A study on autocatalysis through synthetic biology

Exploration of spatiotemporal dynamics in the presence or absence of synthetic autocatalytic Hepatocyte Growth Factor signaling in mammalian cells

Autor: Diego Bárcena Menéndez

TESI DOCTORAL UPF / ANY L'any de la tesi: 2016

DIRECTOR DE LA TESI
Mark Isalan
Imperial College London
Departament Biomedicina

"La biologie est une science comme les autres, soumise aux mêmes lois, aux mêmes règles, à la même évolution ; les mêmes méthodes lui sont applicables. Comme les autres sciences, elle doit être successivement descriptive, analytique et synthétique."

[Leduc, 1906]

"Perhaps there has been an optimisation of the variable parameters and relationships over many cycles in such a sense that the emergence of structure is increasingly favoured. Perhaps cosmic evolution shares this general theme with the evolution of life. Perhaps..."

[Jantsch, 1980]

"Complexity by itself is not the catalyst in which a system might crash. Rather, it is how the complexity emerges in a system that determines whether that system will do what it was intended to do or morph into an unworkable organization clogged by bottlenecks and blockages"

[Samuels, 2013]

Acknowledgments

Firstly, Mark Isalan who conceptually conceived the idea for the project and who has worked magic to maintain the necessary cash flow for the experiments. I would also like to thank all the members of my thesis advisory committee for their guidance and support throughout these years: Jordi Garcia Ojalvo, James Sharpe and Guillaume Fillon.

On the practical side, Jiayi Zhou who contributed significantly to this work during her three month rotation at Imperial College London. I would also like to thank and wish the best of luck to Natalie Scholes Sian who will be taking this project forward. On the theoretical side, Luis Diambra and Mark Sturock, for their invaluable insights into the abstract world of physics and mathematics. Andreas Broedel for his help with proofreading, and structuring of this thesis.

A special thank you goes to David Kriegleder with whom I time and again peeked outside of the scientific ivory tower. To all the "Geoffrey's" in the world who have kept us alive, and to all the destitute people we've met in our journeys I have this to say: I thank you for teaching me the privilege of being able to do a PhD.

Lastly to my wife Alicja, who more than anybody has had to endure the emotional swings of a PhD student.

Resumen

Una de las preguntas sin resolver en la biologia celular, es como las células mamíferas logran generar patrones estables como organos o seres vivos, en un entorno variable. El concepto matemático del bucle de retreoalimentación, es una herramienta que puede generar orden. En esta tesis, presento dos proyectos que forman parte de una idea iterativa para recrear patrones biológicos sintéticamente en células mamíferas. En la primera parte, presento la creación de una linea celular que funge como detector de niveles de la hormona HGF a través de una reportero transcripcional. En la segunda parte, demuestro la reprogramación de esta células con fin de producir HGF en respuesta a HGF, en efecto creando un bucle de retroalimentación positivo. En ambos proyectos, utilizo microscopia cuantitativa espaciotemporal para analyzar y medir la evolución dinámica de las células en respuesta a un estímulo de HGF.

Abstract

One unanswered riddle in biology is how can mammalian cells organize to generate ordered patterns such as organs and living beings, in an ever changing environment. An underlying mathematical principle for the generation of order is given by feedback motifs. Here, I present two projects which are part of an effort to recreate stable ordered patterns in a cellular system through information encoded in DNA. In the first part, I present a receiver mammalian cell line which can accurately sense the diffusible Hepatocyte Growth Factor (HGF) through a transcriptional reporter. In the second part I reprogrammed this cell line so that it produces more HGF in response to HGF, in effect creating an autocatalytic positive feedback. In both cases, I have used spatiotemporal quantitative microscopy analysis to monitor the dynamic evolution of the cell lines in response to a HGF stimulus.

Contents

Part I

Introduction

Chapter 1

BIOLOGICAL SYSTEMS: ORDER, DISORDER AND INFORMATION

In this chapter, I will start by elaborating on a definition of biological systems through systems theory. This will serve as a primer for the second chapter of the introduction, which will focus on the synthesis of biological form and function. In the third chapter I will introduce the particular system topology which I have attempted to synthesize, and its particular manifestations and properties in systems of all kinds.

The advent of systems biology has shown us that indeed an interaction of $(A+B)$ is not always stoichiometrically equivalent to the result (AB), a conclusion with far reaching consequences in the design of synthetic biological systems. In this context, I have worked on engineering a cellular signaling system which may help future scientists study the properties of non-linear systems.

As a biologist, it is truly exceptional times to do science. On the one hand, biological systems theory provides an excellent conceptual framework through which to filter our understanding of cells. On the other hand, the tools available to modify and interact with these systems are becoming ever more powerful, as our understanding of nature deepens. Lastly, the rule-book for the behavior of systems show remarkable similarity across scales, thus systems science is emerging as an interdisciplinary bridge. I will use advances from different scientific disciplines in the work presented here, in an attempt to recreate and analyze complex patterns encoded in simple instructions. The motivation of this work was to try to bring order to complexity, so in the next chapters I will give a brief definition of complex systems.

1.1 Classifying biological complexity

That biological systems are immensely complex is undeniable. However, for several reasons, we are still lacking a universally accepted quantitative measure of the degree of complexity of systems. In order to describe the world around us, the traditional reductionist western scientific approach will dissect an object to be studied and describe it by division into elemental units of measurement (length, mass, time, temperature, electric charge). This approach to simplify the world serves as a filter for unnecessary information, so choosing the appropriate measurement unit strongly depends on what we are trying to achieve.

When it comes to measuring complexity, the attempts thus far have been algorithmic in nature, meaning that truly random processes, regardless of how easily they are defined, will have infinite complexity (see [Li and Vitányi, 2009]). Needless to say, these measures of complexity leave myriads of questions unanswered: What information if any should a unit of complexity contain? If we were to compare a crystal made of a rare mineral and a dog, can we capture which is more complex? Should a measure for complexity be a common **integrating** measurement of all the accepted units of measurement? If there is a unit, will there also be a natural unit such as for length, speed, mass etc.. for complexity? Are there limits or is complexity infinite?

We choose to categorize objects in terms of the objective we want to achieve. For example, in a ship's log book, the objective will be not to overload, thus a unit of weight will help us decide how much we can carry on a barge to cross a river. In this case, a crystal and a dog with weights of 10kg, will be indistinguishable from each other. If however, the objective is to measure displacement potential, then clearly there will be differences in our two objects.

Biology presents us with a further conundrum: Let us say we take our dog, and measure the exact weight of every atom in his body at a given time-point, and we then place all the atoms one-by-one in a container. It is not to be expected that these components will arrange themselves back into a dog. Physical and chemical systems follow the laws of thermodynamics and progressively move towards more and more disordered states. Conversely, it appears that, instead of becoming more disordered as time progresses, the components of biological systems seem to have an intrinsic drive towards order. This would seem to go against the second law of thermodynamics and its principle of irreversibility in time. The attempts to reconcile the biological evolution of form and function and the second law are as old as the theories themselves. Erwin Schrödinger popularized the question in his famous book *What is Life?*, where he claimed that life feeds on *negentropy*, or negative entropy. And life's apparent negentropic drive is maintained in the micro-scales, too. For example a ribosome will always only order amino acids with an mRNA template to make a more ordered protein structure and never the

reverse. If we look at larger biological sizes, we see that the bias towards order of biological system is conserved across scales; barring mutations, a fertilized egg will always strive towards forming a higher ordered organism from the chemical components it encounters in disordered states. An adult being will, however, never revert to a fertilized egg, although there is no known physical law that forbids this.

Biological systems require, of course, energy to drive this curious channeling towards order. So where do biological systems take this free energy from? The solution to this conundrum lies on the cosmological scale. Yes, life does seem to tend towards order, but the energy is being fed by nuclear fusion occurring in the sun and converted into ordered structures on earth. The net balance in entropy is maintained overall. Nonetheless, the amount of entropy produced by planet earth as a whole might well be different than if life was not present: an interesting thought experiment when it comes to the search of "life as we don't know it" [Azua-Bustos and Vega-Martínez, 2013].

It is then not but a mere curiosity that chemical processes have evolved biology to incorporate this energy and transform it into structure and form. Even so, curiosity is one of the key drivers of human innovation, and in the following section I will highlight some of the historical efforts to describe these processes. It will become evident that one common feature of systems that do work with energy to generate order, whether alive or not, is that they all must have autocatalytic processes, also known as positive feedbacks. Since engineering positive feedbacks using a synthetic biology approach is the main research topic of this thesis, it is important to place these in a wider context.

