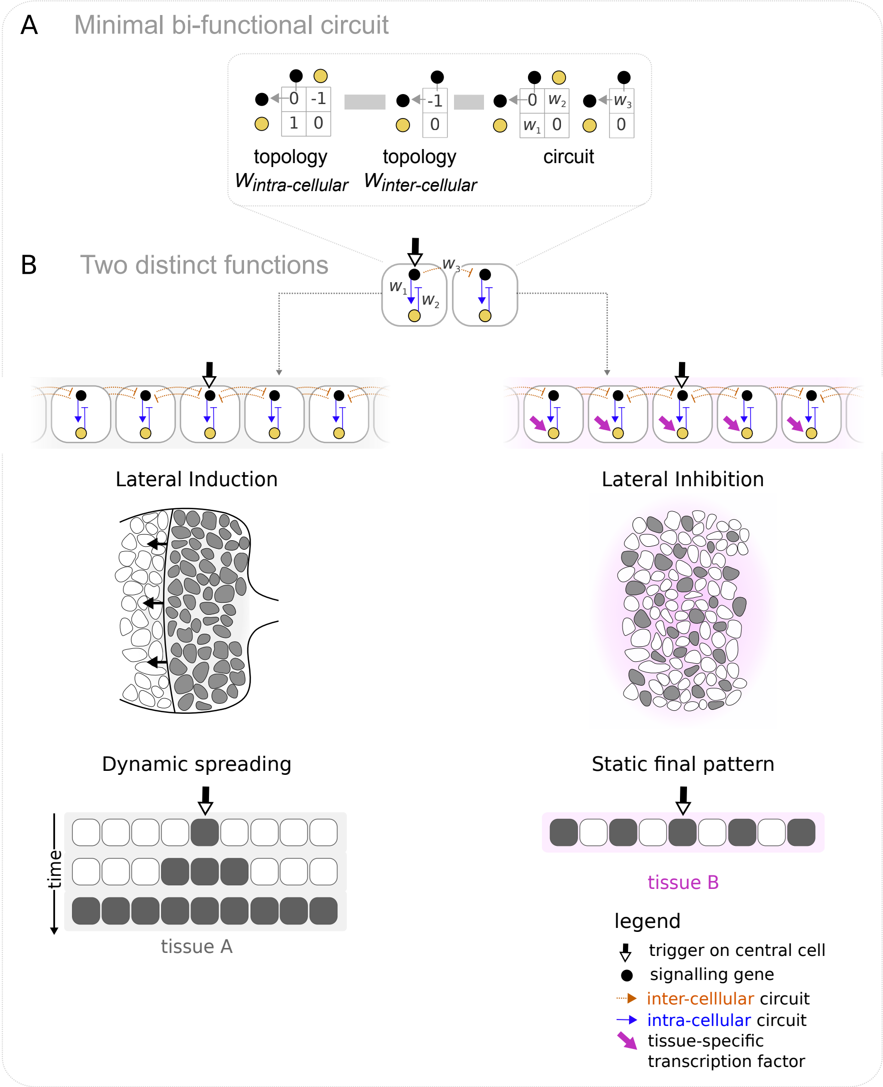

A
Minimal bi-functional circuit
0 -1
1 0
topology
$W_{intra\text{-}cellular}$
-1
0
topology
$W_{inter\text{-}cellular}$
0 w_2
w_1 0
w_3
0
circuit
B
Two distinct functions
w_1
w_2
w_3
Lateral Induction
Lateral Inhibition
Dynamic spreading
Static final pattern
time
tissue A
tissue B
legend
trigger on central cell
signalling gene
inter-celllular circuit
intra-cellular circuit
tissue-specific transcription factor

Figure 6.2: **Defining a bi-functional circuit.** (A) Gene expression is regulated through intra-cellular and inter-cellular circuits. An homogeneous signal specific to the tissue affects the production rate of one of the genes in all cells. Throughout the study, we use a simplified representation of a circuit consisting of the central cell and one of its neighbours. For simplicity, the inter-cellular circuit is only partially shown. (B) A multi-functional circuits switches between two qualitatively distinct multicellular patterns according to the tissue context it is embedded in. Analogous to biological processes such as the progression of the morphogenetic furrow in *Drosophila* [Sato et al., 2013], lateral induction leads to the propagation in time and space of a given gene expression state. Instead, lateral inhibition describes processes such as neurogenesis, where a fine-grained pattern of alternating cell fates is formed.

6.2.2 Flexibility of the regulatory function

As previously shown in the introduction, the regulatory function describes how genes respond to the various inputs they receive –integrated at the level of their *cis*-regulatory regions. We use an innovative function ϕ that includes a parameter – α— allowing for distinct regulatory logic. α controls the position of the steepest part of the sigmoidal curve and allows the effective logic of the integration function to be adjusted (Fig. 6.3). Depending on the range of values, alpha controls whether the gene responds to low or high levels of activation, i.e. whether it acts like and AND gate or and OR gate. It also allows adjustment of whether the gene shows a basal constitutive expression (in the absence of input values). In this manner, each gene within the circuit can adopt a range of distinct regulatory behaviours. This versatility is key to the results obtained.

6.3 Tissue specificity

Here we set out to identify a clear theoretical example in which two distinct patterning functions are performed by a small circuit, without any changes to its topology or modulation of regulatory parameters. As we have shown in the introduction, a single topology can achieve distinct patterning functions by modulation of its regulatory parameters. Instead, we propose that the circuit is kept unchanged. The change in function is triggered by the type of environment the circuit is embedded in (Fig. 6.2B). This way, as distinct patterning functions occur in distinct tissues, we seek circuits able to switch function depending on which tissue they are active in –provided by the context signal C. Tissue specificity is provided by the minimal conceivable influence on the circuit –just altering the basal expression level of one gene in the circuit –, which in a real embryo could be controlled by a tissue-specific transcription factor such as a Hox protein. Indeed, Hox proteins, are known to control cellular patterning in many contexts including the segmental organization of the hindbrain [Narita and Rijli, 2009].

6.4 Objective functions: lateral induction and lateral inhibition

Analogous to biological processes such as the progression of the morphogenetic furrow in *Drosophila* [Sato et al., 2013], lateral induction leads to the propagation in time and space of a given gene expression state. A circuit achieves induction when it causes the progressive spread of a trigger received by the central cell of the tissue.

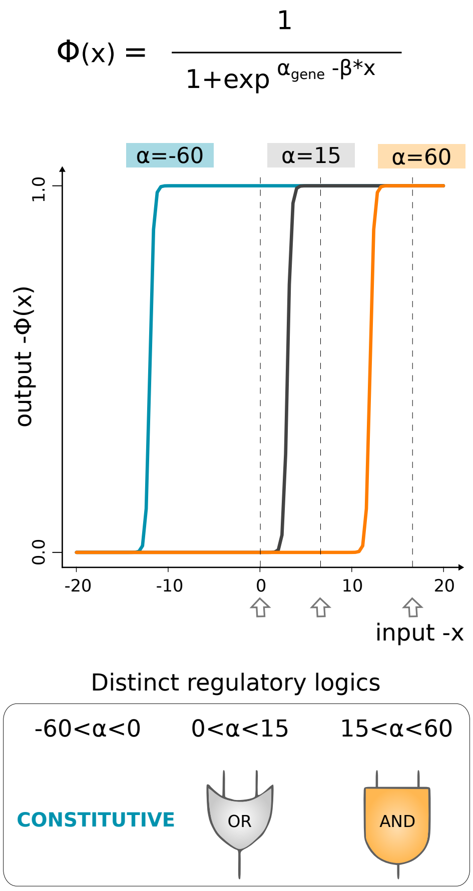


Figure 6.3: **Versatility of the regulatory function.** A regulatory function describes the relationship between the total input into a gene and its output concentration. The regulatory function we use holds gene-specific parameters α that allow for distinct regulatory logics. The value of α controls the position in the x-axis of the steepest part of the curve. α ranges from 60 to -60 and discriminates between distinct regulatory logics. For a [-60:0] range, a gene is constitutively expressed -if the gene does not receive input it is still transcribed. When α belongs to [0:15] range, the regulatory logic is analogous to a boolean OR gate: only a relatively small amount of input is necessary for the gene to be expressed. Last, for a [15:60] range, the regulatory function is analogous to an AND gate: large amounts of input are necessary for transcription.

Instead, lateral inhibition describes processes such as neurogenesis, where a fine-grained pattern of alternating cell fates is formed. A circuit achieves inhibition when it causes consecutive cells to be in alternating gene expression states. Hence, an induction pattern is assessed dynamically throughout the simulation while an inhibition pattern is assessed once the pattern has reached equilibrium.

An induction pattern occurs when the central signal S expands through the tissue. At specific time steps –every 50 time steps–, we measure the expansion of a gene as the number of cells adjacent to the central cell for which the expression level is high. During the simulation, the expansion level needs to increase at least 5 times in order to prove a progressive spreading through the tissue. Finally, at equilibrium, the expansion level is equal to the total number of cells, i.e. all cells express homogeneous high expression level for the gene (Fig. 6.2B).

An inhibition pattern occurs when a given gene sees its steady state expression level alternate between high and low expression states at least 14 times. In a model composed of 33 cells, a perfect 'salt-and-pepper'pattern of single-cell wavelength would require 16 single-cell alternations. Hence, we do not select only for perfect patterns but accept imperfections where two consecutive cells are found in the same state (Fig. 6.2B).

Besides these two particular patterns –induction and inhibition–, Notch-Delta orchestrates a variety of other processes, from the formation and sharpening of a boundary –such as the establishment of vein thickness in the adult *Drosophila* wing[Huppert et al., 1997, de Celis et al., 1997] or the formation of the dorsoventral boundary in *Drosophila* wing imaginal disc [de Celis and Bray, 1997, Panin et al., 1997]–to the regulation of the mouse segmentation clock during somitogenesis [Dequéant et al., 2006, Horikawa et al., 2006]. However, we believe these more complex patterns results from combinations of the two core induction and inhibition behaviours. For example, boundary specification uses both inhibition and induction in a complex fashion to control for the thickness of the veins [de Celis and Bray, 1997, Huppert et al., 1997] while during somitogenesis lateral induction facilitates synchronized oscillations [Dequéant et al., 2006, Horikawa et al., 2006]. Thus, from the many Notch-Delta driven patterns, we chose to study induction and inhibition as they constitute basic building blocks to build a variety of patterns and organs.

Chapter 7

RESULTS

Our goal is to understand the design properties of bi-functional circuits – in particular, the degree to which they are composed of distinct sub-circuits, i.e. their modularity. We follow a two-step process. First, we explore circuit space to detect all circuits performing lateral inhibition or lateral induction (7.1). This search is done in the absence of tissue-specific signals. As the switch between functions is triggered by a change in the tissue environment, these circuits are thus mono-functional. Second, we use the results obtained in this first search to find bi-functional circuits (7.2). For that, we subject the initial pool of mono-functional circuits to a different tissue environment and select those able to switch function, i.e. bi-functional.

7.1 Mono-functional circuits

To identify mono-functional circuits able to perform induction or inhibition, we performed an unbiased search through parameter space by enumerating all the 1,200 possible two-gene topologies and sampling large numbers of parameter sets (10^7 per topology). This search is done in a two-step process analogous to (2.4).

We selected successful mono-functional circuits (601 topologies were found for lateral induction and 655 for lateral inhibition) and constructed a complexity atlas [Cotterell and Sharpe, 2010] (Fig. 7.1A). As discussed in 0.3.3, the layout of the atlas is designed to illustrate relationships between circuit designs (pairs of topologies are linked if differ by the addition/deletion of a single interaction), and is arranged such that the simplest circuits (fewer gene-gene interactions) are found at the bottom of the atlas. In this way, the minimal designs corresponding to different ways to achieve each function (dynamical mechanisms) are found in different 'stalactites'[Cotterell and Sharpe, 2010] (Fig. 7.1A).

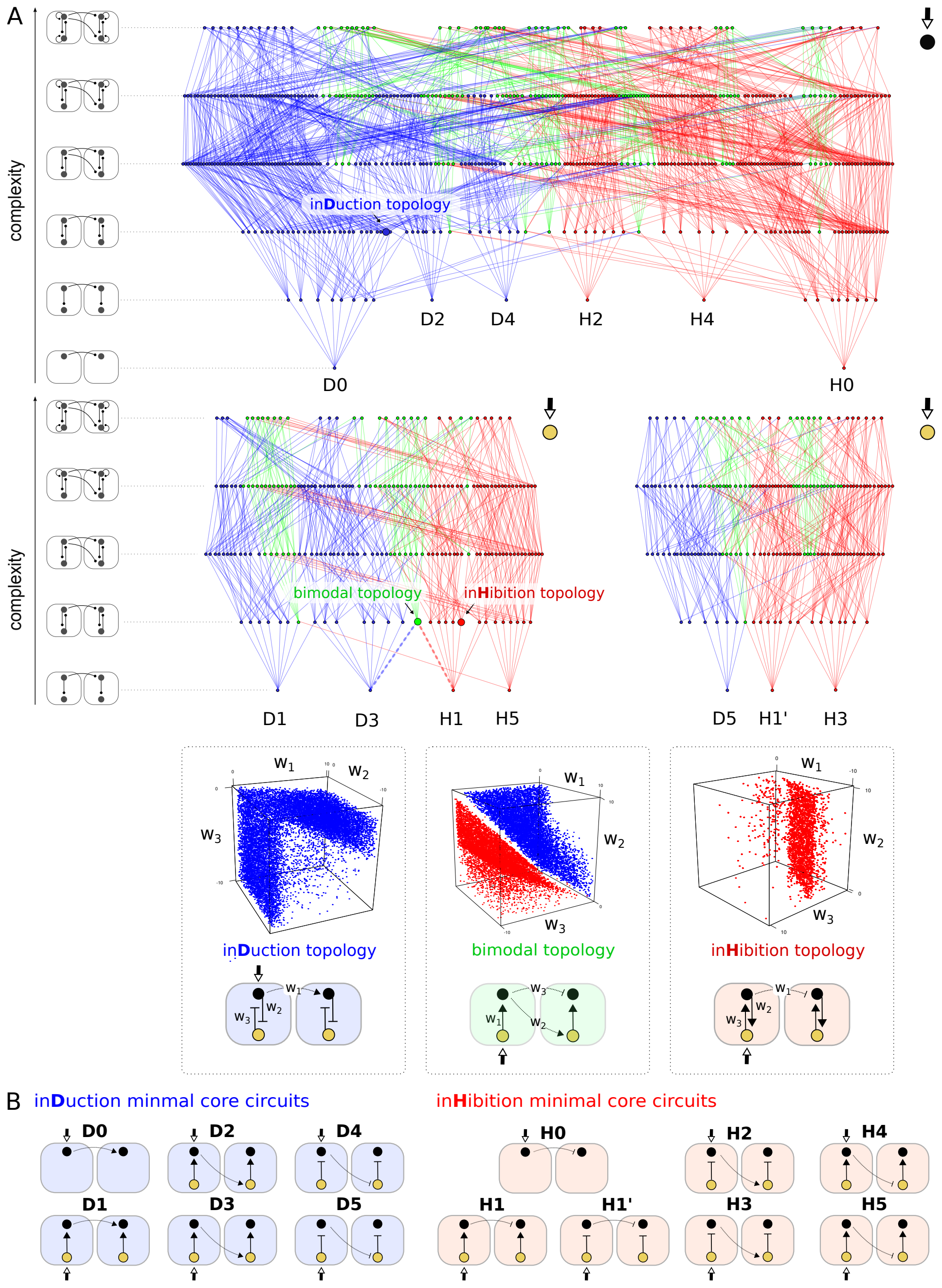

A
complexity
inDuction topology
D2
D4
H2
H4
D0
H0
complexity
bimodal topology
inHibition topology
D1
D3
H1
H5
D5
H1'
H3
w_1
w_2
w_3
inDuction topology
bimodal topology
inHibition topology
B
inDuction minmal core circuits
inHibition minimal core circuits
D0
D2
D4
D1
D3
D5
H0
H2
H4
H1
H1'
H3
H5

Figure 7.1: **Color-coded complexity atlas of mono-functional circuits.** A) Topologies -nodes- are coloured according to their function -blue and red exclusively hold circuits capable of induction or inhibition respectively and green bi-modal topologies hold circuits with both behaviours.(B) Stalactites help identify basic core circuits achieving induction (D0 to D5) and inhibition (H0 to H5)

In this case, our atlas is colour-coded to illustrate the two different patterning functions: blue for lateral induction, red for lateral inhibition and green for those topologies that can perform both functions. Figure 7.1A shows three distinct topologies among which two exclusively contain circuits performing induction or inhibition. A third topology is bi-modal, i.e. it holds within a single architecture circuits with both behaviours depending on parameter values. Indeed, we observe that induction and inhibition occupy distinct regions in parameter space.

From the structure of stalactites we identify the minimal induction [D0 to D5] and inhibition [H0-H5] circuits (Fig. 7.1B) that make use of six distinct dynamical mechanisms –three for each function (Fig. 7.2). Notice that within a mechanism, we find at least two equivalent circuits –for example D0 and D1, depending on which gene receives the central signal. They essentially use the same dynamical strategy. In order to achieve induction, three mechanisms are found, using distinct gene expression dynamics in time and space. The unique time-course of gene expression profiles is plotted under each mechanism. *Auto-activation* leads to an independent expansion of the signalling gene. Instead, *activate-activator* results in a synchronous expansion of both genes. Last, a non-overlapping complementary expansion is unique of *inhibit-inhibition*, as the signalling gene inhibits its intra-cellular inhibitor, 'pushing-away'its expression. Within inhibition mechanisms, inter-cellular *auto-inhibition* of the signalling gene causes the neighbour cell to lose its inhibitory potential on next cell, which in turn adopts a constitutive high concentration. In *activate-inhibitor*, intra-cellular inhibition causes single cells to adopt opposite concentration levels for each gene, leading to an alternating expansion. Last, if the signalling gene inhibits its intra-cellular activator, single cells adopt high expression levels for both genes in a coupled expansion.

At this point, we aim to clarify the role of noise in our model. Noise can drive lateral inhibition without the need of an initial trigger. Indeed, noise helps initial asymmetries self-amplify and can trigger the formation of a fine-grained patterns [Collier et al., 1996, Barad et al., 2011]. In Figure 7.3 we show for the same circuit, how lateral inhibition is formed with and without an initial trigger. Therefore, an underlying mechanism that is not detectable as a 'stalactite'is noise. Although noise can drive lateral inhibition on its own, it has also been suggested that it has a beneficial role into refining lateral inhibition patterns [Barad et al., 2011]. However, with respect to lateral induction, we discard noise as having an essential role in pattern formation, as a progressive spreading from the centre of the tissue cannot be triggered by random perturbations.

In a nutshell, our findings propose a handful of mechanisms likely to be used in real biological systems. Indeed, some of the mechanisms found –marked with

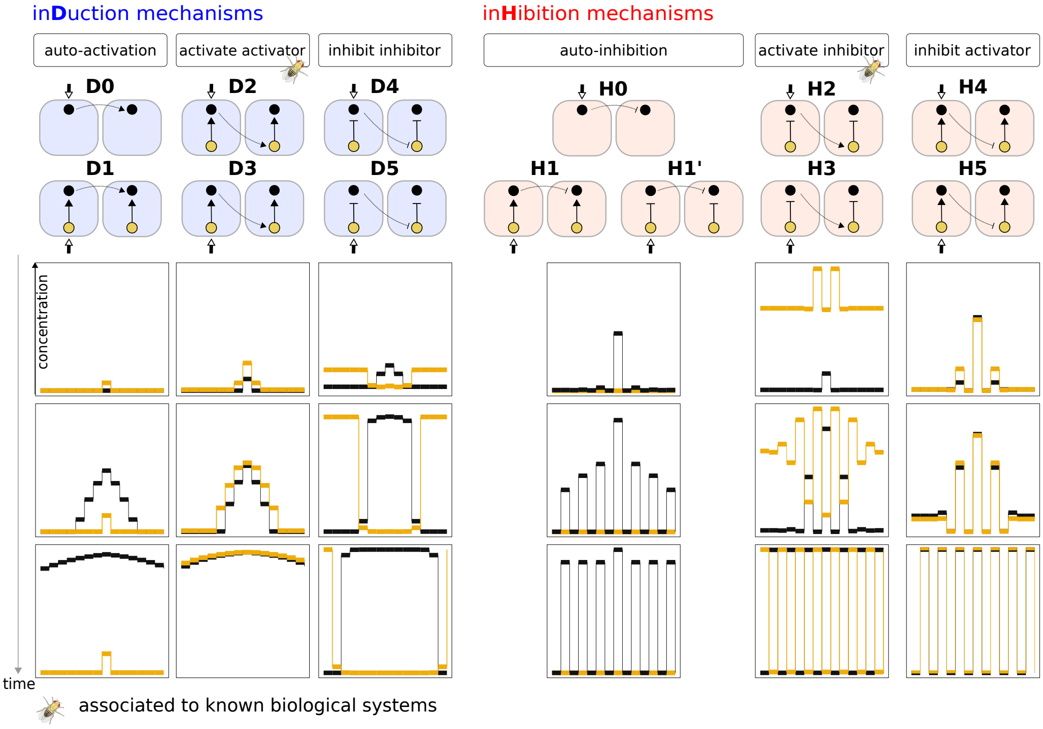

inDuction mechanisms
inHibition mechanisms
auto-activation
activate activator
inhibit inhibitor
auto-inhibition
activate inhibitor
inhibit activator
D0
D2
D4
H0
H2
H4
D1
D3
D5
H1
H1'
H3
H5
concentration
time
associated to known biological systems

Figure 7.2: **Alternative mechanisms to achieve induction and inhibition.** Core circuits are classified into three distinct mechanisms for each function. Each mechanism makes use of a different dynamical strategy, captured in the unique final profile. Within a given mechanism we find at least two equivalent topologies, depending on which gene receives the central trigger. Three distinct strategies are found for each of the functions. Among the core circuits found, some have been proposed to orchestrate several biological processes and are marked with a *Drosophila* cartoon. Although the simulation takes places in one a one-dimensional row of 33 cells, most graphic representations show 15 cells for a matter of clarity.

a *Drosophila* cartoon in Figure 6.3–are associated to known biological systems: *activate-activator* directs dorsoventral boundary formation in the *Drosophila* wing [de Celis and Bray, 1997, Panin et al., 1997] while *activate-inhibitor* has been proposed to orchestrate several biological processes such as regulation of neurogenesis [Lewis, 1996], morphogenesis of *Drosophila* wing veins [Huppert et al., 1997] or synchronization of oscillations during somitogenesis [Horikawa et al., 2006].

The diversity of mechanisms found here does depend on the versatility of our regulatory function, which is capable of AND-logic, OR-logic or constitutive activity (Fig. 6.1). Indeed, the mechanisms each use unique combinations of regulatory

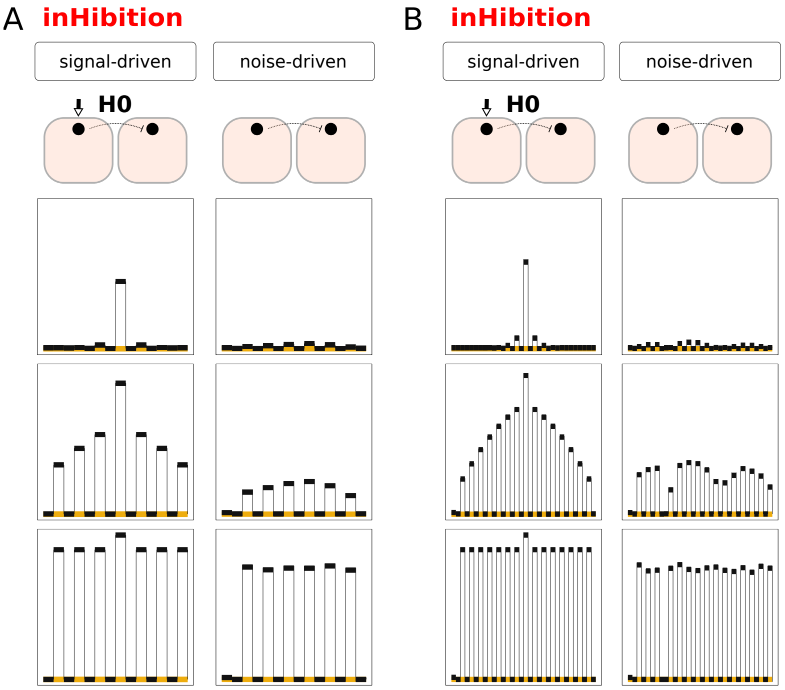


Figure 7.3: **Noise can drive lateral inhibition.** A signal in the center of the tissue is not necessary to obtain a fine-grained pattern. Indeed, it has already been observed that noise helps initial asymmetries self-amplify, thus triggering the formation of a fine-grained pattern [Collier et al., 1996, Barad et al., 2011]. We show how the same circuit can achieve inhibition with or without an initial trigger. Simulations are shown with distinct numbers of cells: 15 (A) or 33 (B).

logic within their circuitry. In Figure 7.4 we plot the range of α parameter for each minimal induction and inhibition circuit. First, we observe that within a mechanism, minimal circuits share a common arrangement of regulatory logic –for example circuits H2 and H3 or H4 and H5 hold the same combination of regulatory logic. Second, genes which receive only negative inputs –for example the signalling black gene in D4-H2-H3 or the golden gene in H4-H5– require constitutive expression in order for the circuit to function.

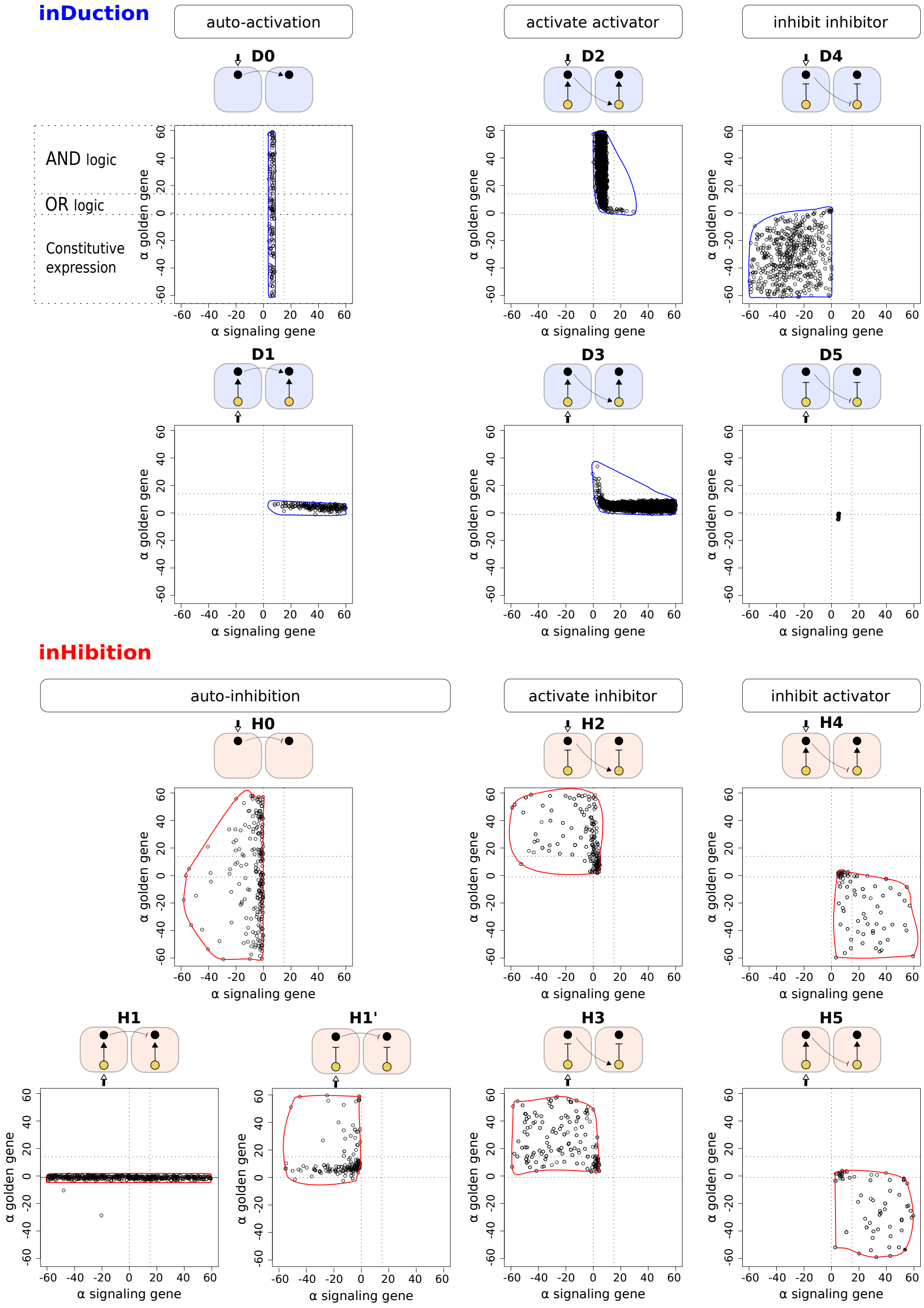


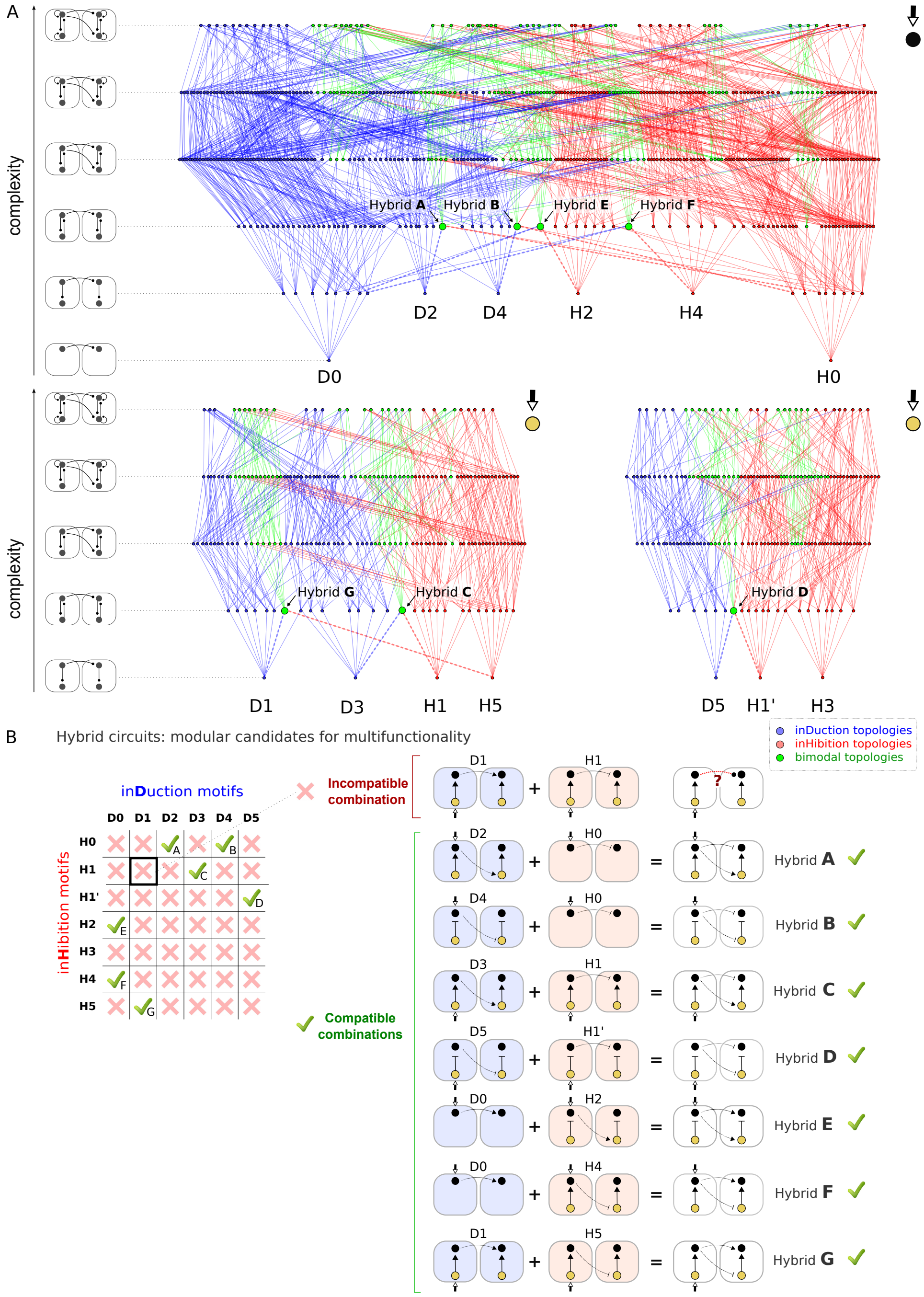
Figure 7.4: **Distinct mechanisms make use of distinct regulatory logic.** The high diversity of mechanisms found is allowed by the versatility of the regulatory function.