1.1.1 Biological systems - the theory of dissipative systems

Ilya Prigogine devised the theory of so called dissipative systems in the 1970s for which he was awarded the Nobel prize in Chemistry in 1977. Prigogine's stroke of genius was to see that non-equilibrium in microscopic systems, far from being transitory states of equilibrium states, could be a source of macroscopic order. It was an observation which at the time contrasted with the understanding of physical systems, which were always thought to tend towards stable states [Prigogine, 1977].

Dissipative systems were defined by Prigogine so that they do not violate the second law of thermodynamics. Up until that time, the reversibility of chemical reactions was thought to depend solely on the law of mass action to maintain stoichiometric balance in chemical reactions. Furthermore, thanks to Boltzmann, it was theoretically known that molecules left on their own would very rarely spontaneously form ordered chemical reactions. The caveat to these effects: all reactions had to occur in closed systems. By imagining and describing systems which were open to the environment, Prigogine opened a world which would have enor-

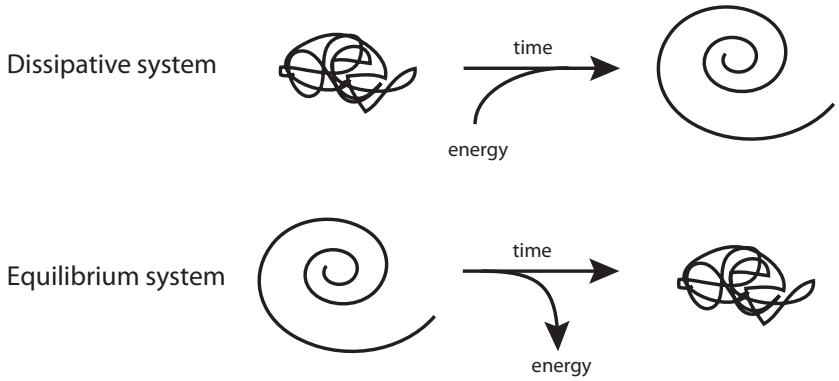


Figure 1.1 – A dissipative system will take energy from the environment and transform it into order. A thermodynamic stable system will release energy in time at the expense of order.

mous repercussions. His main finding was the discovery of certain microscopic processes under specific conditions far from equilibrium could very well be irreversible and thus follow the "arrow of time" (**Figure 1.1**). In order to do this, these systems had to contain at least one autocatalytic reaction. The work presented here is an attempt to generate and study one such reaction in a mammalian cell system.

In contrast to closed systems, where energy, number of particles, and total volume must remain constant, dissipative systems are open and these parameters will widely vary. This allows dissipative systems to absorb energy to maintain continuous entropy production, and to *dissipate* the accumulated entropy. Thus, a dissipative system is by definition intrinsically in a non-equilibrium state, maintained as long as energy exchange occurs. Far from being constrained to this non-equilibrium state though, these systems are in constant fluctuation. Given high enough disturbances in the system, these fluctuations will lead to transitions between non-equilibrium states. The disturbances needed are dynamic and depend solely on each particular dissipative structure.

An interesting conclusion from this is that a system may be described by the total amount of order it contains and the total amount of energy and order it can process from outside such that:

$$dS = d_e S + d_i S \tag{1.1}$$

$$d_i S \geq 0 \tag{1.2}$$

Where dS - the total entropy in the system - is the summation of the change in the entropy due to the exchanges between the environment and the system $d_e S$, and the intrinsic internal entropy of the system $d_i S$. If $d_i S$ is higher than that

which the environment can absorb, then the entropy may manifest itself in the form of macroscopic structure: the spontaneous emergence of order.

In his book, *The Self-Organizing Universe*, Erich Jantsch argues that biological systems particularly favor dissipative systems in the following ways:

> –They are linked with their environment by energy exchange which permits the maintenance of the structure far from equilibrium.
>
> –They include a large number of chemical reactions and transport phenomena, the regulation of which depends [...] on non-linear factors of molecular origin.
>
> –They are in high non-equilibrium, not only from the point of view of energy, but of matter exchange, since the reaction end products are either eliminated from the system or are transported to other locations in order to fulfill their functions there.
>
> [Jantsch, 1980]

Although the outcome of events are unpredictable, the individual components of a dissipative structure follow self-consistency rules, which means that whatever comes into being has to be consistent with itself and its parts. This is particularly interesting as it makes it possible in theory to represent and simulate these systems mathematically.

1.1.2 Arguments against dissipative structures

Ilya Prigogine's explanation for dissipative systems is thought to be a close approximation to reality. However a main paradox remains, as first noted by Daniel R. Brooks et. al, in [Weber et al., 1988] and Collier in the same book. Namely, in Prigogines theory, there must exist a preordained configuration of a system to trap energy, and differentiate it from the outside world. This is accurate when performing experiments with chemical systems, which are defined by the experimentalist. However the application to biological systems might not be so straightforward. Brooks and others (for a more recent attempt see [Pascal et al., 2013]), take a more cosmological approach, and try to determine whether biological systems need be closed from the environment physically, or whether in theory it is possible to generate form-replicating systems without the need for boundaries.

Brook's main thesis is that the equivalent for the second law of thermodynamics in biology is a combination of Dollo's irreversibility law and natural selection. The former, proposed by Louis Dollo in the 1890s [Dollo, 2016], states that evolution is irreversible and has inherent ordering constraints; the latter states that the environment orders and places constraints on life. This irreversibility principle,

however, is not without its own controversy and evidence against it may be found in nature.

To understand Daniel Brooks' reasoning we can follow the example presented in [Weber et al., 1988] for a developing embryo. Although life exists at different pressures and temperatures, during the lifetime of a biological system these usually do not vary, and thus we can write entropy in relation to the free Gibbs energy of a system:

$$\Delta(PV/T) \propto S \tag{1.3}$$

Where $V =$ the number of cells, $P =$ the environmental pressure on the organism, $T =$ the metabolic temperature, and $S =$ the entropy of the system. In contrast to dissipative systems, here V is not a boundary condition to which the system must conform, but a key determinant in the evolution of the system. Thus, in Brooks' model the amount of entropy is directly proportional to the number of components of a system for a general system where all components are equal. As for the right side of the proportionality, namely the entropy of the system, if one takes the general accepted formula for closed systems:

$$S = k(log_e(W)) \tag{1.4}$$

where k is Boltzmann's constant (Energy/Temperature) and W is the number of possible micro states of the system. And if one assumes that particles in a system are not homogeneous (as is usually the case) then entropy is:

$$S = -\sum_i [(p_i) log_e(p_i)] \tag{1.5}$$

where p_i is the frequency of occurrence of the *i*th particle. One way of measuring the complexity in the system is proposed by Brooks as "information", and here is where the conceptual difference with dissipative structures lies. Brooks proposes a Hierarchical Information Theory (HIT) to try to generate a general measure to describe complexity. Briefly, HIT is constructed by equating the thermodynamic entropy to the Shannon-Weaver entropy. In a K-dimensional system, composed of N^k parts, which can assume M states, the total number of possible states is M^{N^k}. Brooks argues that the information content of the system can be defined generally as:

$$I = S_{max} - S \tag{1.6}$$

With that assumption in mind, for our M^{N^k} possible states, one can measure the Shannon-Weaver entropy or information capacity of the system for a particular state as:

$$C_I = N^k \frac{((N^k + 1)}{2} log_2(M) bits \tag{1.7}$$

Now of course, one can easily argue that for a biological developing cell, not all developmental fates are accessible. Firstly because most states possibly lead to death of the system (and thus $dS < 0$) and secondly because $d_i S$ is naturally also constrained. We are only now starting to map out the possible pathways a cell can take (for a recent exciting computational approach see: [Rackham et al., 2016]), but the options will be limited. Brooks foresaw this dilemma, and suggests two forms of the information capacity for a system:

$$C_I = C_O + C_D \tag{1.8}$$

Where C_O is the capacity for order or "information" of the system and C_D is the capacity for disorder or "entropy" of the system. Thus, Brooks is reconciling the second law by introducing information content. A second important conceptual difference to dissipative structures is that HIT recognizes that "the whole may be greater than the sum of its parts".