7.2 Bi-functional circuits

7.2.1 Candidates for bi-functionality

Which design features would a multi-functional circuit tend to employ? As we pointed out, the atlas illustrates that certain topologies (green nodes) are bi-modal, i.e. depending on parameter values they can be a mono-functional induction circuit, or alternatively a mono-functional inhibition circuit (Fig. 7.1A). We observe that bi-modal topologies are often the union of a core induction and inhibition circuits (for example, the union of D3 and H1 in Figure 7.1A, and all cases shown in Figure 7.5A). Could bi-modal topologies contain genuinely bi-functional circuits? In other words, could a version of these circuits be found in which the switching between induction and inhibition depends only on changing the basal expression level of one of the genes, rather than requiring a modification of the parameter values?

To explore this hypothesis we first determined all the possible *hybrid* circuits. Not all pair wise combinations of a D core circuit and an H core circuit are topologically possible. For example, D1 and H1 cannot be combined, because they have opposite signs for the same regulatory interaction. On this basis we found that of all the 42 hypothetical combinations, 7 are compatible which we label A to G (Fig. 7.5B). Hybrids conform modular candidates as they are composed of two separable sub-circuits. Each module, or sub-circuit, is composed of a subset of interactions. From a modular perspective, hybrid circuits could make use of each of their sub-structures to perform each function.


A
complexity
Hybrid A
Hybrid B
Hybrid E
Hybrid F
D2
D4
H2
H4
D0
H0
complexity
Hybrid G
Hybrid C
Hybrid D
D1
D3
H1
H5
D5
H1'
H3
B
Hybrid circuits: modular candidates for multifunctionality
inDuction topologies
inHibition topologies
bimodal topologies
inDuction motifs
inHibition motifs
Incompatible combination
Compatible combinations
D1
H1
?
D2
H0
Hybrid A
D4
H0
Hybrid B
D3
H1
Hybrid C
D5
H1'
Hybrid D
D0
H2
Hybrid E
D0
H4
Hybrid F
D1
H5
Hybrid G

	D0	D1	D2	D3	D4	D5
H0	✗	✗	✓A	✗	✓B	✗
H1	✗	✗	✗	✓C	✗	✗
H1'	✗	✗	✗	✗	✗	✓D
H2	✓E	✗	✗	✗	✗	✗
H3	✗	✗	✗	✗	✗	✗
H4	✓F	✗	✗	✗	✗	✗
H5	✗	✓G	✗	✗	✗	✗

Figure 7.5: **Modular candidates for multi-functionality.** (A) Hybrid circuits are the compatible union between a core induction and inhibition circuit (B) Hybrids (labelled A to G) combine within their topology core induction and inhibition circuits.

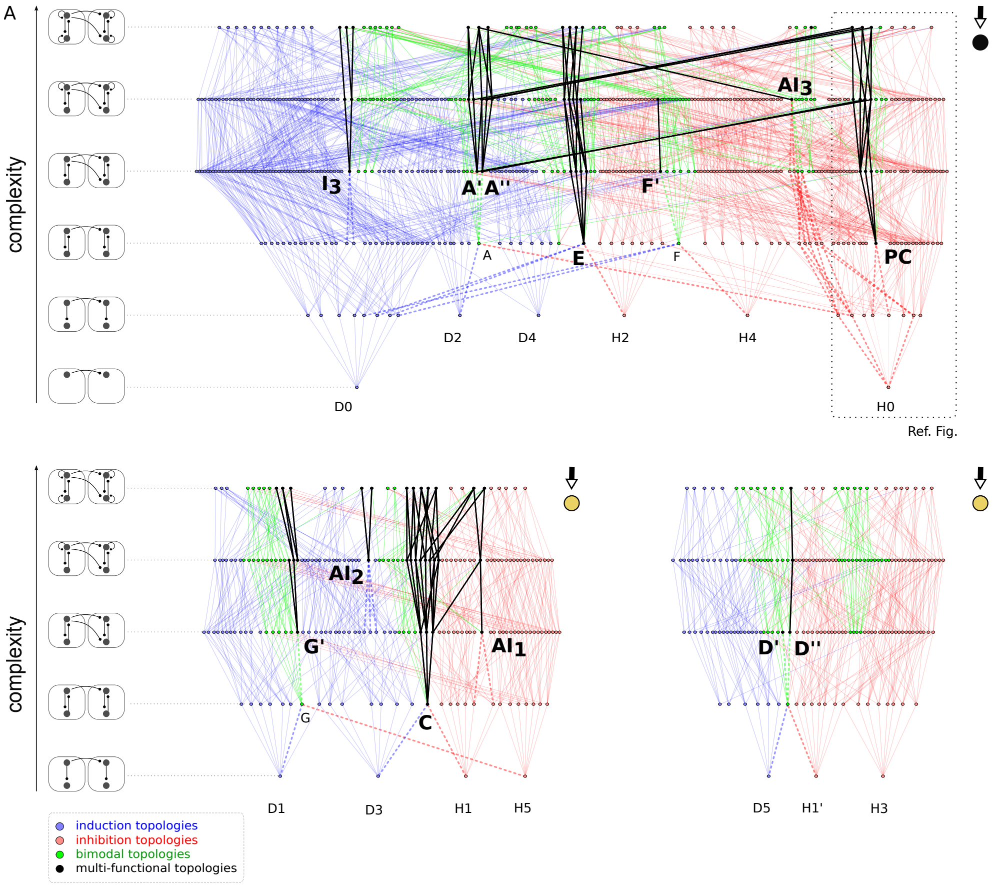
7.2.2 Search for bi-functional circuits

To search for bi-functional circuits, we subject all mono-functional circuits to a new tissue context. The tissue context consists in an external input signal that activates the expression level of one of the genes in all the cells of the tissue. While in the search for mono-functional circuits, this tissue context was not present (i.e. context signal equals to 0), in the new tissue the context is present (i.e. context signal equals to 0) and boosts the basal expression level of one of the genes. We simulate all mono-functional circuits in the new tissue environment and select those able to switch function. In other words, a successful bi-functional circuit must perform lateral induction in the first tissue, and lateral inhibition in the other, or *vice versa*. This analysis identified 72 different topologies capable of bi-functional behaviour (1130 circuits in total). Bi-functional designs can now be highlighted within our atlas (thick black lines in Figure 7.6A), showing that most bi-functional designs are within bi-modal regions (green). A full classification of the minimal designs for bi-functional circuits is given in Figure 7.6B.

We can now examine whether the hybrid designs (A-G) are bi-functional. Not every minimal hybrid gave bi-functional behaviour. For example, C (which is the hybrid of D3 and H1) does produce a bi-functional circuit, while G does not. However, we found that for all non bi-functional hybrids, the addition of an extra regulatory link could render the circuit successfully bi-functional. For example, the addition of positive auto-regulation to the golden gene of G, produces a bi-functional circuit which we label G'. All these modified hybrids are found one level higher in the atlas, and are listed as A' to G' (Fig. 7.6B).

However, more intriguingly, we found a second class of bi-functional circuits which are not hybrids of mono-functional designs. Instead they show a non-modular design which is not an intuitive composite of two separate circuits. Within this second class we find abundant *Activation-Inhibition* circuits that depend on intra-cellular negative feedback loops (Fig. 7.6B).

We have found two distinct classes of circuits –*hybrid* and *emergent*– which building principles allow us to visually detect them from their location within the complexity atlas (Fig. 7.7). While hybrid designs can always be seen at the point where two mono-functional stalactites unite, emergent designs emerge from within a mono-functional stalactite. The ‘emergent’, non-modular design of these circuits can be appreciated from their position within the complexity atlas. Instead of arising at the point where two mono-functional stalactites intersect, they emerge in the centre of a single mono-functional region. Their bi-functionality is thus not reducible to two distinct sub-circuits.


A
complexity
I3
A'A''
F'
AI3
A
E
F
PC
D2
D4
H2
H4
D0
H0
Ref. Fig.
AI2
G'
AI1
G
C
D1
D3
H1
H5
D'
D''
D5
H1'
H3
induction topologies
inhibition topologies
bimodal topologies
multi-functional topologies

B Table of multi-functional motifs

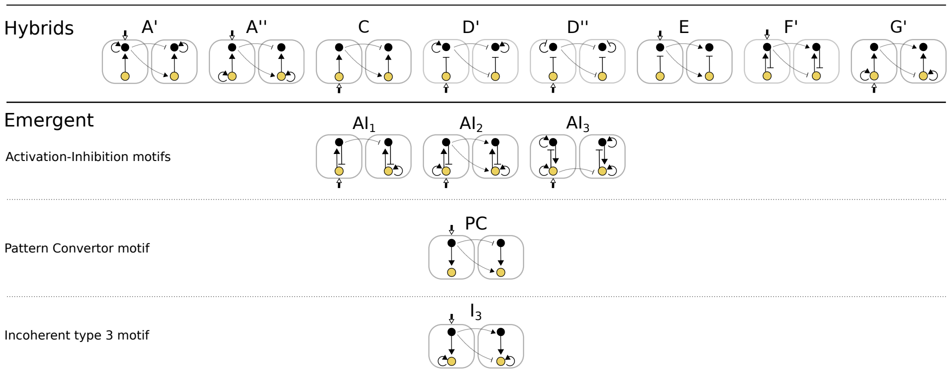
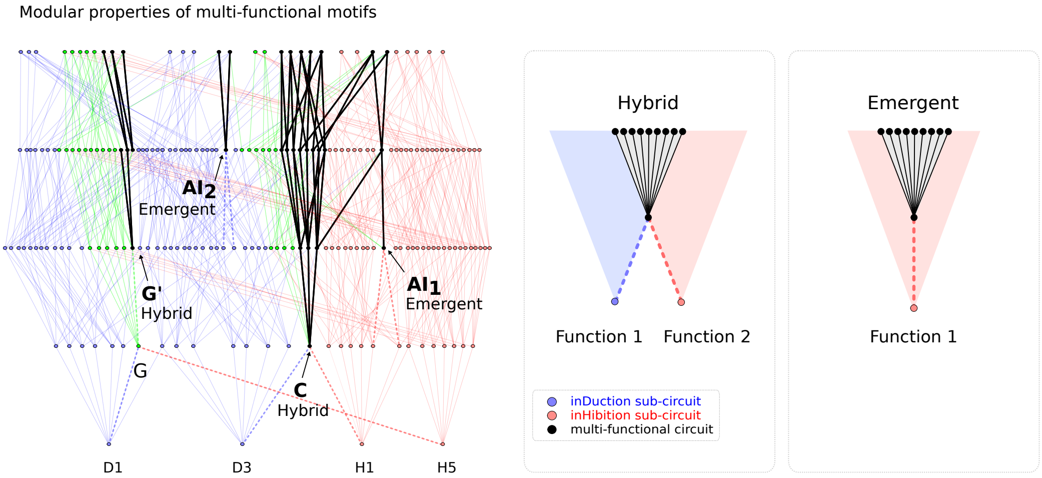

Hybrids
A'
A''
C
D'
D''
E
F'
G'
Emergent
Activation-Inhibition motifs
AI1
AI2
AI3
Pattern Convertor motif
PC
Incoherent type 3 motif
I3

Figure 7.6: **Complete atlas of multi-functional circuits.** (A) Atlas showing multi-functional circuits with 72 underlying topologies –coloured in black. The black stalactites helped us identify 13 core circuits. (B) Table of multi-functional core circuits classified according to their modular properties into hybrids and emergent.



Modular properties of multi-functional motifs
AI2
Emergent
G'
Hybrid
AI1
Emergent
G
C
Hybrid
D1
D3
H1
H5
Hybrid
Function 1
Function 2
inDuction sub-circuit
inHibition sub-circuit
multi-functional circuit
Emergent
Function 1

Figure 7.7: **Properties of hybrid and emergent multi-functional circuits.** Extract of the complete atlas (Fig. 7.6C) showing 4 out of the 13 multi-functional circuits. Multi-functional circuits show distinct modular properties. Hybrid circuits visually appear as the sum of two induction and inhibition circuits while emergent circuits emerge as more complex forms of mono-functional circuits. In this manner, the modular properties of multi-functional circuits can be visually detected.

When comparing both classes, while the design of hybrid circuits was predictable –as the union of two mono-functional circuits –that of emergent designs is not intuitive to predict. Therefore, it might seem that such non-intuitive, tightly overlapping emergent designs might be expected to have lower robustness. We show in Table 7.1 the robustness of every minimal bi-functional circuit. Robustness is measured as parameter robustness, i.e. the number of successful multi-functional circuits out of 10^7 parameter sets sampled. It provides an intuition on the dimensions of the parameter space for that specific circuit and often correlates to mutational robustness –how probable is it to maintain the phenotype –multi-functional capacity– as gene interactions are mutated. Table 7.1 shows no dramatic difference between the classes, highlighting that both design strategies may be equally plausible in real biological systems.

Hybrid A’	63
Hybrid A”	56 336
Hybrid C	7
Hybrid D’	1
Hybrid D”	2
Hybrid E	1
Hybrid F’	1
Hybrid G’	8

Emergent AI_1	10
Emergent AI_2	35
Emergent AI_3	2
Emergent PC	2
Emergent I_3	2

Table 7.1: **Robustness of multi-functional hybrid and emergent motifs.** Parameter robustness is measured as number of successful multi-functional genotypes out of 10^7 sampled.

7.3 Mechanisms for function switching

How does a small change in the tissue context drive a change in pattern? Here we explore the changes in phase portrait of the system upon the change in basal expression level for the context gene, and compare the underlying mechanisms used by hybrid and emergent circuits in order to switch between functions (Fig. 7.8-7.10). We use hybrid C (Fig. 7.8) and emergent circuits AI_1 and Pattern-Convertor (Fig. 7.9-7.10) as representative of each type of multi-functional circuit.

For such simple circuits, we use a 2-cell model and follow the concentrations of each of the species in time. For simplicity, we refer to the black signalling gene as Delta and to the golden gene as Notch.

7.3.1 Underlying mechanisms of hybrid circuits

We first study the underlying switching mechanisms of hybrid C. We can follow how the concentration of the four species evolves in time (Fig. 7.8B):

$$\frac{dN_1}{dt} = \frac{1}{1 + exp^{\alpha_N - 5\,(S + w_B D_2)}} - \lambda N_1 \tag{7.1}$$

$$\frac{dD_1}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(C + w_A N_1 + w_C D_2)}} - \lambda D_1 \tag{7.2}$$

$$\frac{dN_2}{dt} = \frac{1}{1 + exp^{\alpha_N - 5\,(w_B D_1)}} - \lambda N_2 \tag{7.3}$$

$$\frac{dD_2}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(C + w_A N_2 + w_C D_1)}} - \lambda D_2 \tag{7.4}$$

where $N_1(t)$, $D_1(t)$, $N_2(t)$ and $D_2(t)$ are the concentrations of Notch and Delta in the central cell –$cell_1$- and its immediate neighbour –$cell_2$.

We can then numerically solve this system for distinct values of C –two distinct tissues– to obtain $N_1(t)$, $D_1(t)$, $N_2(t)$ and $D_2(t)$. The species of interest are $D_1(t)$ and $D_1(t)$, as Delta is the gene which gene expression switches form inhibition to induction.

Thus, we aim at drawing a 2-dimensional phase portrait where each dimension corresponds to D_1 and D_2. For that, we reduce the dimension of the system. We impose equilibrium for Notch genes:

$$\frac{dN_1}{dt} = 0 \rightarrow N_1 = \frac{1}{\lambda\,(1 + exp^{\,\alpha_N - 5\,(S + w_B D_2)})} \tag{7.5}$$

$$\frac{dN_2}{dt} = 0 \rightarrow N_2 = \frac{1}{\lambda\,(1 + exp^{\,\alpha_N - 5\,(w_B D_1)})} \tag{7.6}$$

then replace N_1 and N_2 into equations (7.2) and (7.4) to obtain a reduced two-dimensional system:

$$\frac{dD_1(D_1, D_2)}{dt} = \frac{1}{1 + exp^{\,\alpha_D - 5\,(C + \frac{w_A}{\lambda\,(1 + exp^{\,\alpha_N - 5\,(S + w_B D_2)})} + w_C D_2)}} - \lambda D_1 \tag{7.7}$$

$$\frac{dD_2(D_1, D_2)}{dt} = \frac{1}{1 + exp^{\,\alpha_D - 5\,(C + \frac{w_A}{\lambda\,(1 + exp^{\,\alpha_N - 5\,(w_B D_1)})} + w_C D_1)}} - \lambda D_2 \tag{7.8}$$

For this reduced system, a phase portrait with the corresponding nullclines and steady states [Strogatz, 2014] can be drawn with *Mathematica*.

We show the trajectory of the system (black-dotted line) in two distinct phase portraits each corresponding to a different tissue. For a given phase portrait, specific stable states can represent a given pattern: for example, in the current [D_1, D_2] portrait, a stable state in the upper corner represents lateral induction as it corresponds to high expression level for the same species in two consecutive cells. Likewise, in the same portrait, a [high, low] or [low, high] stable state would represent lateral inhibition (Fig. 7.8C).

We compare the phase portraits under distinct tissue-context and observe that the tissue signal radically changes the phase portrait by changing the number and stability of stable states. For example, for hybrid C, in the first tissue the system is attracted to a [high D_1, low D_2] inhibition state, while in the second tissue it switches to a [high D_1, high D_2] induction one (Fig. 7.8C).

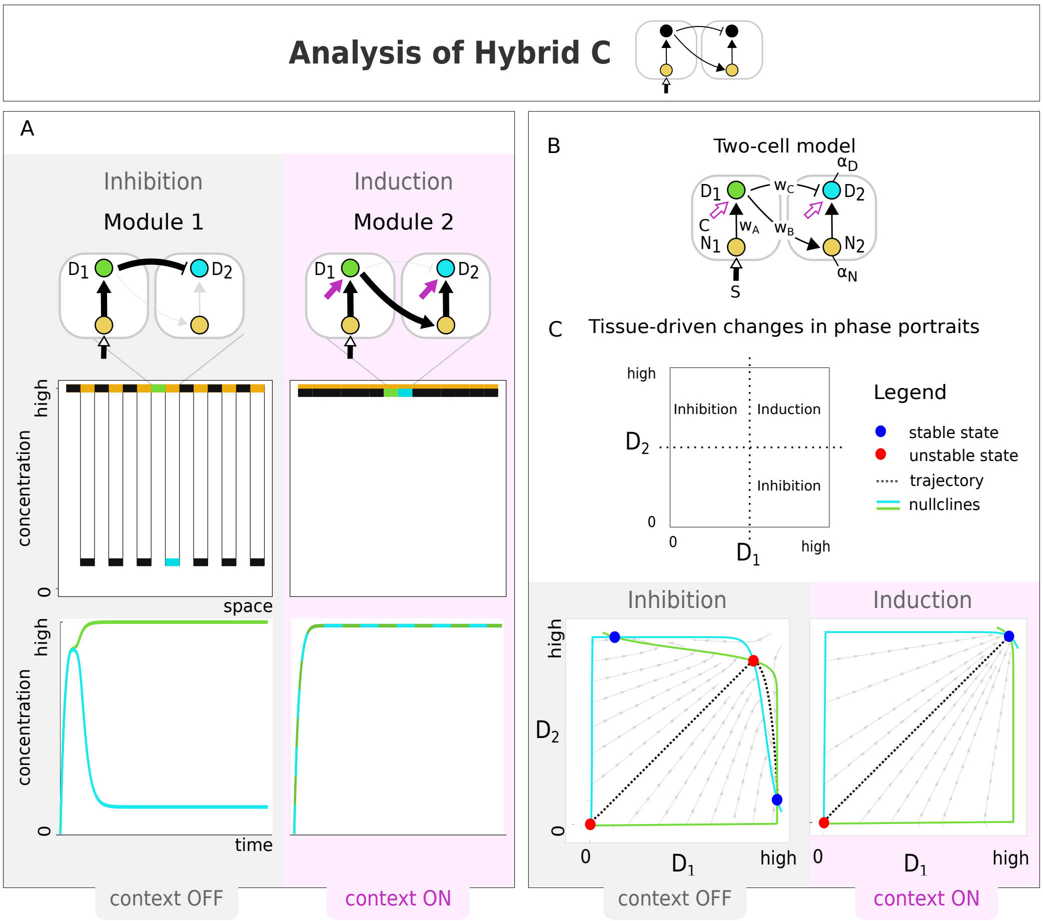


Figure 7.8: **Underlying mechanisms of hybrid circuits.** We model a hybrid C circuit with parameters $w_A = 0.41$, $w_B = 5.49$, $w_C = -0.30$, $\alpha_N = 6.93$ and $\alpha_D = 12.79$. (A) The tissue signal balances the weight of each module or sub-circuit in order to achieve each pattern. In the absence of the signal –tissue A, the first sub-circuit – core H1– drives inhibition: concentrations of delta in two consecutive cells (D_1, D_2) diverge to reach high and low alternating states. By contrast, in the presence of the signal –tissue B, the circuit switches to the second module -core D3- to produce induction: both cells show high expression levels for delta. (B-C) Changes in the phase portrait of the system upon distinct contexts. Each dimension in the phase portraits represents a species (D_1 and D_2). Attractors represent distinct possible final patterns and the trajectory of the system -black dotted line- traces how $D_1(t)$ and $D_2(t)$ evolve in time. The tissue context changes the number and stability of steady states. The trajectory is modified to reach inhibition or induction attractors.

7.3.2 Underlying mechanisms of emergent circuits

Emergent Activation-Inhibition circuits

We chose to study the underlying dynamics of emergent AI_1 as representative for Activation-Inhibition circuits (Fig. 7.9). In this circuit, gene Delta switches from induction to inhibition as the tissue context changes. Ideally, we would like to plot [D_1, D_2] phase portraits for each tissue context. Hence, we write the equations of the system:

$$\frac{dN_1}{dt} = \frac{1}{1 + exp^{\,\alpha_N - 10\,(S + w_D N_1 + w_B D_1)}} - \lambda N_1 \tag{7.9}$$

$$\frac{dD_1}{dt} = \frac{1}{1 + exp^{\,\alpha_D - 10\,(C + w_C N_1 + w_A D_2)}} - \lambda D_1 \tag{7.10}$$

$$\frac{dN_2}{dt} = \frac{1}{1 + exp^{\,\alpha_N - 10\,(w_D N_2 + w_B D_2)}} - \lambda N_2 \tag{7.11}$$

$$\frac{dD_2}{dt} = \frac{1}{1 + exp^{\,\alpha_D - 10\,(C + w_C N_2 + w_A D_1)}} - \lambda D_2 \tag{7.12}$$

We can once more numerically solve the system and follow its trajectory –$N_1(t)$, $D_1(t)$, $N_2(t)$ and $D_2(t)$. However, this complex system cannot be reduced to a two-dimensional [D_1, D_2] system. Instead, we can benefit from studying a [N_2, D_2] system. Independently of the pattern achieved –induction or inhibition– the central cell 1 reaches high-expression levels for both of its genes N_1 and D_1. We thus focus on cell 2 to explain the switch in pattern. Indeed, cell 2 solely receives a single inter-cellular input from D_1. In order to reduce the system to a two-dimensional [N_2, D_2] system, we treat the input from D_1 as a fixed external signal Ext:

$$\frac{dN_2}{dt} = \frac{1}{1 + exp^{\,\alpha_N - 10\,(w_D N_2 + w_B D_2)}} - \lambda N_2 \tag{7.13}$$

$$\frac{dD_2}{dt} = \frac{1}{1 + exp^{\,\alpha_D - 10\,(C + w_C N_2 + w_A Ext)}} - \lambda D_2 \tag{7.14}$$

Next, we can draw a series of instantaneous [N_2, D_2] phase portraits at particular time steps (t_1 to t_{final}) where, for each time step, D_1 holds a fixed Ext value (Fig. 7.9).

We observe that the dynamics of emergent circuits appear more complex than that of hybrid circuits (Fig. 7.8-7.9). Indeed, while in hybrid dynamics stable states are maintained fixed in a given tissue environment, in emergent dynamics we observe movements of attractors leading to complex trajectories. Here we briefly summarize how lateral induction and lateral induction are achieved by AI_1 (Fig. 7.9A). Depending on the tissue context, the circuit implements distinct dynamical behaviours termed captures and pursuits.

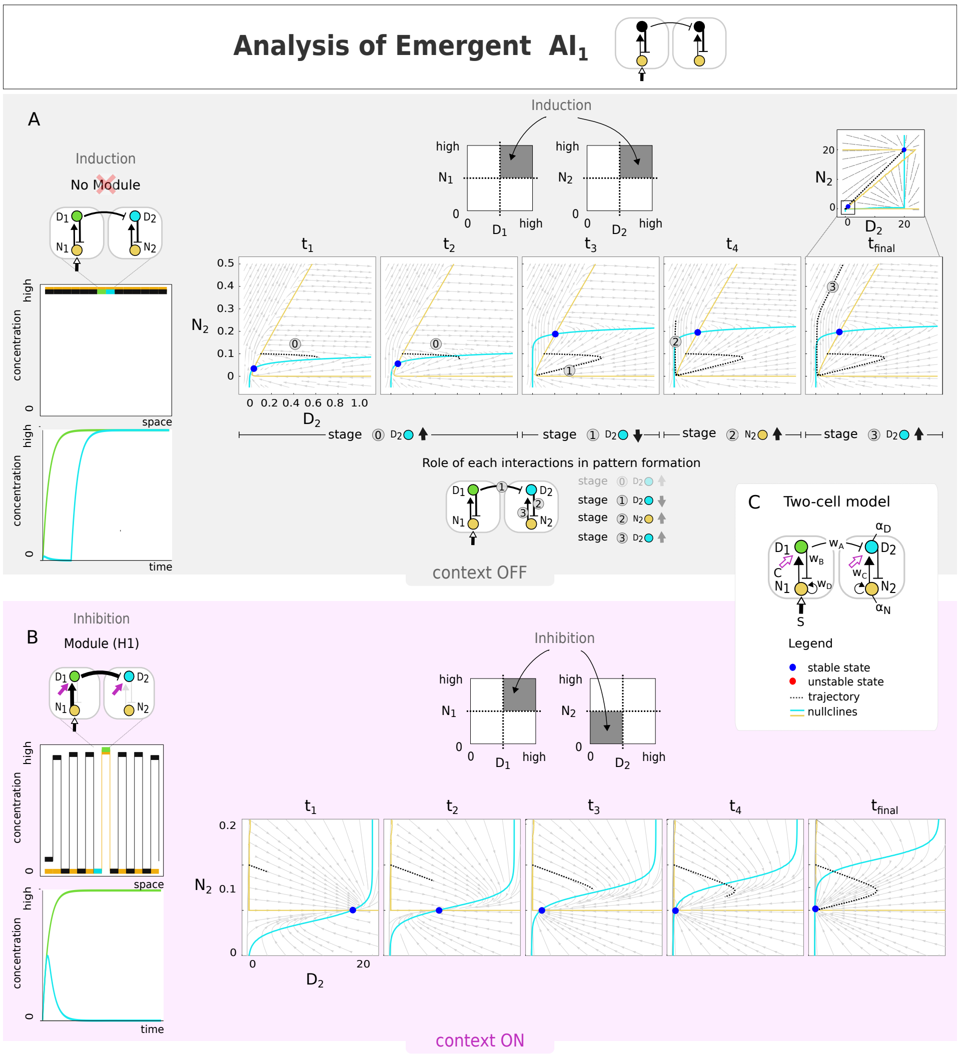


Figure 7.9: **Underlying mechanisms of emergent Activation-Inhibition circuits.** We model a particular AI_1 circuit with parameters $w_A = -0.05$, $w_B = -7.98$, $w_C = 6.47$, $w_D = 9.61$, $\alpha_N = 6.40$ and $\alpha_D = 6.81$. We draw a series of instantaneous [N_2-D_2] phase portraits. At different time points (t_1-t_final) D_1 is treated as and external fixed value.(A) During this capture event, the vertical movement of the attractor causes the trajectory to be deviated and finally reach a [high N_2-high D_2] induction state. Each interaction is responsible for a sudden change in the trajectory (1) D_1 inter-cellular inhibition on D_2 causes D_2 to decrease (2) accordingly, D_2 inhibition on N_2 is weakened and N_2 increases (3) therefore, N_2 activation on D_2 is powered, as a result D_2 increases. (B) During this pursuit event, the trajectory is deviated towards the moving attractor. The moving attractor is finally reached at a [low N_2-low D_2] state that corresponds to lateral inhibition.

First, lateral induction pattern results from a capture behaviour. A capture event occurs when the trajectory at its beginning and its end is attracted towards different attractors. In Figure 7.9A we show how the movement of a particular attractor causes the trajectory to be overtaken by a moving separatrice and then recruited into a new basin of attraction. As such, the trajectory is deviated and finally reaches a distinct attractor in the upper corner of the phase portrait. The reached attractor corresponds to a induction pattern as N_2 and D_2 reach high expression levels and the final state corresponds to [highD_1, highD_2]. Importantly, as the trajectory progresses, each individual interaction is responsible for one sudden change in its direction. The contributions of all interactions to form a pattern illustrate the non-decomposable properties of emergent circuits as all interactions –and not only a sub-set- are essential to generate the pattern.

Second, lateral inhibition pattern results from a pursuit behaviour. During a pursuit event, the trajectory is attracted towards a single moving attractor that is finally reached. The movement of the attractor shown in Figure 7.9B alters the direction of the trajectory. The system finally reaches [lowN_2, lowD_2] state which, considering the states of both cells, represents an inhibition pattern [highD_1, lowD_2].

Summarizing, following complex dynamics, emergent circuit AI_1 switches pattern by transitioning from a capture to a pursuit mechanism. This type of dynamical mechanism are specially relevant in biology. Remarkably, there exist specific biological examples where pattern formation is shaped by pursuit mechanisms, such as the transient domain shifts observed in gap genes in *Drosophila* [Manu et al., 2009a, Manu et al., 2009b].

Emergent circuit Pattern Convertor

A different case of an emergent circuit is *Pattern-Convertor*. As a characteristic feature of emergent circuits, Pattern-Convertor is connected to two lower-complexity mono-functional circuits (Fig. 7.10A). Its function-switching mechanism can be understood following a simple toy model (Fig. 7.10B-C). On one hand, the auto-inhibition H0 core circuit found within its architecture, drives an alternating pattern for the signalling gene, independently of the tissue context. However, when the context is present/ON, the amplitude of the pattern –difference in concentration of the black gene in two consecutive cells– increases. On the other hand, the golden gene feeds from the black gene and functions as a convertor.

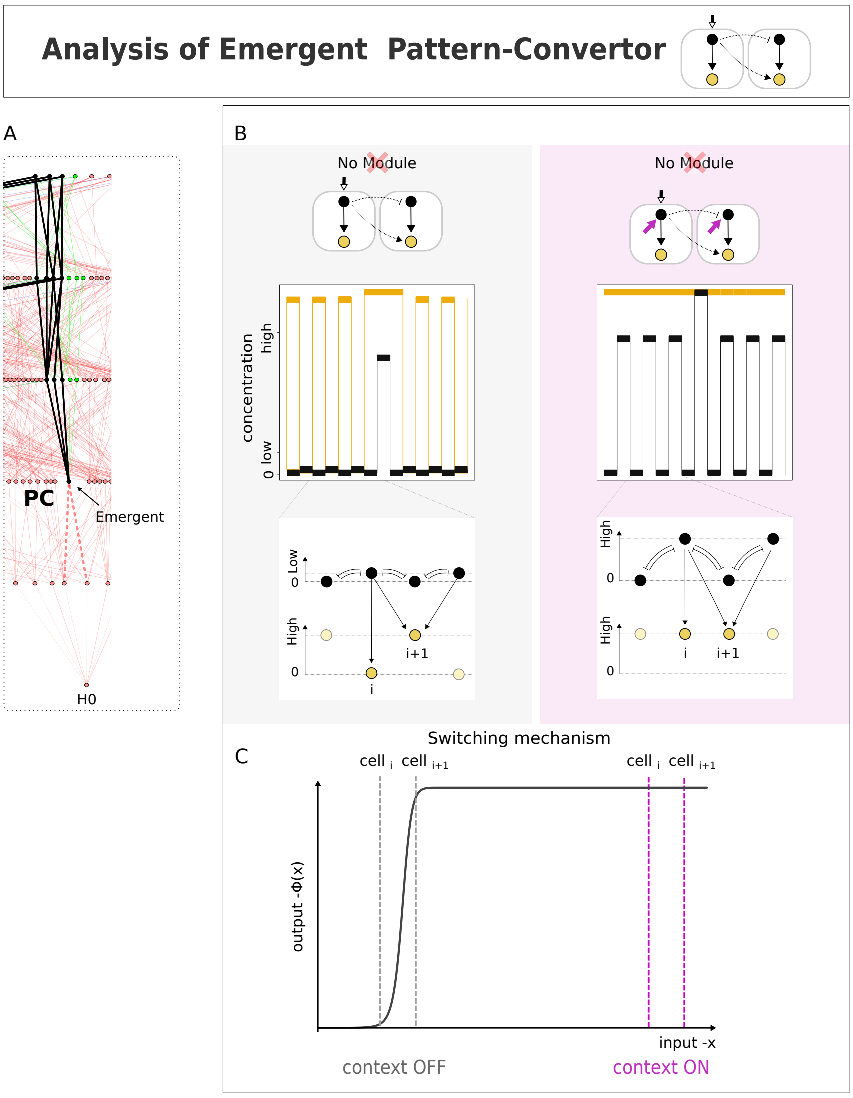

Analysis of Emergent Pattern-Convertor
A
PC
Emergent
H0
B
No Module
No Module
concentration
0 low
high
Low
0
High
0
i+1
i
High
0
High
0
i
i+1
Switching mechanism
C
cell $_i$ cell $_{i+1}$
cell $_i$ cell $_{i+1}$
output -Φ(x)
input -x
context OFF
context ON

Figure 7.10: **Underlying mechanism of emergent Pattern Convertor.** (A) Emergent circuit connected to two lower-complexity mono-functional nodes (extracted from Fig. 7.6). (B) The black gene holds low and high-amplitude alternating patterns depending on the tissue context. The golden gene converts the low-amplitude pattern into an inhibition pattern and the high-amplitude pattern into an induction one. The golden gene reads-out three positive inputs from the black gene: two from neighbouring cells and one from the same cell. (C) We map the inputs received by the golden gene in two consecutive cells in both tissue contexts. The shape of the regulatory function leads to the following switching-mechanism: when the context is OFF the input received by the golden gene in one of the cells is sufficient for activation, while in its neighbour cell it is not (inhibition); instead when context is ON both cells receive sufficient input (induction).