1.1.3 Dynamics: Reaction-Diffusion systems

All the above examples try to reconcile the fact of entropy dissipation in biological systems from the thermodynamic point of view. In reality, biological systems do consume and move energy, however this energy is usually stored and moved in matter (chemical potentials, etc.). Thus, it has been more intuitive to describe the flows in biological systems, through the dynamics of diffusion of matter, expressed through reaction-diffusion equations of the type:

$$\delta_t u = D\delta_x^2 u + R(u) \tag{1.9}$$

where u is the diffusing chemical, D is the diffusion coefficient in m^2/s and $R(u)$ is the reaction kinetics of u. One of the key conceptual changes in this field was brought by the mathematician Alan Turing in his seminal paper on chemical diffusion [Turing, 1952]. Turing expanded the field by introducing a second component v which would inhibit u and with different diffusion coefficient D so that:

- u catalyzes itself and v and diffuses slowly
- v inhibits u and diffuses quickly

As we saw with the theory of systems in previous sections, the addition of feedback to a system will produce order that is somewhat unexpected from a classical thermodynamic perspective. Thus, perhaps unsurprisingly to us now, Turing's simple addition to the theory of diffusion can give rise to a seemingly

endless number of patterns. Nowadays, the conceptual framework set by Turing has been applied to model a myriad of chemical and biological processes.

In the particular case of biochemical systems, where reactants are expected to be produced and degraded, one can represent the Turing system as:

$$u = D_u\Delta^2 + p_u f(u, v) - \mu_u u \tag{1.10}$$

$$v = D_v\Delta^2 + p_v g(v, u) - \mu_v v \tag{1.11}$$

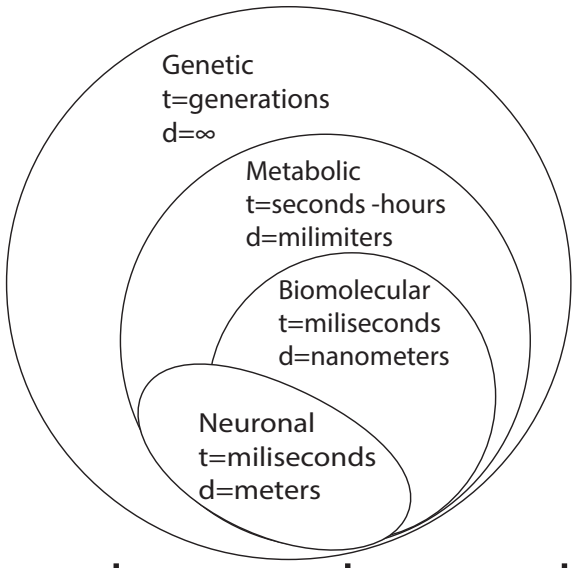
Where p and μ are functions describing production and degradation respectively. We recently showed [Diambra et al., 2015] that it is theoretically possible for a system with a single function p to generate Turing instabilities; yet another demonstration of the possibilities of reaction-diffusion systems.

1.2 The theory of communication in biology

Biological systems undergo constant exchange of energy and entropy with their environments and with each other. As described in section 1.1.2, it is possible to equate energy and information. Thus, both information and energy exchange might be the driver of order we observe in biology. Jantsch [Jantsch, 1980] dissects communication into four hierarchical domains shown in **Figure 1.2a** depending on the spatio-temporal scales at which exchange of information takes place. The four domains are: **Genetic, Metabolic, Biomolecular, Neuronal** (examples are given below). Jantsch proposes that all dissipative systems possess this intrinsic communication hierarchy. He further argues that, for communication between systems to occur, there must be a level of cognition and recognition at the hierarchical level between two systems.

Let us examine this in a practical worked example. Vervet monkeys (*Chlorocebus pygerythrus*) are a small primate species which lives in highly complex social groups of between 10 and 70 individuals. We will look at the different levels of communication in two systems, the individual vervet and the community, and I will highlight the hierarchy of communication in italics. Let the start of the signaling cascade be a Vervet identifying a menacing shadow lurking above (*Neuronal:Biology*), a sudden adrenaline rush (*Metabolic:Biology*) invades it and it immediately starts to make a series of whistles that are partly innate (*Genetic:Biology*) and partly learned from others in the community(*Genetic:Society*). The signal is received and interpreted correctly immediately by other vervets (*Neuronal:Biology* learned through (*Genetic:Society*)). Some monkeys encourage each other to whistle further in a group to amplify the signal (positive feedback) to make sure that every individual receives the message (*Metabolic:Society*), while others scramble for cover.

(a)



(b)

	Genetic	Metabolic	Neuronal
Society	Culture, Technological innovation	Economy, money	Face to face communication, Internet
Biology	Evolution, Instinct, DNA	Glycolisis, Kinase signalling, Hormones	Nervous system, Electric communication (biofilms)
Climate	Decadal Oscillations, Seasons	Carbon/Nitrogen cycles, water vapour	Weather, Tornadoes, hurricanes

Figure 1.2 – Communication hierarchies in 1.2a are shown as self encompassing forms as envisioned by Erich Jantsch, and in 1.2b a few examples are listed in three different systems.

One can see several things from this example. First, the signals can be accurately described and these seem to act in synchrony across the hierarchies. This link between the different levels is key for the survival of dissipative structures. Second, the signals have clear sender and receiver modules, which can only interpret certain signals whilst ignoring all other inputs. And third, communication is restricted to systems, i.e. the adrenaline will be sensed and interpreted by a specific receptor module in the body, the whistles can be heard by other organisms, but will not be interpreted, etc. Communication events which occur outside the system can be called interactions. Last, although it might not be evident, feedback mechanisms play crucial roles in interpreting and transferring the signals into ordered responses.

In **Figure 1.2b** I have listed other examples of hierarchical communication in different systems. As this classification of systems can apply to different scales

at least qualitatively, it would be beneficial to find common quantitative measures that describe and help to predict such systems. Generating a reproducible system where one can test the predictions from mathematical models in a relatively simple setup was the fundament of creating the sender-receiver system presented in **Part II**.

1.2.1 Information theory applied to biology

To explain communication between biological cells, scientists initially postulated a theory of an invisible force driving communication between cells which they called Mitogenic Rays [Lorenz, 1934]. This theory was never accepted, since the rays were never found. A similar philosophical conundrum arises whilst trying to quantify information in biological communication. The debate revolves around the question of whether or not signals are carriers of transmitted information (TI). On the one side, there are those who ascribe informational value in signals, whereas on the other side some argue that information is not transmitted at any time; rather, it is only assigned this property because of a perceiver's correct interpretation thereof. Whichever side of the argument you take, it is nonetheless useful to postulate information signals and quantify biological communication.

Mathematical definitions of information

The first to define information mathematically was Claude Shannon, in his seminal work published together with Warren Weaver [Shannon and Weaver, 1949]. In this section I would like to highlight its main tenets and what I believe to be its main consequence for biological systems. According to Shannon, a variable is defined to carry information when knowing the exact value of the variable reduces the uncertainty of the observer. As such, information transfer between a sender and a receiver will reduce uncertainty for both actors. Uncertainty, in this case being the sum of probabilities for a set of elements.

To define the uncertainty of a variable, Shannon used Boltzmann's description of thermodynamic particles (**see equation1.5**) to ascribe the amount of information contained in these particles. From this, just as is the case with energy in the second law of thermodynamics, it follows that information by itself transmitted through a channel cannot increase as it is transmitted, it can only decrease or remain the same.

As to the quantifiable measure of the uncertainty of a variable, Shannon called it Entropy. For example for a random variable x the entropy is defined as:

$$H[x] = -\sum_{x}(p_x)log(p_x) \tag{1.12}$$

In this case p_x is the probability distribution of the variable x and the logarithm can be chosen arbitrarily (log2 makes the unit bits, the natural logarithm makes the units nats - natural units). In other words, the entropy of a variable x is a measure of the information gained from a given observation given a probability distribution p_x. In general, if $H[x]$ is high, there is a higher uncertainty on the information carried by element xi. If, however the value for x is known, then the entropy will be zero as there will be very low uncertainty in the system.