In the first tissue, we observe that the golden gene converts the low-amplitude inhibition pattern into a high-amplitude inhibition pattern. In the second tissue, the golden gene converts a high-amplitude inhibition pattern into an induction pattern. By plotting the inputs received by the golden gene on the regulatory function, we provide intuitive understanding on this process. In the first tissue, when the context is OFF, the input received by the golden gene in one of the cells is sufficient to activate is expression, while in its neighbour cell it is not (leads to inhibition). Instead, when the context is ON, both cells receive sufficient input to reach high expression levels (leads to induction).

Characteristic of emergent circuits, the behaviour of Pattern-Convertor cannot be reduced to that of two separable sub-circuits.

7.4 Continuous versus discontinuous pattern transitions

The tissue context has been interpreted up to now as characterizing two distinct tissues. However, context can also represent different developmental stages of the same tissue. In order to study the transient nature of tissue environment we treat the context as a time-dependent cue. This study is thus particularly relevant for biological systems that are exposed to a changing environment or variable tissue context (e.g. signalling input).

Precisely, in real biological systems induction and inhibition can occur in distinct tissues but also succeed each other in a single tissue: in the retina of *Drosophila* a wave of cellular differentiation known as the *morphogenetic furrow* progresses through the tissue –induction– to later give rise to a fine-grained pattern of R8 photoreceptor cells –inhibition– [Sato et al., 2013]. Likewise, in the chick's inn-ear ear induction first leads to a continuous domain of precursor cells , i.e. patch of prosensory cells. Subsequently, within each patch inhibition mediates differentiation into hair cells and supporting cells [Daudet and Lewis, 2005, Petrovic et al., 2014]. In both cases, lateral induction precedes lateral inhibition. Indeed, lateral induction seems adequate to first establish the progenitor field of cells, i.e. maintaining a given regulatory state in a field of cells, while lateral induction promotes further differentiation in a later step. We believe this succession of events is biologically relevant as the establishment of the progenitor field of cells is one of the first steps of organogenesis.

These two real biological examples suggest that the key molecule driving a change in pattern is a time-dependent cue that gradually accumulates in the tissue. We thus explore how the dynamic nature of tissues directs pattern transitions. Here, instead of using a binary signal –context OFF/0 or ON/1–, we mimic time-dependent cues by gradually increasing the value of the tissue signal from 0 to 1. We find examples of both continuous and discontinuous pattern transitions (Fig. 7.11-7.12). Like we did in section 7.3, we draw phase portraits for Hybrids C and E, this time gradually increasing the context value C.

Hybrid E (Fig. 7.12B) is described by the following equations:

$$\frac{dN_1}{dt} = \frac{1}{1 + exp^{\alpha_N - 5\,(C + w_D D_2)}} - \lambda N_1 \tag{7.15}$$

$$\frac{dD_1}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(S + w_A N_1 + w_B D_1 + w_C D_2)}} - \lambda D_1 \tag{7.16}$$

$$\frac{dN_2}{dt} = \frac{1}{1 + exp^{\alpha_N - 5\,(C + w_D D_1)}} - \lambda N_2 \tag{7.17}$$

$$\frac{dD_2}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(w_A N_2 + w_B D_2 + w_C D_1)}} - \lambda D_2 \tag{7.18}$$

Analogous to 7.3.1, we impose equilibrium for Notch genes in equations (7.15) and (7.17) and obtain a two dimensional reduced system:

$$\frac{dD_1(D_1, D_2)}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(S + \frac{w_A}{\lambda\,(1 + exp^{\alpha_N - 5\,(C + w_D D_2)})} + w_B D_1 + w_C D_2)}} - \lambda D_1 \tag{7.19}$$

$$\frac{dD_2(D_1, D_2)}{dt} = \frac{1}{1 + exp^{\alpha_D - 5\,(\frac{w_A}{\lambda\,(1 + exp^{\alpha_N - 5\,(C + w_D D_1)})} + w_B D_2 + w_C D_1)}} - \lambda D_2 \tag{7.20}$$

Both hybrids (Fig. 7.11-7.12) show a type of bifurcation –pichfork– in which attractors tend to appear and disappear in symmetrical pairs. As such, the context behaves as a bifurcation parameter leading two different types of bifurcation. The continuity of pattern transition depends on the type of bifurcation.

For hybrid C, context directs a supercritical pitchfork bifurcation where two stable states move towards each other, later collide and mutually annihilate to create a new stable state (Fig. 7.11). Because of the shift between attractors is continuous, the pattern also transitions continuously. We observe how from an initial [highD_1, lowD_2] inhibition state, the systems follows the moving attractor as D_1 is kept high and D_2 increases to transition to [highD_1, highD_2] inductive state. This way, an inhibition pattern gradually reduces its amplitude to continuously transition to induction.

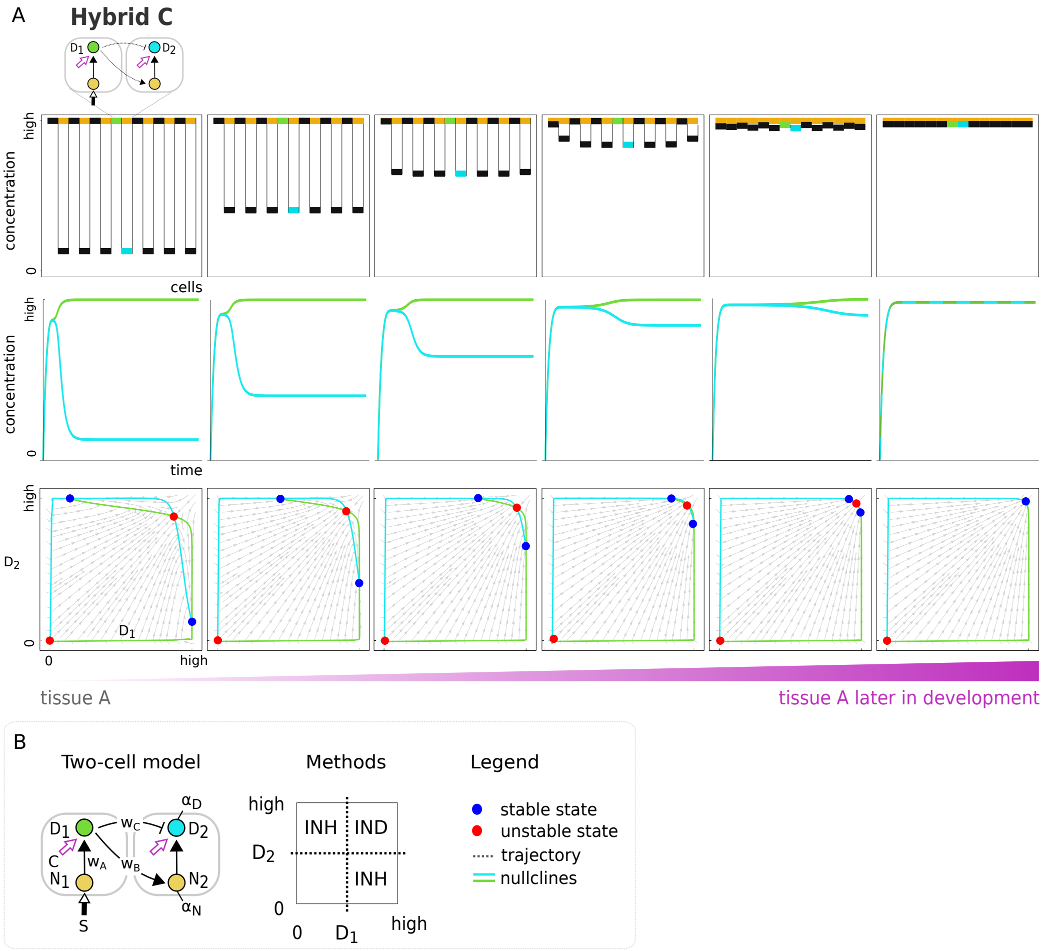


Figure 7.11: **Continuous pattern transitions.** For this figure and figure 7.12, at different context values, three panel rows show the final multicellular pattern, concentrations $D_1(t)$ and $D_2(t)$ and the corresponding phase portraits, respectively. We use different values of C $\{0; 0,2; 0,5; 0,7; 0,9; 1\}$ that are replaced in the systems of equations (7.7-7.8) and (7.19-7.20). We observe how $D_1(t)$ and $D_2(t)$ evolve as C increases. Hybrid C shows a type of bifurcation -pitchfork- in which attractors tend to appear and disappear in symmetrical pairs: the context parameter C directs a supercritical pitchfork bifurcation where two stables states move towards each other, later collide and mutually annihilate to create a new stable state. Because the shift between attractors is continuous, the pattern transitions continuously. From an initial [high D_1 -low D_2] inhibition state, the system follows the moving attractor as D_1 is kept high and D_2 increases to transition to [high D_1 -high D_2] inductive state. This way, an inhibition pattern gradually reduces its amplitude to continuously transition to induction.

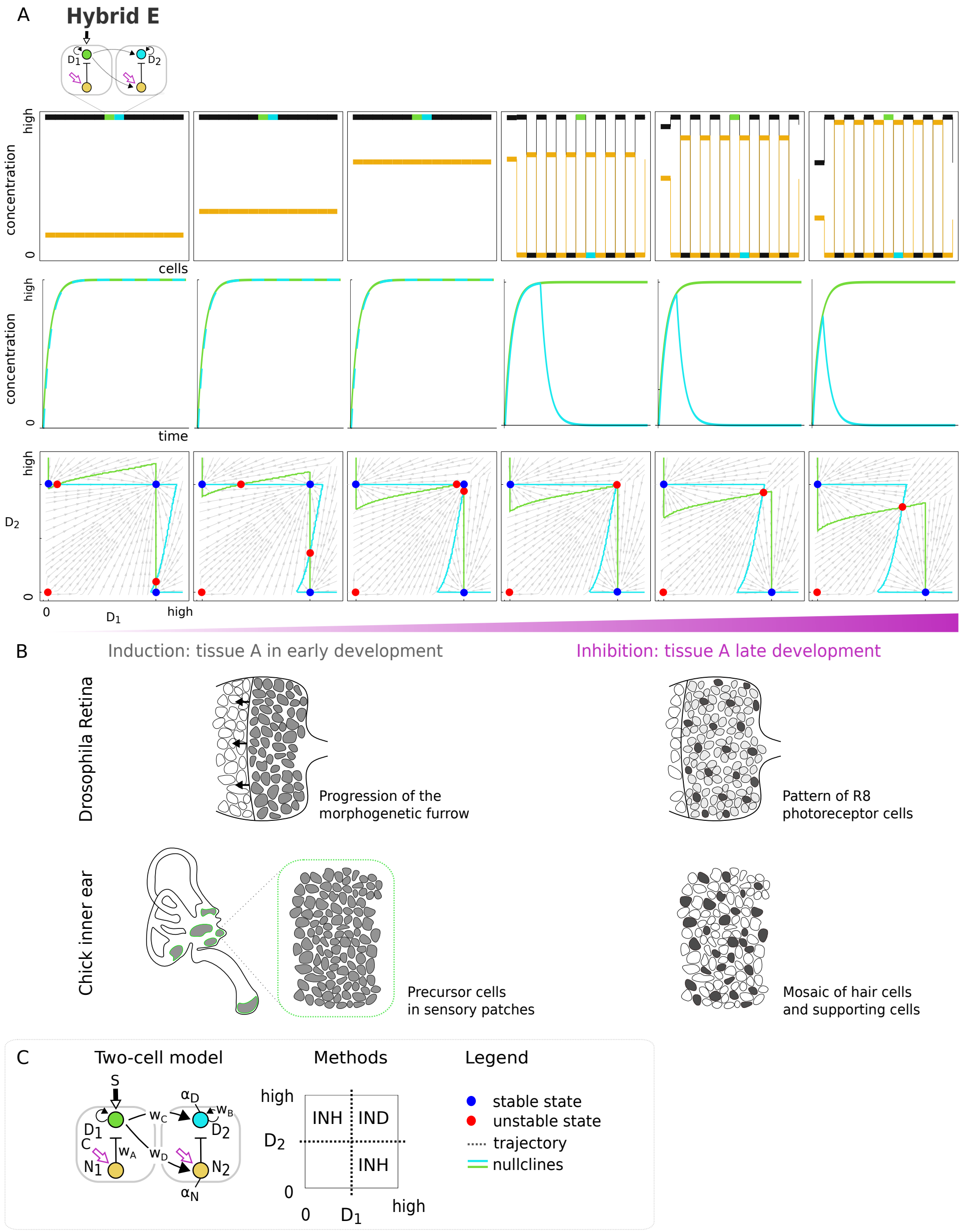


Figure 7.12: **Discontinuous pattern transitions.** Circuit with parameters $w_A = -4.93$, $w_B = 3.22$, $w_C = 0.43$, $w_D = 0.14$, $\alpha_N = 15.86$ and $\alpha_D = 8.62$. The system undergoes a subcritical pitchfork bifurcation: two unstable states coalesce into a new fixed point that changes its stability from stable to unstable, causing an abrupt and discontinuous pattern transition. This model mimics known biological processes where lateral induction precedes lateral inhibition in the same tissue such as in the *Drosophila* retina [Sato et al., 2013] or the Chick inner ear [Daudet and Lewis, 2005, Petrovic et al., 2014].

Instead, for hybrid E, the system undergoes a subcritical pitchfork bifurcation: two unstable states coalesce into a new fixed point that changes its stability from stable to unstable (Fig. 7.12). This change of stability makes the system –initially at [highD_1, highD_2] induction state– to suddenly switch to an attractor in a different position in the phase portrait –[highD_1, lowD_2] inhibition state–. This sudden switch leads to an abrupt and discontinuous pattern transition. Interestingly, hybrid E mimics the previously mentioned biological processes where lateral induction precedes lateral inhibition. This alternative scenario suggests that multi-functional circuits can achieve each of their function guided by time-dependent signals.

7.5 Conservation of design principles in higher-order circuits

The above analyses focused on minimal two-node circuits and identified simple circuits that are sufficient to achieve two distinct biological functions. But are the designs found universal, i.e. unique? In other words, are the identified 13 minimal circuits the foundation of all possible bi-functional -induction/inhibition- circuits, or are there more complex higher-order solutions that do not contain these minimal circuits? More importantly, are the design principles of hybrid and emergent circuits conserved among higher-order circuits? To explore this question we expanded our study and consider all possible three-node circuits. Again, for each circuit topology we sampled large numbers of parameter sets. Because the number of topological designs increases from 1200 two-gene topological designs to 731,250 possible three-gene circuits, the computational cost of the simulation explodes and we had to sample fewer parameter sets (from the 10^7 previously sampled for two-gene circuits to 30.000 for three-gene circuits). Due to the high number of solutions (32,956 topologies), we show a truncated complexity atlas that includes only low-complexity multi-functional circuits. The atlas shows dozens of stalactites (Fig. 7.13A-B) –core circuits. For all of them, we compared their topology to that of the basic multi-functional two-gene circuits. Strikingly, all topologies contain at least one of the 13 minimal circuits or hold equivalent forms. Within these equivalences, the third gene performs a variety of roles (Fig. 7.13C): from simple roles such as a downstream read-out or an upstream moderator, to more complex roles such as intermediate positions or the specific transition from a 2-gene negative feedback to a 3-gene repressilator [Elowitz and Leibler, 2000].