The next logical step in Shannon formalism was to define the relationship of two (or more) variables whence they depend on each other through the conditional entropy of $H(y|x)$. Such that if we know the distribution p_x, then the additional information gained by a second variable p_y is given by $log_2(p(y|x)$ and the additional entropy is:

$$H[x, y] = -\int\int p(y, x) log_2 p(y|x) dy dx \tag{1.13}$$

Intuitively, this tells us that the information needed to describe a variable y alone can be estimated from the information provided by x plus the information to specify y given x. In such a system, it is then possible to calculate the amount of shared information both from the probability distributions:

$$MI(X;Y) = \sum_X \sum_Y \frac{p(x, y) log_2 p(x, y)}{p(x)p(y)} dx dy \tag{1.14}$$

and from the individual and joint entropies, such that the total shared information per unit time of measurement is the summation of the information on the input x plus an output y, less the joint entropy:

$$MI(X;Y) = H[X] + H[Y] - H[X, Y] \tag{1.15}$$

The mutual information between two events is also known as the rate of transmission of information between x and y.

Shannon and Weaver's mathematical description of information theory was widely applied to telecommunication, where bits of information in the form of electrical currents can be driven through copper wires and transmitted to a receiver on geographical distance scales. It has also been used extensively in analytic linguistics to describe language rules. Surprisingly, in the research for this thesis I found that for all its scope, Shannon's definition for information, fails to accommodate for the effects of feedback. In view of the incredible speed and reliability of microelectronic devices to transmit signals and communicate with one another, it has been a subject of intense investigation to understand how Shannon's laws behave in the presence of feedback [de Ronde et al., 2010, Tostevin and ten Wolde, 2009]. Furthermore, in light of the apparent function of

feedbacks in the translation of environmental information within systems, it might provide some insight to apply information theoretical analysis to these interesting network topologies.

Examples of information theory applied to biology

Intuitively, just as with dissipative systems and diffusible systems, information must be a scalable property. Although - perhaps as a particular warning for future scientists who might be tempted to ascribe predictive properties to this robust mathematical model - even Shannon himself is reported to have found "applications of his work outside of communication theory to be suspect" [Ritchie, 1986]. Perhaps, Shannon failed to see that the real world could very well be described by his theories.

Cells undoubtedly communicate with each other, and communication can be broken down into three parts: a sender, a receiver, and a channel. When a cell is not in its normal state, say when it enters a cancerous state, the information processing capacities will change [Levchenko and Nemenman, 2014]. For homeostasis and order to be maintained in multicellular organisms, there needs to be a constant awareness and adaptation to a changing environment. To achieve this, the signaling molecules used by multicellular organisms have evolved to transmit and receive information over short-(e.g. the notch signaling system) mid-(e.g. morphogens such as WNTs and Growth/Morphogenic factors) or long-ranges (hormones in the blood, insulin, small molecules). If biological systems have evolved from chemical manifestations of feedback motifs, as some have suggested [Kauffman, 2011], they might provide an interesting tool to reconcile Shannon's mathematical description of communication with Prigogines definition of dissipative systems.

In **Part II**, of this thesis I have first focused on constructing a cell line which can detect a diffusible signal. Then in **Part III** I expanded this cell line, to include an autocatalytic positive feedback. This work was partly inspired as an approach to describe a biological system both from the point of view of transport of matter (diffusion) and transfer of information (bitrate). This system will hopefully serve as a fundament for future studies of these properties and the extent to which different network topologies affect them.

In [Barcena Menendez et al., 2015] we reviewed some systems that have been constructed through synthetic biology and propose how information theory could prove to become a powerful tool to construct and help understand biological processes. Furthermore, in **section 6.2.3** I describe an information theoretical approach to analyze a biological feedback which could perhaps serve as a basis for further research into this area.

Feedbacks and information transmission

Interestingly, some work has been done in regard to the capacity of feedbacks to induce long-term information transmission, e.g. memory in systems (reviewed in [Purcell and Lu, 2014]). Evidently, in rationally designed biological systems both negative and positive feedbacks can confer some sort of memory effects, similar as to those observed in computer memory.

It is my opinion that biology does not function as a set of transistors transferring information by RNA-polymerases, although this view has recently found exciting applications (see: ([Nielsen et al., 2016])). Biological matter is more complex than electrons transferring information in chips. Unlike transistors, biological systems are disspiative structures which must remain flexible whilst challenged by an ever changing environment. If feedbacks do indeed function as memory devices, these effects will be nothing similar to what microchips exhibit both in scale and heterogeneity. Further down I will go into the technical details of measuring memory effects of feedbacks (section 3.3).

Chapter 2

SYNTHESIS OF BIOLOGICAL SYSTEMS

In the previous chapter two different approaches were proposed to categorize a system, which seemingly converge in the term entropy, a measure of information content in communications theory, and a measure of disorder in thermodynamics. This thesis is based on constructing a positive feedback to further our understanding of these interesting network topologies in biology. The act of "constructing" something in biology has recently being crystallized into a new field termed synthetic biology. In this chapter I will give a very brief overview of the field, and I will defend why I believe synthetic biology provides an excellent approach for studying biological phenomena beyond genome editing. I will begin by describing the first attempts to synthesize biological form and function by the french chemist Stephane Leduc in the beginning of the 20th Century. I will use the conceptual framework set by Stephane Leduc to categorize distinct fields of synthetic biology. This section will serve as an introduction to the final chapter in which I will introduce the main matter of my dissertation: the generation and characterization of a synthetic positive feedback cell line.

2.1 Stephane Leduc: The first synthetic biologist.

The first recorded mention of synthetic biology is found at the beginning of the 20th century when a French chemist named Stephane Leduc published a book titled *"La Biologie Synthetique"* [Leduc, 1912]. This book followed his first more controversial book *"Theorie Physico-chimique de la Vie et Generations Spontanee?"* [Leduc, 1906] which translates to "Physical-Chemical theory of life and spontaneous generations [emergence]" and was translated to English in 1914 by the British doctor W. Deane Butcher [Leduc, 1914].

Leduc presented a very interesting idea, namely that any human science seemed to go through three distinct and ordered phases of discovery:

1. **Observational** - Describe a phenomena
2. **Analytical** - Dissect the phenomena
3. **Synthetic** - Reproduce the phenomena

Leduc main thesis in his books is that biological sciences were at the time deep in the observational phase, and slowly entering the analytic one. For example, scientists were discovering that the basic components of life were of chemical nature. The "Chromatin theory of inheritance" based on Walther Flemming's discovery of chromosomes by staining with basic dyes [Flemming, 1880], hinted strongly at the chemical nature of biological cells. Still, beyond carbon and some organic compounds, the actual chemical composition of life eluded researchers at the time. Morphological form, it seemed, was conserved and inherited from macroscopic organisms to the microscopic world within cells. But how?

To make sense of this, Stephane Leduc attempted to synthesize some of the forms and functions he could observe. He first defined what he believed were three distinctions within synthetic biology of which only one was being pursued: **synthetic organic chemistry**. Scientists working in this field were proving once and again that it was possible to synthesize chemicals in the lab, which were previously thought to exist solely in natural organisms (e.g. amino acids). Then there were two divisions of synthetic biology that had yet to be studied and Leduc intended to establish himself at the forefront: **morphogeny**, which dealt with the synthesis of form and **physiogeny** which consisted of the synthesis of function. Leduc was fascinated by these two ideas and particularly intent on whether he could synthesize morphological aspects of life with the simple mixture of chemicals compounds in solution. He attempted to reconstitute conserved biological forms both microscopically as shown in **Figure 2.1a**, and macroscopically as shown in **Figure 2.1b**.

To the contemporary observer, it is obvious that Stephane Leduc had deep misconceptions about his observations. Take for example the mechanistic explanations he put forth, where he believed that the forms he was generating could be explained by "field-like forces" underlying diffusion, which would act similarly to Faraday's electromagnetism, or the 'mitogenic waves' I mentioned in **section 1.2.1**. But it is perhaps unfair to criticize him in hindsight, because his larger vision was truly remarkable.

As Leduc's experiments evidence, biological systems are not alone in generating "order" (as covered in **chapter: 1**) which we perceive as aesthetically pleasing (e.g. stalagmites, clouds), and which sometimes can have clear functions

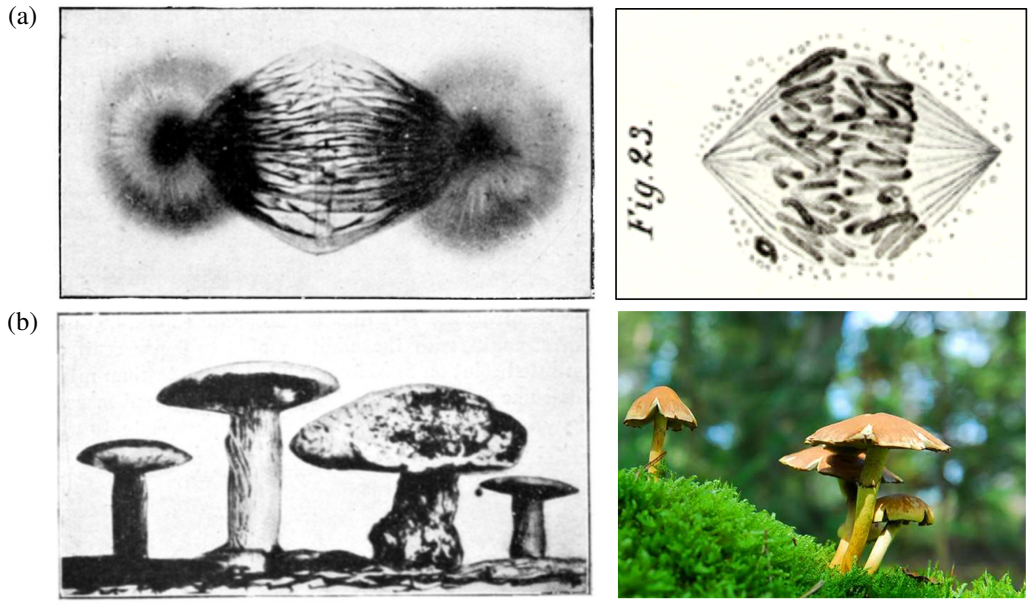


Figure 2.1 – Stephane Leduc believed that he could synthesise biological form and function. In the left panel of **a)** Leduc placed a drop of ink in the middle of the figure, and then a drop of lightly coloured hypertonic salt solution on either side. The pictures were taken after a few hours and one can clearly observe "spindles" [Leduc, 1912]. Leduc believed he was recreating what histologist Walther Flemming had described as chromosomes in salamander cells **a)** right panel [Flemming, 1880]. On a macroscopic level, he experimented with what he called "osmotic growth", by mixing CaCl2 and MnCl2 at different concentrations in alkaline solutions, he observed what he believed where synthetic fungi [Leduc, 1912] **b)** left panel.

or properties (stones, sand dunes, the sun). Our current understanding of living things requires that both form and function be reproduced and passed on through generations. Furthermore, they must adapt an organism to its environment. In the face of intense scientific scrutiny, Stephane Leduc's experiments were quickly forgotten. After Leduc, there was no more mention of Synthetic Biology until the 1980s. But the idea thereof was brewing as early as the 1970s. At the same time as Ilya Prigogine was devising his theory of dissipative systems (see **section 1.1.1**), a pair of Chilean scientists who had met in Harvard, Maturana and Varela, proposed for the first time that biological systems could be viewed as "autopoietic machines" that reproduced intrinsically [Maturana and Varela, 1972]. They provided a definition that first captured the singularity of life's drive to reproduce itself, and second to differentiate living systems from non living machines, which they would refer to as "allopoietic" (reproduced extrinsically). Maturana and Varela put forth the suggestion of "introducing perturbations in an autopoi-

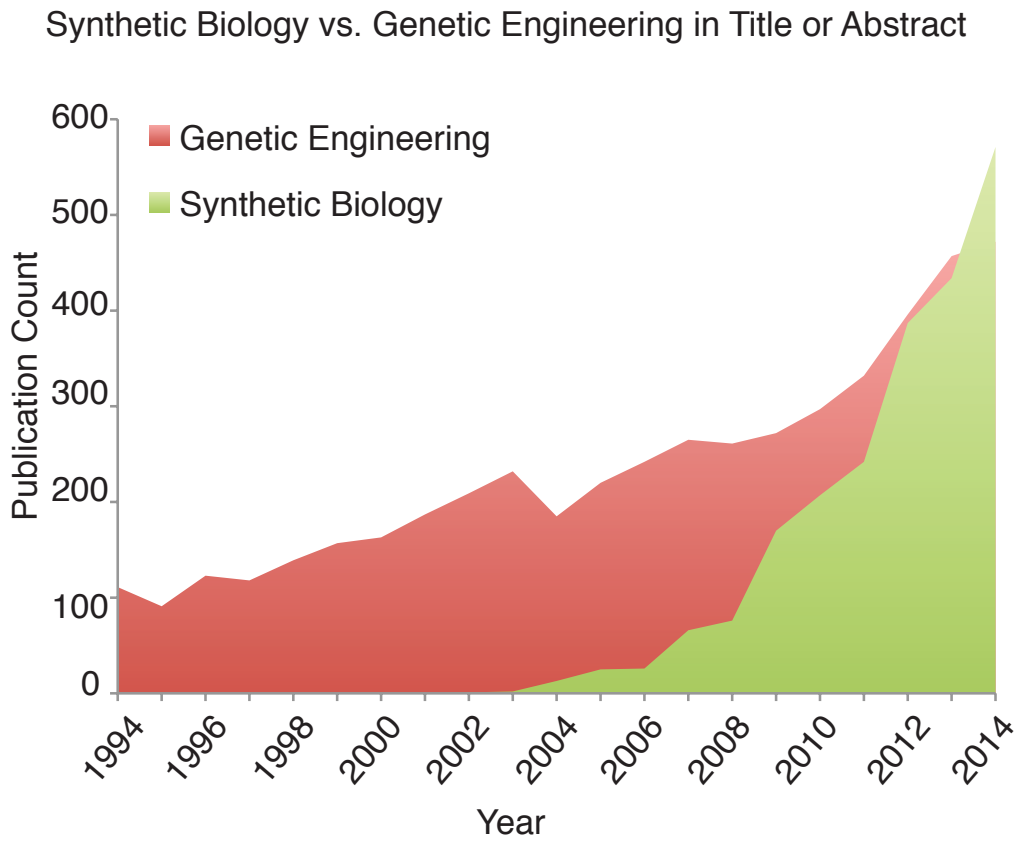


Figure 2.2 – Number of publications per year with the term "Synthetic Biology" either in the abstract or in the title. Source: **Pubmed.gov**

etic machine so as to make it allopoietic", which is very close to both Stephane Leduc's and a modern definition of synthetic biology.

Subsequently, the scientific community used the term "genetic engineering" in reference to efforts being made to treat biological life as machines. As one can see in **Figure 2.2**, it was in the middle of the 2000's when the term synthetic biology saw exponential growth. The term "synthetic biology" is nowadays used widely and even whole institutes and journals have sprung up under this umbrella. There is possibly an equally wide range of scientists doing synthetic biology who may be reticent to use this name to describe their work, thus making it difficult to put an exact number on the size of this field.

In the following chapters I will briefly review three examples that are clearly synthetic biology according to Stephane Leduc's categories of **Synthetic organic chemistry, Physiogeny and Morphogeny**.

2.1.1 Physiogeny: Genetic engineering, or changing the force within

Unbeknownst to Stephane Leduc, both biological form and function is encoded in the DNA of every living organism. In what would become one of the most ground-breaking discoveries in biology, Avery and colleagues isolated "a highly polymerised and viscous form of deoxyribonucleic acid" which was responsible for the "chemically induced alterations of **structure and function**" responsible for transforming avirulent Type II pneumococcus into the virulent Type III [Avery et al., 1944]. Years later, after this DNA structure had been chemically modeled and after DNA had been found to code for Proteins [Crick et al., 1961], these pioneering biologists discovered something even more fascinating: novel functions could be written in the language of biology, and this language was (almost) universal. Our understanding has grown substantially in the past 50 years, however it is still impossible to write completely *de novo* instructions to create simple dividing organisms [Isalan, 2012] . Thus, to further our understanding in the complexity of life, it is an imperative to go beyond the study of what is already there, and to generate novel forms and functions.

The major target of editing life is of course the DNA molecule itself, either exogenously by reintroducing it into an organism, or endogenously within the organism. Our capacity to do this has dramatically increased with the discovery and re-purposing of DNA binding proteins. Most noticeably, the clustered regularly inter-spaced short palindromic repeat (CRISPR) sequences and associated proteins, a bacterial adaptive immune system of sorts, has been shown and exploited to edit nucleated cells (**[Mali et al., 2013]**). We are now at a point where the genome might be edited just as a book might be modified with word processor.