These results indicate that hybrid and emergent designs are conserved in high-dimensional circuits and conform the two prevalent forms of information compression. Interestingly, the conservation of design principles as the dimensionality of a system increases has already been observed in a number of studies, for example, in circuits achieving biochemical adaptation. Indeed, as Ma et al. explored gene circuits performing adaptation [Ma et al., 2009] –a function observed in many sensory systems that consists on the ability of that system to transiently respond after input stimulus then reset itself back to its original steady-state output level–, they observed that higher-order circuits contained within their topology lower-order core circuits. These results highlight the value of minimal circuits in understanding the designs principles underlying biological processes.

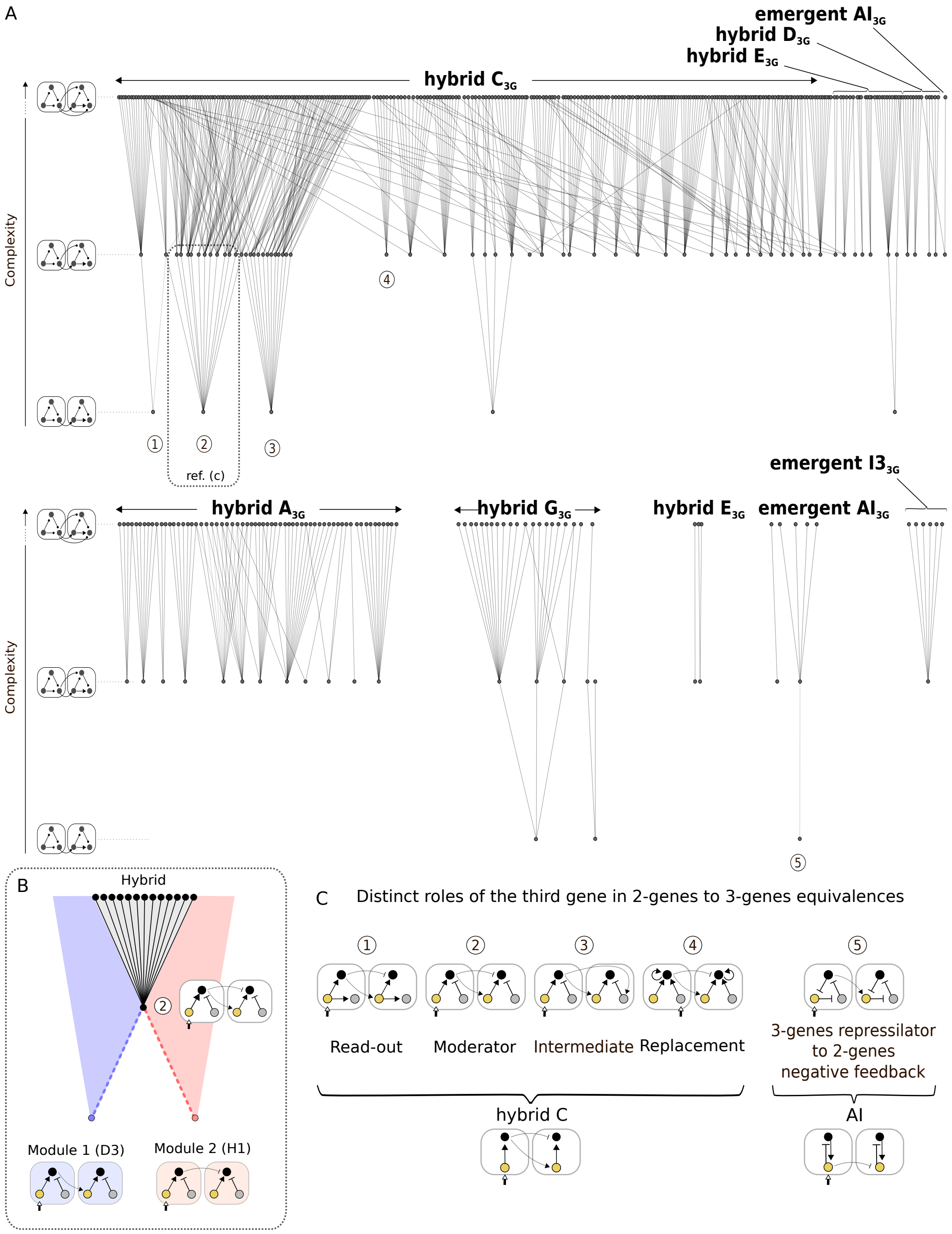

A
emergent AI$_{3G}$
hybrid D$_{3G}$
hybrid E$_{3G}$
hybrid C$_{3G}$
Complexity
①
②
③
④
ref. (c)
emergent I3$_{3G}$
hybrid A$_{3G}$
hybrid G$_{3G}$
hybrid E$_{3G}$
emergent AI$_{3G}$
Complexity
⑤
B
Hybrid
②
Module 1 (D3)
Module 2 (H1)
C
Distinct roles of the third gene in 2-genes to 3-genes equivalences
①
②
③
④
⑤
Read-out
Moderator
Intermediate
Replacement
3-genes repressilator
to 2-genes
negative feedback
hybrid C
AI

Figure 7.13: **Universality of multi-functional designs.** (A-B) Complexity atlas of three-gene bi-functional circuits. Atlas truncated to show only low complexity circuits -3 to 5 interactions. Bi-functional 3-gene circuits are grouped according to their equivalence with Hybrid or Emergent two-node circuits. (C) The role of the extra gene can be simple (downstream read-out or upstream moderator) or more complex (intermediate role, role replacement, equivalence between a 2-gene negative feedback and a 3-gene repressilator).

Chapter 8

DISCUSSION

It is not obvious how a circuit could be designed to perform 2 distinct functions such that both functions depend on all genes in the circuit. Here we have taken inspiration from the Notch-Delta signalling system to create a purely theoretical model to explore this question. The goal of this model is not to understand the details of Notch-Delta signalling in a real biological system (which is another on-going area of research [Collier et al., 1996, Formosa-Jordan and Ibanes, 2009, Petrovic et al., 2014, Palau-Ortin et al., 2015]), but instead to provide an abstract but biologically relevant model of bi-functionality. Using this model we have been able to demonstrate concrete cases of bi-functionality where the same minimal circuit of genes is able to perform two distinct patterning functions without any modulation of its regulatory parameters. The switch in behaviour is triggered just by changing the basal expression level of one of the two genes. In a real system, this could be accomplished by differential expression of a transcription factor in a tissue-specific or time-specific manner.

Gene circuits are believed to be decomposable into structural modules. Due to the advantages structural modularity provides, there exist a bias towards picturing networks as composed of such non-overlapping structures [Solé and Valverde, 2008, Clune et al., 2013, Ellefsen et al., 2015, Kashtan and Alon, 2005, Kashtan et al., 2009] (Fig. 5.2A). However, observations such as the prevalence of pleiotropy in development force us to revisit the concept of structural modularity. For that, we performed a theoretical search to reveal minimal designs where all components are essential to perform two distinct functions. These minimal designs stand as concrete examples of complete circuit overlap (Fig. 5.2C). Revealing such compact multi-functional circuits has the following three consequences for our understanding of modularity.

First, by finding such compact circuits, we reveal the limits of the concept of structural modularity, and strengthen the need to consider instead functional modularity [Kholodenko et al., 2002, Eisen et al., 1998, Segal et al., 2003, Ten Tusscher and Hogeweg, 2011, Irons and Monk, 2007, Alexander et al., 2009]. We propose to change our notion of modularity –rather than being structurally distinct, the modules can now only be defined by which set of gene interactions are responsible for which function, thus transitioning from 'set of genes'to 'set of interactions'. In that sense, among multi-functional designs, hybrid circuits are indeed modular. They can be decomposed into distinct sub-circuits that underlie function where each module consist on a distinct gene interaction path.

Second, the compact structure of multi-functional circuits has an impact on motif detection. Indeed, a particular kind of study detects network motifs within large transcriptional regulatory networks [Shen-Orr et al., 2002, Milo et al., 2002, Alon, 2007]. Networks motifs are detected as patterns that occurred much more often than would be expected in random networks [Alon, 2007]. Once detected, the cellular function a motif performs is extrapolated from its topology. Motifs correspond to particular groups of interacting genes. In that sense, they are analogous to a structural module. However, the lack of structural modularity of our compact circuits will make them very hard to detect by motif-finding methods. Instead, methods based on the dynamic properties of networks [Irons and Monk, 2007, Alexander et al., 2009] will be more adequate to detect multi-functional circuits.

Third, the compact structure of multi-functional circuits has consequences for evolution: while structural modularity could explain the ability of two traits to evolve autonomously, the existence of compact overlapping modules could instead account for their covariation [Wagner and Altenberg, 1996]. We believe that the real structure of gene circuits combines this two distinct scenarios.

In addition to revealing concrete examples of tightly overlapping bi-functional circuits, we have also found two distinct design principles. In the first case, the circuit directly combines 2 mono-functional circuits. These *hybrid* designs are intuitively understandable from the functionality of their two sub-modules, and in the atlas they can always be seen as the point where two mono-functional stalactites unite (Fig. 7.7). In the second case, bi-functionality is *emergent* because the circuit cannot be deconstructed into 2 different sub-circuits – at least one of the functions is dependent on every regulatory interaction within the circuit. The location of these circuits within the complexity atlas reveals how this second design type always emerges from within a single mono-functional stalactite. The design of these circuits can be considered by starting with a minimal mono-functional circuit at the base of a stalactite, and then adding specific extra regulatory links such that a new non-reducible mechanism is created for the second function (Fig. 7.7).

Interestingly, there is no clear difference between the two design classes in terms of their robustness to parameter variations, and thus we suggest they are equally plausible to occur in real biological systems. Indeed, although modular solutions are more widely discussed than non-modular ones, our unbiased search finds that both designs are as efficient. Non-modular solutions are rarely discussed, as their integrated and complex wiring makes them non-intuitive to understand. However, the low-dimensionality of our model (only two genes) allows phase portraits to be visualized providing an intuitive understanding of their dynamical mechanisms.

We believe that the approach of mapping out landscapes of dynamical mechanisms, using tools such as the complexity atlas, will remain essential to our engineering attempt to design new circuits synthetically [Matsuda et al., 2012, Matsuda et al., 2015, Schaerli et al., 2014]. In that sense, the finding of bi-functional designs is of relevant interest in synthetic biology. Already, distinct circuits have been successfully engineered that propagate a signal throughout a cell population under induction [Matsuda et al., 2012] or inhibition [Matsuda et al., 2015] modes, where D2 and H2 core circuits have respectively been implemented. Despite the many technical challenges, the simple designs found here can potentially be synthetically built as pattern-switch circuits able to achieve distinct patterns upon a tunable input.

Part IV

Global Discussion

Both projects of this thesis are related to the concept of developmental constraints, a concept first defined in the 1980s by Maynard-Smith as the ‘biases on the production of variant phenotypes ... caused by the structure, character, composition, or dynamics of the developmental system ’[Maynard Smith et al., 1985]. While our first project focuses on dynamic constraints, the second does consider structural ones.

The results of our first project show that the dynamics of gene circuits itself constitutes a developmental constraint, as it restricts the phenotypic novelties available from a particular genotype. Indeed, from the broad perspective of developmental biology, the concept of evolvability [Pigliucci, 2008] focuses on the generation of novel phenotypes (i.e. a potential), while the concept of developmental constraints [Maynard Smith et al., 1985] refers to restrictions on the production of certain phenotypes (i.e. a limitation). Independent of the context, both evolvability and developmental constraints describe the available novelties.

In order to better understand the concept of developmental constraints, we consider the following analogy: many human inventions can achieve the function of flight: aeroplanes, hot-air balloons, helicopters and other curious creations (Fig. 8.1). From a functional perspective, all inventions achieve the same phenotype –transporting humans by flight– in a distinct way, i.e. using a different mechanism. If we were to tinker with these designs, or ‘mutate ’them, to invent a new mode of transport, each of them would result in a distinct type of vehicle. Intuitively, the transition from Leonardo da Vinci ’s flying machine to a space craft is constrained while that between an aeroplane and a space craft is easier to achieve. The same principle applies to gene regulatory circuits in biological systems. In order to generate a given gene expression pattern a gene circuit can use few mechanisms which correspond to distinct dynamical trajectories. We observe that the dynamical mechanisms of a circuit determines which phenotypes are easier or harder to reach. We have provided a theoretical proof that indeed the dynamics of a gene circuit constitute a specific type of developmental constraint on the search for new phenotypes.

Regarding the second project, in order to understand the capacity of gene circuits to perform multiple function we have imposed the following constraint: gene circuits need to encode a high number of functions with a minimal set of genes, providing a form of information compression. Most previous studies on multi-functionality consider large gene circuits [Kashtan and Alon, 2005, Kashtan et al., 2009, Clune et al., 2013, Ellefsen et al., 2015, Espinosa-Soto and Wagner, 2010, Ten Tusscher and Hogeweg, 2011]. The dimension of those circuits allows them to split into independent mono-functional circuits where different sets

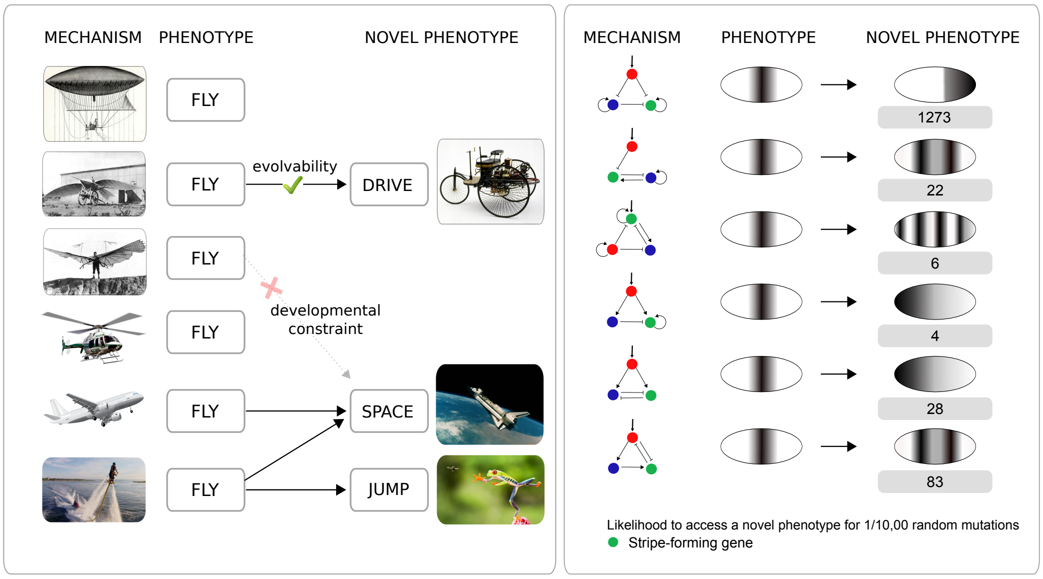

MECHANISM
PHENOTYPE
NOVEL PHENOTYPE
FLY
FLY
evolvability
DRIVE
FLY
developmental constraint
FLY
FLY
SPACE
FLY
JUMP
MECHANISM
PHENOTYPE
NOVEL PHENOTYPE
1273
22
6
4
28
83
Likelihood to access a novel phenotype for 1/10,00 random mutations
● Stripe-forming gene

Figure 8.1: **Impact on evolvability of alternative mechanisms to produce a phenotype.** Distinct flying devices can achieve the *flying* phenotype. Similarly, distinct gene circuits can produce the *stripe* phenotype. As an analogy, the use of a particular device or a particular dynamical mechanism both impact on the novel phenotypes accessible.

of genes perform distinct tasks or functions. However, the knowledge that the same set of genes generates a diversity of functions imposes a certain degree of functional overlap. The necessity of a structural overlap is a constraint in itself. Hence, the bi-functional designs revealed are found under the limitation/potential for compact encoding.