The relative ease for a skilled person to target and modify DNA sequences inside a living cell is being harnessed to create tailor made immune cells to identify and eliminate cancerous cells [Su et al., 2016] and indeed even to eliminate whole species from this planet by tinkering with the sex chromosomes of a species (**[Hammond et al., 2016]**), which some fear might be used to edit human embryos to confer novel functions (**[Kang et al., 2016b]**). The power of this technology cannot be overstated and parallels can be drawn to the discovery of nuclear fusion to generate energy, both positive and negative. Indeed the discoverers and future users of CRISPR (whomever history will ascribe this decorum to) might yet do well to heed Robert Oppenheimer's famous words from the Bhagavad Gita after watching the detonation of the first atomic bomb: "Now I am become Death, the destroyer of worlds."

During my PhD, I considered very carefully whether to attempt to introduce the functional DNA sequences required directly into the genome through CRISPR-mediated editing. However, I stumbled upon an unexpected hurdle: given a set of

"instructions" of >3,000,000,000,000 letters plus the 3-dimensional configuration of these, it is not at all trivial to choose an appropriate position where to insert novelty without disrupting the system too much. Although 'neutral' genetic insertion loci such as ROSA26 or AAVS1 (**[Perez-Pinera et al., 2012]**) are being developed in other organisms, the canine cells we use did not have these well-defined. Therefore, this work is based on random genomic integration of synthetic genetic constructs and selection for function, rather than site-directed genome editing.

2.1.2 Synthetic organic chemistry: x DNA and x Amino Acids

Most organic compounds have been synthesized chemically in the time since Leduc lived. Using the machinery of life, including properties such as evolution, it is nowadays possible to generate novel functional proteins or alter the functions of existing enzymes ([Esvelt et al., 2011]). So what remains to be done? If one takes the basis of life as being DNA and protein, then perhaps the greatest conceptual breakthrough in recent years has been the introduction of new synthetic oligonucleotides and synthetic amino acids into a biological organisms' repertoire.

For example, scientists have recently demonstrated that not only is it possible to insert novel nucleotides into the genetic code, but that these synthetic DNA molecules (or xDNAs) are inherited given the right starting material and available machinery [Pinheiro et al., 2012]. Interestingly, further research has found that when incorporated into DNA aptamers, these synthetic bases can be tuned to specifically bind target proteins, with possible implications in cancer therapy [Kimoto et al., 2013].

On the protein side, initially, scientists used the "promiscuous" activity of tRNA-acyl synthetases to trick the cells into incorporating non-naturally occurring amino acids, that were structurally similar to their natural analogs. Later on it was shown that rarely-used codons, also known as amber codons, could be efficiently used as templates for orthogonal t-RNA/tRNA synthase pairs ([Wang et al., 2001]). After this, a deluge of novel functional unnatural amino acids has emerged, and more than 200 unnatural amino acids [Sun et al., 2014] have been successfully incorporated into bacteria, nematodes, insects [Chin, 2014] and, most recently, mammalian cells [Elsässer et al., 2016]. Functionally, there are a myriad of examples, from non-natural amino acids that fluoresce under specific protein conformations and thus help studying protein structural changes ([Summerer et al., 2006]), to amino acids that degrade under specific conditions, to study kinase pathways *in vivo* [Gautier et al., 2011]. Advances in this field have been enormous yet, as with the genome editing technologies, work remains if we ever want to achieve the fidelity of natural orthologues [Aerni et al., 2015] (reviewed in [Wang et al., 2012]).

2.1.3 Morphogeny: from transplantation to 3D printing of organs

As we have seen in the examples mentioned in the two previous sections, it is possible to generate novel chemical form and function within biological systems. If we refer back to the holoarchy of communication (**see Figure 1.2a**, we observe that synthesis at higher levels will affect a system more profoundly, as this is impossible to do without modifying lower level parts.

Form is then a spatial property of, or within, an object (for example the shape of a hand) and, more deeply, the DNA sequence which holds the instructions for the hand to form. Function can be seen as the temporal dimension of the object and in some cases the interaction of the form with its environment. In the example of the hand, this will be the specific function of the appendage during a specific timeframe, in a specific environmental context. Thus, form and function are inseparable properties, but also independent from each other.

From the above definitions, it follows then that morphogeny may be subcategorised further. Firstly, one can attempt to mimic form and functions that already exist. An example of this is the synthesis of organs in vitro, such as skin to treat burn injuries, or recent advances in 3D printing of cells together with scaffolds to create long-lasting transplantable organs [Kang et al., 2016a]. The second aspect of morphogeny would deal with generating completely novel forms and functions such as gliding jelly-fish made of muscle cells, as shown in [Nawroth et al., 2012].

Major questions in biology remain not fully explained: how can one cell, containing a set of DNA instructions, have all that is needed to create living forms? How can some of these cells retain their functions even when explanted? Substantial advances to answer these questions have been made by studying biological forms outside their context. We now know that almost any cell in our body can be reprogrammed to its primordial stage or to any other cell, moreover these cells can later on be steered to generate organoid structures which resemble organs morphologically (**reviewed extensively in: [Lancaster and Knoblich, 2014]**).

Novel forms and functions can also be generated artificially. In my thesis I have worked with synthetic biological organoid structures, as cells cultivated in extracellular matrix (ECM) grow in different forms to cells grown in a plastic dish. The distinct forms appear to interpret the function I programmed in their DNA differently; however, even when new forms were generated (see: **section 6.3.2**) remnants of the the underlying genetic functions were observed. In the final Chapter, I will describe in detail the genetic function that I aimed to engineer.

Chapter 3

FEEDBACKS: FUNDAMENTS OF LIVING SYSTEMS

The purpose of this work is to use the conceptual framework provided by synthetic biology (Chapter 2) to expand our understanding of the underlying rules and principles of form and function in dissipative systems (Chapter 1). It is an attempt to generate a system with both microscopic and macroscopic function. I will show in the results section, how I have generated new biological functions with information encoded in the form of DNA, and novel forms by growing cells under different environmental conditions. Functions encoded in DNA are conserved when cells are grown in specific forms. Choosing a well-characterized system was crucial for my attempt at synthesizing new functions. It is in the synthetic biologist's best interest to use this cumulative knowledge of the past to explore beyond forms and functions.

In this chapter, I will expand on the importance of positive feedbacks in systems. The main objective of my four year pre-doctoral research has been to design and characterize such a topology in a mammalian cell system. As I described in Chapter 1, feedbacks seem to fulfill a fundamental function in systems: to transform and translate the energy from outside a system into order and structure within. In information theory, feedbacks can confer systems with properties such as memory, signal amplification and digital-analog signal conversions (section 1.2.1). This chapter will lead directly into the Results section of this thesis, where I will show the experimental work to create and characterize a synthetic positive feedback in a mammalian cell line.

But first, what defines a feedback exactly?

3.1 General definition of a positive feedback

A feedback can be defined as an autocatalytic reaction of two components X and Y, whereby Y gets converted to X, such as:

$$X + Y \rightarrow 2X \tag{3.1}$$

with rate equations such that:

$$[X] = -p[X][Y] + \mu[X]^2 \tag{3.2}$$

$$[Y] = p[X][Y] - \mu[X]^2 \tag{3.3}$$

If one plots the concentration of X over time, one will observe that the increase of X is non-linear, namely it will increase exponentially *ad infinitum*. There is of course a dependence on Y, and a reaction with unlimited supply of Y would be an infinite positive feedback. In contrast, a negative feedback is an autocatalytic reaction where one of the reactants inhibits the production of the other. Ilya Prigogine proposed a reaction scheme called the Brusselator which contains both positive and negative feedback:

$$\begin{aligned} 2X + Y &\rightarrow 3X \\ A &\rightarrow X \\ B + X &\rightarrow Y + D \\ X &\rightarrow E \end{aligned} \tag{3.4}$$

As we can see, in this case the concentration of X is not expected to grow exponentially, as it is transformed into Y and D in the third step. Thus, feedbacks can also serve to control the amount of order in systems.

Feedbacks are an efficient way for a closed system to increase utilization of available information and energy. It is likely that when multicellular organisms first evolved, the majority of cells' immediate surroundings contained other cells; cells had to communicate in a coordinated manner to produce responses to external stimuli and so naturally fed back information. Eventually, when multicellular organisms began communicating with each other, feedbacks guided the way towards a balance between self and environment. Instead of inventing a new mechanism for transmitting this information, biological systems became masters of controlling and utilizing feedbacks.

In technology, feedbacks have become pivotal in the last two centuries. As well as guiding our machines, they have given us a better understanding of reality

by helping us grasp non-linear dynamics. Once we learned how to tame electrical signals with feedbacks, we were quick to transmit messages across enormous distances with unimaginable speeds and reliability. Furthermore, the increasing processing power of interconnected devices is allowing engineers to develop ever more precise feedbacks that help existing devices become more efficient. For example, the Swedish miner LKA utilizes trains that feed back energy generated when going downhill (with full loads of iron ore) towards a port. This mechanism makes the net energy of the round trip of the train essentially close to zero.

Feedbacks are time-scalable and inherent to systems which generate information through the consumption of energy. But, as we shall see, it is not trivial to put a quantitative measure on the magnitude, or velocity, of a feedback. Particularly in molecular and cell biology, there has been a focus on studying feedbacks that are strong in nature, and thus easy to define, analyze and synthesize. As I will present in the next sections, studies in other disciplines demonstrate that weak feedbacks can have measurable effects on systems. Moreover, these feedbacks display certain additive features which could be possibly mimicked through synthetic biology. Thus, observations from reproducible feedback systems can potentially be translated to other scientific disciplines which study these motifs.

In fact, one of the main inspirations for the present work is that the importance of these motifs are being recognized as critical in two other scientific disciplines which have a high impact on civilization and which also study systems: Climate and Finance. As these systems do not lend themselves to easy manipulation and synthesis without affecting the lives and livelihoods of people, it will be useful to understand and study feedback mechanisms in organisms so as to gain a better control of the world we live in. Vice versa, methodologies used in other scientific fields can be used to attempt to describe feedbacks in biology.

3.2 Positive feedbacks in climate

The Earth's climate is an example of a dissipative system with different communication hierarchies (see figure 1.2b). The planet receives radiative energy from the sun, and reflects this radiation in the form of excess heat. Towards the end of the 19th century, Svante Arrhenius noticed that for the planet to have the temperature to support life, some of that energy must be captured in the atmosphere and not reflected, thus came the discovery of the so called greenhouse gases (such as CO_2) [Arrhenius and Holden, 1897]. What Arrhenius and his contemporaries could not know at the time, was that this captured energy, also gave rise to multiple "ordered" time-variant structures, which we observe as weather, climate and environmental landscapes.

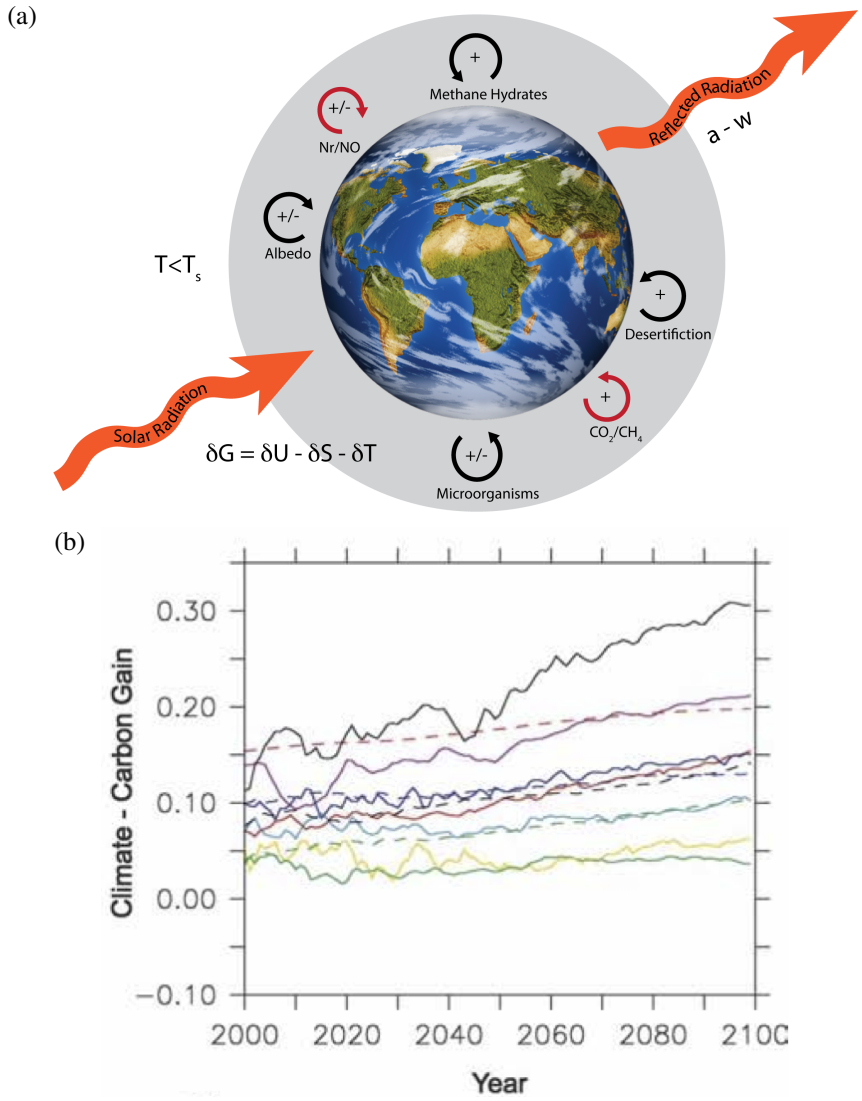


Figure 3.1 – Climate Feedbacks between Temperature and Carbon emissions. A schematic depiction of Earth's climatic system with some of its feedbacks is shown in 3.1a. The signs inside the feedbacks are taken from the literature. The two feedbacks shown in red which correspond to the Carbon and Nitrogen cycles, are heavily influenced by anthropogenic action. The 11 models compared in [Friedlingstein et al., 2006] are shown in 3.1b, which show the increase in the "gain" factor between temperature and climate within the next century.

As a dissipative system, we expect the climate to contain multiple feedback mechanisms at any of the hierarchies. Understanding these feedbacks is of utmost urgency for our species, as we have at last reached population and greenhouse gas

emission levels which directly influence the global climate system. Exceeding thresholds of which we know very little, could lead to a potentially catastrophic "runaway feedback". On the upside, this also means that we can, for the first time in our history, directly predict and possibly intervene in near-to-long term future global climate scenarios.

The United Nations International Panel on Climate Change (IPCC) was conceived to warn us that anthropogenic carbon emissions are leading towards an unknown path of planetary warming. The IPCC and, more recently, the COP21 nations have set specific temperature rise boundaries beyond which it is believed that life on earth as we know will be unviable. One of the main reasons these limits or thresholds were set, is the general respectful fear of the power of feedbacks. Unfortunately, scientists are warning of the thermodynamic reality of our planet: policies to mitigate emissions have been lackluster and are highly influenced by economic principles based on infinite growth, and prosperity through energy consumption.

3.2.1 Holoarchical feedbacks in climate

The debate on feedbacks in the climate system is not centered around whether these exist but around the accuracy of quantitative measurements of the feedbacks' contributions to the balance of temperature and carbon in the planet. The first proposition to measure the strength of these feedback mechanisms in climate was put forth by James Hansen in [Hansen et al., 1984]. Here, Hansen describes what could be referred to as the most basic holon of the climate system. In this seminal paper, Hansen defines several mechanisms which might exert temperature feedbacks on the Earth's climate, some of which are shown in figure 3.1a. Hansen proposes a simplified model to capture these feedbacks in which the earth's temperature is at equilibrium ΔT_eq ; that is there is a constant difference in radiated energy received and reflected from the sun. This measure can be decomposed to the Earth's temperature at departure from equilibrium ΔT_o, plus the addition of some temperature feedback $\Delta T_{feedback}$ so that:

$$\Delta T_{eq} = \Delta T_o + \Delta T_{feedback} \tag{3.5}$$

From this, Hansen proposed two parameters. Firstly, the feedback factor f which defines the relation of ΔT_{eq} to ΔT_o, and a gain factor which describes the ratio between the feedback contribution and the equilibrium temperature.

$$\begin{aligned} \Delta T_{eq} &= f\Delta T_o \\ g &= \frac{\Delta T_{feedback}}{\Delta T_{eq}} \end{aligned}$$

A feedback is then positive when $f > 1$ and $g > 0$. Thermodynamics were originally devised to describe the conversion of heat into work. As we have seen in previous chapters, the relationship g was derived by Hansen from its analogue in thermodynamics, namely the thermal efficiency η, which measures the efficiency between output and input energy in a heat engine, and is also known as Carnot's theorem:

$$\eta = \frac{W}{Q_E} = 1 - \frac{T_C}{T_H} \tag{3.6}$$

Here, W is the work a system performs and loses through heat, and Q_E is the input energy into the system. This is equal to the difference in temperatures from a cold reservoir T_C and a hot reservoir T_H.