The search for a multi-functional behaviour constitutes an innovative perspective. Indeed, developmental biologists studying pattern formation often restrict their analysis to a single function. This way, the currently studied developmental gene regulatory circuits are mostly constructed by gathering the minimal number of proteins, genes and molecules sufficient for the formation of a single organ or particular part of an embryo [Raff and Conway Morris, 1996, von Dassow et al., 2000, Davidson et al., 2002, Ma et al., 2006, Alvarez-Buylla et al., 2010]. Instead of studying biological functions independently, we propose to study gene regulatory circuits performing multiple functions, considering the bigger picture and provide intuitive understanding on multi-functionality.

An interesting aspect of both projects, is the use of such small 2 and 3-noded circuits to explore relevant biological questions. On one hand, small circuits show a handful of complex dynamical mechanisms. Their impact on evolvability can be observed even for such an apparently simple patterning task and using very simple criteria to define novel patterns. On the other hand, it is notable how contrasting and qualitatively distinct functions can be encoded within such a small set of genes.

Our finding of compact multi-functional circuits provides and answer to the G-value paradox [Hahn and Wray, 2002]. Indeed, as the number of genes does not increase with organismal complexity, compact circuits provide a plausible solution. However, a broad approach to the G-value paradox has been to search for further sources of complexity within the genome [Mattick et al., 2010, Schad et al., 2011]– other distinct molecular 'components '– such as modulation of chromatin [Kouzarides, 2007, Cairns, 2009], alternative splicing [Kim et al., 2007], the multi-functionality of proteins [Jeffery, 1999] or the newly discovered regulatory functions of dozens of types of non-protein coding RNAs such as lncRNAs or miRNAs [Sempere et al., 2006, Taft et al., 2007]. In principle, additional elements found within the genome could account for the acquisition of additional biological functions – the complexity of an organism might be in some way 'proportional 'to the list of distinct molecular components in the genome. Instead, we have proposed and alternative scenario where organismal complexity does not directly follow from molecular complexity, but instead arises as the emergent behaviour of dynamic regulatory circuits [Kauffman, 1993b, Solé and Fernández, Pau and Kauffman, Stuart A,]. Rather on focusing on the number and various details of molecular components, we give a concrete theoretical example of tightly compact encoding of biological information – multiple distinct functions encoded within a small set of genes – suggesting that organismal complexity shows a complex and non-trivial relationship to genomic complexity. Hence, we believe that circuit designs can show interesting and complex behaviours that help us understand biological processes. Furthermore, we believe that the approach of mapping out genotype-phenotype landscapes using tools such as the complexity atlas, will remain essential to reveal the function of biological circuits.

Bibliography

[Aguirre et al., 2011] Aguirre, J., Buldú, J. M., Stich, M., and Manrubia, S. C. (2011). Topological structure of the space of phenotypes: the case of rna neutral networks. *PLoS One*, 6(10):e26324.

[Ahn and Joyner, 2004] Ahn, S. and Joyner, A. L. (2004). Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell*, 118(4):505–516.

[Alberch, 1991] Alberch, P. (1991). From genes to phenotype: dynamical systems and evolvability. *Genetica*, 84(1):5–11.

[Aldana et al., 2007] Aldana, M., Balleza, E., Kauffman, S., and Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *J Theor Biol*, 245(3):433–448.

[Alexander et al., 2009] Alexander, R. P., Kim, P. M., Emonet, T., and Gerstein, M. B. (2009). Understanding modularity in molecular networks requires dynamics. *Sci Signal*, 2(81):pe44.

[Alon, 2006] Alon, U. (2006). *An introduction to systems biology: design principles of biological circuits*. CRC press.

[Alon, 2007] Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8(6):450–461.

[Alvarez-Buylla et al., 2010] Alvarez-Buylla, E. R., Azpeitia, E., Barrio, R., Benítez, M., and Padilla-Longoria, P. (2010). From abc genes to regulatory networks, epigenetic landscapes and flower morphogenesis: making biological sense of theoretical approaches. *Semin Cell Dev Biol*, 21(1):108–117.

[Artavanis-Tsakonas et al., 1995] Artavanis-Tsakonas, S., Matsuno, K., and Fortini, M. E. (1995). Notch signaling. *Science*, 268(5208):225–232.

[Ashe and Briscoe, 2006] Ashe, H. L. and Briscoe, J. (2006). The interpretation of morphogen gradients. *Development*, 133(3):385–394.

[Balaskas et al., 2012] Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K. M., Briscoe, J., and Ribes, V. (2012). Gene regulatory logic for reading the sonic hedgehog signaling gradient in the vertebrate neural tube. *Cell*, 148(1-2):273–284.

[Bandyopadhyay et al., 2010] Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.-K., Chuang, R., Jaehnig, E. J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M., Fiedler, D., Dutkowski, J., Guénolé, A., van Attikum, H., Shokat, K. M., Kolodner, R. D., Huh, W.-K., Aebersold, R., Keogh, M.-C., Krogan, N. J., and Ideker, T. (2010). Rewiring of genetic networks in response to dna damage. *Science*, 330(6009):1385–1389.

[Barad et al., 2011] Barad, O., Hornstein, E., and Barkai, N. (2011). Robust selection of sensory organ precursors by the notch-delta pathway. *Curr Opin Cell Biol*, 23(6):663–667.

[Becskei et al., 2001] Becskei, A., Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J*, 20(10):2528–2535.

[Ben-Tabou de-Leon and Davidson, 2009] Ben-Tabou de-Leon, S. and Davidson, E. H. (2009). Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Dev Biol*, 325(2):317–328.

[Berg and Brown, 1972] Berg, H. C. and Brown, D. A. (1972). Chemotaxis in escherichia coli analysed by three-dimensional tracking. *Nature*, 239(5374):500–504.

[Brandon, 1999] Brandon, R. N. (1999). The units of selection revisited: the modules of selection. *Biology and Philosophy*, 14(2):167–180.

[Bullinaria, 2007] Bullinaria, J. A. (2007). Understanding the emergence of modularity in neural systems. *Cogn Sci*, 31(4):673–695.

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

[Cairns, 2009] Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, 461(7261):193–198.

[Carroll, 2008] Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1):25–36.

[Carroll et al., 2013] Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. (2013). *From DNA to diversity: molecular genetics and the evolution of animal design*. John Wiley & Sons.

[Chau et al., 2012] Chau, A. H., Walter, J. M., Gerardin, J., Tang, C., and Lim, W. A. (2012). Designing synthetic regulatory networks capable of self-organizing cell polarization. *Cell*, 151(2):320–332.

[Ciliberti et al., 2007] Ciliberti, S., Martin, O. C., and Wagner, A. (2007). Innovation and robustness in complex regulatory gene networks. *Proc Natl Acad Sci U S A*, 104(34):13591–13596.

[Clune et al., 2013] Clune, J., Mouret, J.-B., and Lipson, H. (2013). The evolutionary origins of modularity. *Proc Biol Sci*, 280(1755):20122863.

[Collier et al., 1996] Collier, J. R., Monk, N. A., Maini, P. K., and Lewis, J. H. (1996). Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J Theor Biol*, 183(4):429–446.

[Cotterell et al., 2015] Cotterell, J., Robert-Moreno, A., and Sharpe, J. (2015). A local, self-organizing reaction-diffusion model can explain somite patterning in embryos. *Cell Systems*, 1(4):257–269.

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

[Crick, 1970] Crick, F. (1970). Diffusion in embryogenesis. *Nature*, 225(5231):420–422.

[Daudet and Lewis, 2005] Daudet, N. and Lewis, J. (2005). Two contrasting roles for notch activity in chick inner ear development: specification of prosensory patches and lateral inhibition of hair-cell differentiation. *Development*, 132(3):541–551.

[Davidson, 2010] Davidson, E. H. (2010). Emerging properties of animal gene regulatory networks. *Nature*, 468(7326):911–920.

[Davidson and Erwin, 2006] Davidson, E. H. and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science*, 311(5762):796–800.

[Davidson et al., 2002] Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust,

A. G., Pan, Z. j., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. (2002). A genomic regulatory network for development. *Science*, 295(5560):1669–1678.

[de Celis and Bray, 1997] de Celis, J. F. and Bray, S. (1997). Feed-back mechanisms affecting notch activation at the dorsoventral boundary in the drosophila wing. *Development*, 124(17):3241–3251.

[de Celis et al., 1997] de Celis, J. F., Bray, S., and Garcia-Bellido, A. (1997). Notch signalling regulates veinlet expression and establishes boundaries between veins and interveins in the drosophila wing. *Development*, 124(10):1919–1928.

[Dequéant et al., 2006] Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquié, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, 314(5805):1595–1598.

[Di Ferdinando et al., 2001] Di Ferdinando, A., Calabretta, R., and Parisi, D. (2001). Evolving modular architectures for neural networks. In *Connectionist models of learning, development and evolution*, pages 253–262. Springer.

[Dichtel-Danjoy and Félix, 2004] Dichtel-Danjoy, M.-L. and Félix, M.-A. (2004). Phenotypic neighborhood and micro-evolvability. *Trends Genet*, 20(5):268–276.

[Dietrich and Hiiragi, 2007] Dietrich, J.-E. and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development*, 134(23):4219–4231.

[Draghi et al., 2010] Draghi, J. A., Parsons, T. L., Wagner, G. P., and Plotkin, J. B. (2010). Mutational robustness can facilitate adaptation. *Nature*, 463(7279):353–355.

[Economou et al., 2012] Economou, A. D., Ohazama, A., Porntaveetus, T., Sharpe, P. T., Kondo, S., Basson, M. A., Gritli-Linde, A., Cobourne, M. T., and Green, J. B. A. (2012). Periodic stripe formation by a turing mechanism operating at growth zones in the mammalian palate. *Nat Genet*, 44(3):348–351.

[Eisen et al., 1998] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25):14863–14868.

[El-Samad et al., 2002] El-Samad, H., Goff, J. P., and Khammash, M. (2002). Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective. *J Theor Biol*, 214(1):17–29.

[Eldar and Elowitz, 2010] Eldar, A. and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173.

[Ellefsen et al., 2015] Ellefsen, K. O., Mouret, J.-B., and Clune, J. (2015). Neural modularity helps organisms evolve to learn new skills without forgetting old skills. *PLoS Comput Biol*, 11(4):e1004128.

[Elowitz and Leibler, 2000] Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338.

[Espinosa-Soto and Wagner, 2010] Espinosa-Soto, C. and Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS Comput Biol*, 6(3):e1000719.

[Esser et al., 2006] Esser, A. T., Smith, K. C., Weaver, J. C., and Levin, M. (2006). Mathematical model of morphogen electrophoresis through gap junctions. *Dev Dyn*, 235(8):2144–2159.

[Ferrada and Wagner, 2010] Ferrada, E. and Wagner, A. (2010). Evolutionary innovations and the organization of protein functions in genotype space. *PLoS One*, 5(11):e14172.

[Fontana, 2002] Fontana, W. (2002). Modelling 'evo-devo' with rna. *Bioessays*, 24(12):1164–1177.

[Formosa-Jordan and Ibanes, 2009] Formosa-Jordan, P. and Ibanes, M. (2009). Diffusible ligand and lateral inhibition dynamics for pattern formation. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03019–.

[Formosa-Jordan et al., 2012] Formosa-Jordan, P., Ibañes, M., Ares, S., and Frade, J. M. (2012). Regulation of neuronal differentiation at the neurogenic wavefront. *Development*, 139(13):2321–2329.

[François and Siggia, 2008] François, P. and Siggia, E. D. (2008). A case study of evolutionary computation of biochemical adaptation. *Phys Biol*, 5(2):026009.

[François and Siggia, 2012] François, P. and Siggia, E. D. (2012). Phenotypic models of evolution and development: geometry as destiny. *Curr Opin Genet Dev*, 22(6):627–633.

[Fujimoto et al., 2008] Fujimoto, K., Ishihara, S., and Kaneko, K. (2008). Network evolution of body plans. *PLoS One*, 3(7):e2772.

[Gardner et al., 2000] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342.

[Gerhart and Kirschner, 2007] Gerhart, J. and Kirschner, M. (2007). The theory of facilitated variation. *Proc Natl Acad Sci U S A*, 104 Suppl 1:8582–8589.

[Goodwin, 1982] Goodwin, B. C. (1982). Development and evolution. *Journal of Theoretical Biology*, 97(1):43–55.

[Green, 2002] Green, J. (2002). Morphogen gradients, positional information, and xenopus: interplay of theory and experiment. *Dev Dyn*, 225(4):392–408.

[Guerreiro et al., 2013] Guerreiro, I., Nunes, A., Woltering, J. M., Casaca, A., Nóvoa, A., Vinagre, T., Hunter, M. E., Duboule, D., and Mallo, M. (2013). Role of a polymorphism in a hox/pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci U S A*, 110(26):10682–10686.

[Hahn and Wray, 2002] Hahn, M. W. and Wray, G. A. (2002). The g-value paradox. *Evol Dev*, 4(2):73–75.

[Hartwell et al., 1999] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–C52.

[Hayden et al., 2011] Hayden, E. J., Ferrada, E., and Wagner, A. (2011). Cryptic genetic variation promotes rapid evolutionary adaptation in an rna enzyme. *Nature*, 474(7349):92–95.

[Heitzler and Simpson, 1991] Heitzler, P. and Simpson, P. (1991). The choice of cell fate in the epidermis of drosophila. *Cell*, 64(6):1083–1092.

[Horikawa et al., 2006] Horikawa, K., Ishimatsu, K., Yoshimoto, E., Kondo, S., and Takeda, H. (2006). Noise-resistant and synchronized oscillation of the segmentation clock. *Nature*, 441(7094):719–723.

[Hornung and Barkai, 2008] Hornung, G. and Barkai, N. (2008). Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput Biol*, 4(1):e8.

[Huppert et al., 1997] Huppert, S. S., Jacobsen, T. L., and Muskavitch, M. A. (1997). Feedback regulation is central to delta-notch signalling required for drosophila wing vein morphogenesis. *Development*, 124(17):3283–3291.

[Huynen, 1996] Huynen, M. A. (1996). Exploring phenotype space through neutral evolution. *J Mol Evol*, 43(3):165–169.

[Ibañes and Izpisúa Belmonte, 2008] Ibañes, M. and Izpisúa Belmonte, J. C. (2008). Theoretical and experimental approaches to understand morphogen gradients. *Mol Syst Biol*, 4:176.

[Irons and Monk, 2007] Irons, D. J. and Monk, N. A. M. (2007). Identifying dynamical modules from genetic regulatory systems: applications to the segment polarity network. *BMC Bioinformatics*, 8:413.

[Jaeger, 2011] Jaeger, J. (2011). The gap gene network. *Cell Mol Life Sci*, 68(2):243–274.

[Jaeger and Crombach, 2012] Jaeger, J. and Crombach, A. (2012). Life's attractors : understanding developmental systems through reverse engineering and in silico evolution. *Adv Exp Med Biol*, 751:93–119.

[Jaeger and Goodwin, 2002] Jaeger, J. and Goodwin, B. C. (2002). Cellular oscillators in animal segmentation. *In Silico Biol*, 2(2):111–123.

[Jaeger et al., 2007] Jaeger, J., Sharp, D. H., and Reinitz, J. (2007). Known maternal gradients are not sufficient for the establishment of gap domains in drosophila melanogaster. *Mech Dev*, 124(2):108–128.

[Jaeger and Sharpe, 2014] Jaeger, J. and Sharpe, J. (2014). On the concept of mechanism in development. *Towards a theory of development*, pages 56–78.

[Jeffery, 1999] Jeffery, C. J. (1999). Moonlighting proteins. *Trends Biochem Sci*, 24(1):8–11.

[Jung et al., 1998] Jung, H. S., Francis-West, P. H., Widelitz, R. B., Jiang, T. X., Ting-Berreth, S., Tickle, C., Wolpert, L., and Chuong, C. M. (1998). Local inhibitory action of bmps and their relationships with activators in feather formation: implications for periodic patterning. *Dev Biol*, 196(1):11–23.

[Karlebach and Shamir, 2008] Karlebach, G. and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10):770–780.

[Kashtan and Alon, 2005] Kashtan, N. and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A*, 102(39):13773–13778.

[Kashtan et al., 2009] Kashtan, N., Mayo, A. E., Kalisky, T., and Alon, U. (2009). An analytically solvable model for rapid evolution of modular structure. *PLoS Comput Biol*, 5(4):e1000355.

[Kauffman, 1969] Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*, 22(3):437–467.

[Kauffman, 1993a] Kauffman, S. A. (1993a). *The origins of order: Self organization and selection in evolution*. Oxford University Press, USA.

[Kauffman, 1993b] Kauffman, S. A. (1993b). *The origins of order: Self organization and selection in evolution*. Oxford university press.

[Kawaguchi et al., 2008] Kawaguchi, D., Yoshimatsu, T., Hozumi, K., and Gotoh, Y. (2008). Selection of differentiating cells by different levels of delta-like 1 among neural precursor cells in the developing mouse telencephalon. *Development*, 135(23):3849–3858.

[Kholodenko et al., 2002] Kholodenko, B. N., Kiyatkin, A., Bruggeman, F. J., Sontag, E., Westerhoff, H. V., and Hoek, J. B. (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci U S A*, 99(20):12841–12846.

[Kicheva et al., 2012] Kicheva, A., Cohen, M., and Briscoe, J. (2012). Developmental pattern formation: insights from physics and biology. *Science*, 338(6104):210–212.

[Kim et al., 2007] Kim, N., Alekseyenko, A. V., Roy, M., and Lee, C. (2007). The asap ii database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res*, 35(Database issue):D93–D98.

[Kimble and Simpson, 1997] Kimble, J. and Simpson, P. (1997). The lin-12/notch signaling pathway and its regulation. *Annu Rev Cell Dev Biol*, 13:333–361.

[Kirschner and Gerhart, 1998a] Kirschner, M. and Gerhart, J. (1998a). Evolvability. *Proc Natl Acad Sci U S A*, 95(15):8420–8427.

[Kirschner and Gerhart, 1998b] Kirschner, M. and Gerhart, J. (1998b). Evolvability. *Proc Natl Acad Sci U S A*, 95(15):8420–8427.

[Kouzarides, 2007] Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4):693–705.

[Kraut and Levine, 1991] Kraut, R. and Levine, M. (1991). Spatial regulation of the gap gene giant during Drosophila development. *Development*, 111(2):601–609.

[Kubota and Ito, 2000] Kubota, Y. and Ito, K. (2000). Chemotactic migration of mesencephalic neural crest cells in the mouse. *Dev Dyn*, 217(2):170–179.

[Kumar and Bentley, 2003] Kumar, S. and Bentley, P. J. (2003). *On growth, form and computers*. Academic Press.

[Lander, 2013] Lander, A. D. (2013). How cells know where they are. *Science*, 339(6122):923–927.

[Lewis, 1996] Lewis, J. (1996). Neurogenic genes and vertebrate neurogenesis. *Curr Opin Neurobiol*, 6(1):3–10.

[Lewis, 1998] Lewis, J. (1998). Notch signalling and the control of cell fate choices in vertebrates. *Semin Cell Dev Biol*, 9(6):583–589.

[Li et al., 1996] Li, H., Helling, R., Tang, C., and Wingreen, N. (1996). Emergence of preferred structures in a simple model of protein folding. *Science*, 273(5275):666–669.

[Lim et al., 2013] Lim, W. A., Lee, C. M., and Tang, C. (2013). Design principles of regulatory networks: searching for the molecular algorithms of the cell. *Mol Cell*, 49(2):202–212.

[Luscombe et al., 2004] Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A., and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431(7006):308–312.

[Ma et al., 2006] Ma, W., Lai, L., Ouyang, Q., and Tang, C. (2006). Robustness and modular design of the Drosophila segment polarity network. *Mol Syst Biol*, 2:70.

[Ma et al., 2009] Ma, W., Trusina, A., El-Samad, H., Lim, W. A., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4):760–773.

[Maeda and Sano, 2006] Maeda, Y. T. and Sano, M. (2006). Regulatory dynamics of synthetic gene networks with positive feedback. *J Mol Biol*, 359(4):1107–1124.

[Manrubia and Cuesta, 2015] Manrubia, S. and Cuesta, J. A. (2015). Evolution on neutral networks accelerates the ticking rate of the molecular clock. *J R Soc Interface*, 12(102):20141010.

[Manu et al., 2009a] Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M., and Reinitz, J. (2009a). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Comput Biol*, 5(3):e1000303.

[Manu et al., 2009b] Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M., and Reinitz, J. (2009b). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS Biol*, 7(3):e1000049.

[Marcon and Sharpe, 2012] Marcon, L. and Sharpe, J. (2012). Turing patterns in development: what about the horse part? *Curr Opin Genet Dev*, 22(6):578–584.

[Martin and Wagner, 2008] Martin, O. C. and Wagner, A. (2008). Multifunctionality and robustness trade-offs in model genetic circuits. *Biophys J*, 94(8):2927–2937.

[Matsuda et al., 2012] Matsuda, M., Koga, M., Nishida, E., and Ebisuya, M. (2012). Synthetic signal propagation through direct cell-cell interaction. *Sci Signal*, 5(220):ra31.

[Matsuda et al., 2015] Matsuda, M., Koga, M., Woltjen, K., Nishida, E., and Ebisuya, M. (2015). Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat Commun*, 6:6195.

[Mattick et al., 2010] Mattick, J. S., Taft, R. J., and Faulkner, G. J. (2010). A global view of genomic information–moving beyond the gene and the master regulator. *Trends Genet*, 26(1):21–28.

[Maynard Smith et al., 1985] Maynard Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. *Quarterly Review of Biology*, 60(3):265–287.

[Menshykau et al., 2012] Menshykau, D., Kraemer, C., and Iber, D. (2012). Branch mode selection during early lung development. *PLoS Comput Biol*, 8(2):e1002377.

[Milo et al., 2002] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.

[Morris et al., 2010] Morris, S. A., Teo, R. T. Y., Li, H., Robson, P., Glover, D. M., and Zernicka-Goetz, M. (2010). Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc Natl Acad Sci U S A*, 107(14):6364–6369.

[Munteanu et al., 2014] Munteanu, A., Cotterell, J., Solé, R. V., and Sharpe, J. (2014). Design principles of stripe-forming motifs: the role of positive feedback. *Sci Rep*, 4:5003.

[Murray, 1989] Murray, J. (1989). Mathematical biology. *C271*, pages –.

[Muzzey et al., 2009] Muzzey, D., Gómez-Uribe, C. A., Mettetal, J. T., and van Oudenaarden, A. (2009). A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*, 138(1):160–171.

[Narita and Rijli, 2009] Narita, Y. and Rijli, F. M. (2009). Hox genes in neural patterning and circuit formation in the mouse hindbrain. *Curr Top Dev Biol*, 88:139–167.

[Nijhout et al., 2003] Nijhout, H. F., Maini, P. K., Madzvamuse, A., Wathen, A. J., and Sekimura, T. (2003). Pigmentation pattern formation in butterflies: experiments and models. *C R Biol*, 326(8):717–727.

[Ouyang and Swinney, 1991] Ouyang, Q. and Swinney, H. L. (1991). Transition from a uniform state to hexagonal and striped turing patterns. *Nature*, 352(6336):610–612.

[Palau-Ortin et al., 2015] Palau-Ortin, D., Formosa-Jordan, P., Sancho, J. M., and Ibañes, M. (2015). Pattern selection by dynamical biochemical signals. *Biophys J*, 108(6):1555–1565.

[Panin et al., 1997] Panin, V. M., Papayannopoulos, V., Wilson, R., and Irvine, K. D. (1997). Fringe modulates notch-ligand interactions. *Nature*, 387(6636):908–912.

[Panovska-Griffiths et al., 2013] Panovska-Griffiths, J., Page, K. M., and Briscoe, J. (2013). A gene regulatory motif that generates oscillatory or multiway switch outputs. *J R Soc Interface*, 10(79):20120826.

[Payne and Wagner, 2013] Payne, J. L. and Wagner, A. (2013). Constraint and contingency in multifunctional gene regulatory circuits. *PLoS Comput Biol*, 9(6):e1003071.

[Payne and Wagner, 2014] Payne, J. L. and Wagner, A. (2014). The robustness and evolvability of transcription factor binding sites. *Science*, 343(6173):875–877.

[Petrovic et al., 2014] Petrovic, J., Formosa-Jordan, P., Luna-Escalante, J. C., Abelló, G., Ibañes, M., Neves, J., and Giraldez, F. (2014). Ligand-dependent notch signaling strength orchestrates lateral induction and lateral inhibition in the developing inner ear. *Development*, 141(11):2313–2324.

[Pierfelice et al., 2011] Pierfelice, T., Alberi, L., and Gaiano, N. (2011). Notch in the vertebrate nervous system: an old dog with new tricks. *Neuron*, 69(5):840–855.

[Pigliucci, 2008] Pigliucci, M. (2008). Is evolvability evolvable? *Nat Rev Genet*, 9(1):75–82.

[Pires-daSilva and Sommer, 2003] Pires-daSilva, A. and Sommer, R. J. (2003). The evolution of signalling pathways in animal development. *Nat Rev Genet*, 4(1):39–49.

[Plahte, 2001] Plahte, E. (2001). Pattern formation in discrete cell lattices. *J Math Biol*, 43(5):411–445.

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

[Raff and Conway Morris, 1996] Raff, R. A. and Conway Morris, S. (1996). The shape of life. genes, development, and the evolution of animal form. *Trends in Genetics*, 12(10):430–.

[Raff and Sly, 2000] Raff, R. A. and Sly, B. J. (2000). Modularity and dissociation in the evolution of gene expression territories in development. *Evol Dev*, 2(2):102–113.

[Raj and van Oudenaarden, 2008] Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.

[Raspopovic et al., 2014] Raspopovic, J., Marcon, L., Russo, L., and Sharpe, J. (2014). Modeling digits. digit patterning is controlled by a bmp-sox9-wnt turing network modulated by morphogen gradients. *Science*, 345(6196):566–570.

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

[Salazar-Ciudad et al., 2000] Salazar-Ciudad, I., Garcia-Fernández, J., and Solé, R. V. (2000). Gene networks capable of pattern formation: from induction to reaction-diffusion. *J Theor Biol*, 205(4):587–603.

[Salazar-Ciudad et al., 2003] Salazar-Ciudad, I., Jernvall, J., and Newman, S. A. (2003). Mechanisms of pattern formation in development and evolution. *Development*, 130(10):2027–2037.

[Salazar-Ciudad et al., 2001] Salazar-Ciudad, I., Newman, S. A., and SolÃ©, R. V. (2001). Phenotypic and dynamical transitions in model genetic networks. I. Emergence of patterns and genotype-phenotype relationships. *Evol Dev*, 3(2):84–94.

[Sato et al., 2013] Sato, M., Suzuki, T., and Nakai, Y. (2013). Waves of differentiation in the fly visual system. *Dev Biol*, 380(1):1–11.

[Schad et al., 2011] Schad, E., Tompa, P., and Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biol*, 12(12):R120.

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

[Schlosser and Wagner, 2004] Schlosser, G. and Wagner, G. P. (2004). *Modularity in development and evolution*. University of Chicago Press.

[Schuster et al., 1994] Schuster, P., Fontana, W., Stadler, P. F., and Hofacker, I. L. (1994). From sequences to shapes and back: a case study in rna secondary structures. *Proc Biol Sci*, 255(1344):279–284.

[Segal et al., 2003] Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet*, 34(2):166–176.

[Sempere et al., 2006] Sempere, L. F., Cole, C. N., Mcpeek, M. A., and Peterson, K. J. (2006). The phylogenetic distribution of metazoan micrornas: insights into evolutionary complexity and constraint. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 306(6):575–588.

[Setty et al., 2003] Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A*, 100(13):7702–7707.

[Shah and Sarkar, 2011] Shah, N. A. and Sarkar, C. A. (2011). Robust network topologies for generating switch-like cellular responses. *PLoS Comput Biol*, 7(6):e1002085.

[Shen-Orr et al., 2002] Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of escherichia coli. *Nat Genet*, 31(1):64–68.

[Sheth et al., 2012] Sheth, R., Marcon, L., Bastida, M. F., Junco, M., Quintana, L., Dahn, R., Kmita, M., Sharpe, J., and Ros, M. A. (2012). Hox genes regulate digit patterning by controlling the wavelength of a turing-type mechanism. *Science*, 338(6113):1476–1480.

[Shubin et al., 2009] Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature*, 457(7231):818–823.

[Smith, 1970] Smith, J. M. (1970). Natural selection and the concept of a protein space. pages –.

[Solé and Fernández, Pau and Kauffman, Stuart A,] Solé, R. V. and Fernández, Pau and Kauffman, Stuart A, Journal = arXiv preprint q-bio/0311013, Y. . . P. . . O. . a. T. . . Adaptive walks in a gene network model of morphogenesis: insights into the cambrian explosion.

[Solé and Valverde, 2008] Solé, R. V. and Valverde, S. (2008). Spontaneous emergence of modularity in cellular networks. *J R Soc Interface*, 5(18):129–133.

[Stathopoulos and Levine, 2002] Stathopoulos, A. and Levine, M. (2002). Linear signaling in the toll-dorsal pathway of drosophila: activated pelle kinase specifies all threshold outputs of gene expression while the bhlh protein twist specifies a subset. *Development*, 129(14):3411–3419.

[Stich et al., 2007] Stich, M., Briones, C., and Manrubia, S. C. (2007). Collective properties of evolving molecular quasispecies. *BMC Evol Biol*, 7:110.

[Strogatz, 2014] Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press.

[Taft et al., 2007] Taft, R. J., Pheasant, M., and Mattick, J. S. (2007). The relationship between non-protein-coding dna and eukaryotic complexity. *Bioessays*, 29(3):288–299.

[Takei et al., 2004] Takei, Y., Ozawa, Y., Sato, M., Watanabe, A., and Tabata, T. (2004). Three drosophila ext genes shape morphogen gradients through synthesis of heparan sulfate proteoglycans. *Development*, 131(1):73–82.

[Tanaka et al., 2005] Tanaka, Y., Okada, Y., and Hirokawa, N. (2005). Fgf-induced vesicular release of sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. *Nature*, 435(7039):172–177.

[Ten Tusscher and Hogeweg, 2011] Ten Tusscher, K. H. and Hogeweg, P. (2011). Evolution of networks for body plan patterning; interplay of modularity, robustness and evolvability. *PLoS Comput Biol*, 7(10):e1002208.

[Thomas, 1973] Thomas, R. (1973). Boolean formalization of genetic control circuits. *J Theor Biol*, 42(3):563–585.

[Tosh and McNally, 2015] Tosh, C. R. and McNally, L. (2015). The relative efficiency of modular and non-modular networks of different size. *Proc Biol Sci*, 282(1802).

[Tóth-Petróczy and Tawfik, 2013] Tóth-Petróczy, A. and Tawfik, D. S. (2013). Protein insertions and deletions enabled by neutral roaming in sequence space. *Mol Biol Evol*, 30(4):761–771.

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

[Uzkudun et al., 2015] Uzkudun, M., Marcon, L., and Sharpe, J. (2015). Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. *Mol Syst Biol*, 11(7):815.

[Verd et al., 2014] Verd, B., Crombach, A., and Jaeger, J. (2014). Classification of transient behaviours in a time-dependent toggle switch model. *BMC Syst Biol*, 8:43.

[von Dassow et al., 2000] von Dassow, G., Meir, E., Munro, E. M., and Odell, G. M. (2000). The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192.

[von Dassow and Munro, 1999] von Dassow, G. and Munro, E. (1999). Modularity in animal development and evolution: elements of a conceptual framework for evodevo. *J Exp Zool*, 285(4):307–325.

[Waddington, 1939] Waddington, C. H. (1939). An introduction to modern genetics. *An Introduction to Modern Genetics.*, pages –.

[Wagner and Altenberg, 1996] Wagner, G. P. and Altenberg, L. (1996). Perspective: complex adaptations and the evolution of evolvability. *Evolution*, pages 967–976.

[Wagner et al., 2001] Wagner, G. P., Mezey, J., and Calabretta, R. (2001). *Natural selection and the origin of modules*. na.

[Wagner et al., 2007] Wagner, G. P., Pavlicev, M., and Cheverud, J. M. (2007). The road to modularity. *Nat Rev Genet*, 8(12):921–931.

[Werner et al., 2010] Werner, T., Koshikawa, S., Williams, T. M., and Carroll, S. B. (2010). Generation of a novel wing colour pattern by the wingless morphogen. *Nature*, 464(7292):1143–1148.

[Wolf and Arkin, 2003] Wolf, D. M. and Arkin, A. P. (2003). Motifs, modules and games in bacteria. *Curr Opin Microbiol*, 6(2):125–134.

[Wolpert, 1968] Wolpert, L. (1968). The french flag problem: a contribution to the discussion on pattern development and regulation. *Towards a theoretical biology*, 1:125–133.

[Wolpert, 1969] Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *J Theor Biol*, 25(1):1–47.

[Wolpert et al., 2002] Wolpert, L., Beddington, R., Brockets, J., Jessel, T., Lawrence, P., and Meyerowitz, E. (2002). *Principles of Development*. Oxford University.

[Yamanaka et al., 2010] Yamanaka, Y., Lanner, F., and Rossant, J. (2010). Fgf signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*, 137(5):715–724.