Most IPCC climate models have at their cores this proposition by James Hansen. In the latest IPCC report, several coupled climate-carbon models from different scientific institutions around the world were compared. What they found is indeed worrying: independent of the complexity of the model, when adjusted to data measurements, all models simulated positive feedback [Friedlingstein et al., 2006, Friedlingstein et al., 2011] and figure 3.1b. Furthermore, all models predict an increase in the feedback strength in the coming decades. And yet, the summary report of the IPCC emphasizes that the *"Confidence in the magnitude, and sometimes even the sign, of many of these feed-backs between climate and carbon and other biogeochemical cycles is low."* [Ciais et al., 2013].

Thus, although a feedback of the positive kind in temperature (a high temperature at time t leads to a higher temperature at $t + 1$) can become the sole driver of temperature increase, negative feedbacks are also at play.

The important message in the IPCC reports is that failure to reduce the anthropogenic contribution to these feedbacks, could tip the balance towards runaway positive feedback which could push the planetary temperatures away from the relative stability experienced during the holocene [Clark et al., 2016]. Unfortunately, little is being done in this respect, as the mantra of societal well being through economic prosperity continues to drive policy.

Although some of the feedbacks in figure 3.1a are directly linked to human action, some are clearly not. It is also not clear what the "strength" of these individual feedbacks is. This parameter should however be strongly taken into consideration, as some of these feedbacks such as methane hydrates being released have led to "abrupt climate change" in the past (this being an average temperature rise or decrease over a very short geological time-scale of 100-200 years). Although any reference to this term is absent from IPCC reports or COP21 summaries, evidence of recent methane plume activation from 500m depth sediments off the North American coast have been observed and linked to a rising global temperature [Johnson et al., 2015]. On the other side, cloud cover if more heat,

more evaporation, more clouds. if clouds form horizontally then less sun, negative feedback. if clouds form vertically. more sun

Like climate models, our understanding of natural biological feedbacks is hampered by complexity and difficulties in quantification. Here, synthetic biology can allow us to find a compromise between setting up controlled hypotheses and having a 'black box' signalling environment. Analysis of feedbacks in biology have focused mainly on examining the effects of strong and clear feedbacks in noise modulation ([Dublanche et al., 2006]) and analog to digital signal conversion ([Daniel et al., 2013]). To my knowledge, there has not been any attempt at characterizing the *strength* of feedbacks. Thus, in the results section I will try to use the methodology from climate scientists to measure the gain g and feedback factor f in a biological system.

3.3 Positive feedbacks in Economy and Finance. The Minsky moment

The financial system consists of an intricate network of exchange transactions of services and goods. The agents in financial systems are human beings and their actions and interactions with their environment are the links. Fundamentally, financial systems are like any other system, transformers of energy into entropy. The energy, once again, ultimately comes virtually all from the sun. Along human history has been utilized in different forms, but has come predominantly from carbon-based storage forms such as biomass, coal and, in its most recent iteration, petroleum [Duran and Reyes, 2014]. Thus, growth in economic terms is closely linked with the amount of energy being consumed.

The conversion efficiency by which energy is translated into order by economic feedbacks takes the form of monetary value. Importantly, as I have discussed in previous sections, it is not possible to reverse this process, i.e. generating monetary volume or debt by printing more money cannot by itself generate more energy. This result is is true with one exception: if one decides that time is no constraint, and one initiates the activation of the feedback loops into the future, in a manner defined by Hyman P. Minsky. According to Minsky, there are three distinct forms of financing structures, which can nonetheless all transition into each other. A first form is one in which the future payments are made by a constant income stream and is called a "hedge". A second is in which the future payments are not covered by a constant income stream, which he called "speculative". Finally, a third exists in which future payments are only assured by issuing new commitments, known as "Ponzi". The proposition goes as follows: an economic corporation will convert energy into a product (i.e. reduces

entropy) which is assigned a specific economic value (output). If this scheme shows a profit during a business cycle then there is an expectation that it will generate a profit during the next business cycle and thus, if as Minsky proposed, "Investment determines output, output determines investment," investment will tend to flow towards these corporations. This investment will generate an increase in output and an initial financial "hedge" will be attractive. If the cycle continues, then individuals can take higher risk (more entropy) and invest in a "speculative" manner, and even in a "Ponzi" scheme. Thus, the output of the corporation will be diverted more and more into fulfilling financial commitments at an increasing uncertainty or entropy. It is at this point that Minsky believe that this positive feedback-grown system is at its weakest, and that surprise events could expose this fragility, thus causing the whole interlinked system to come down. Indeed, one of Minsky's key motivations was to answer the question whether an economic depression - such as the one experienced during the 1930's - could happen again [Toporowski, 2005, Minsky, 1982]. The problems have biological parallels in ecology and ecosystem modelling, although it is not yet clear to what extent competing cells or organisms 'borrow from the future' in order to survive natural selection in the now. Nonetheless, the tools of economics have the potential to be applied in biology in a manner that has not yet been sufficiently explored.

3.3.1 Autocorrelation

Economics is a statistical science. This implies that economics is data-rich and mechanism-poor: features become visible only above a certain number of observations. Correlation functions are statistical measures by which one can estimate the similarity of two time-dependent noisy functions and include the "autocorrelation" function. Let R be a function for an autostationary process X, then the autocorrelation function will be:

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2} \tag{3.7}$$

This function is valid for processes with standard variation $0 < \sigma < \infty$. The time variable τ is referred to in the literature as the "lag" between a time t and $t + i$. The result for each lag is an indicator of similarity between -1 and 1 which will be *positive* for correlation and *negative* for anti-correlation.

As stated, the autocorrelation function is used in economics, and can be understood, for example, by the statement that a high GDP at time t will most of the time imply a higher GDP at time s. This function is applied to identify such motifs in noisy economic data and it can also be used to search for oscillatory behaviors (for a biological example, see **Figure 3e** in: [Stewart-Ornstein and Lahav, 2016]).

The specific decay of the autocorrelation function can give an indication of the duration of memory in a process. An important measure to estimate this effect is the $\frac{\tau}{2}$. For example, as the lag increases, the correlation of a noisy function will tend to decrease. In processes which contain positive feedback, it is thought that this decay occurs slower, and thus one can expect a higher half-autocorrelation lag, or longer lasting memory. This can have implications when it comes to making future predictions about time correlated data, such as those presented in the Results section of this thesis.

3.3.2 Hurst Exponent

A second interesting measure used in financial time-series is the Hurst exponent of a time series ([Peters, 1996]). Named after the hydrologist Harold Edwin Hurst, who was exploring methods to reliably predict future water usage in the Nile delta by looking at time-series from the past, to do this, he developed a method by which he could establish long term correlations [Hurst, 1956].

To do this, Hurst proposes to re-scale the range of the data he collected (which included water-gauge levels which went back hundreds of years) to the mean over the whole observations. Dividing the rescaled time series by the standard deviation of the measurements:

$$H = \frac{R_t}{\sigma_t} \tag{3.8}$$

In contrast to being a completely random process, Hurst unexpectedly discovered that there were long-lasting memory effects in the data, namely, if the reservoirs in the Nile had more water than the mean over a very long time period, then a tendency existed to accumulate more water over time, conversely, if the water levels were below the mean, then the tendency would be towards less water in the future. Thus, by comparing how a function relates to itself, Hurst looked at the change in the cumulative deviations from the mean in time and found a surprising correlation.

The regularities that Hurst uncovered had a further property, namely that they were scalable. It took a few decades until re-known mathematician Benoit Mandelbrot applied the methods used by Edwin Hurst to describe the laws governing the self-similar scalable functions he developed. It was Mandelbrot himself who named the exponent after Edwin Hurst.

There is a simple relationship between the Autocorrelation function and the Hurst exponent [Lillo and Farmer, 2004]. For long-term memory processes, the decay of the correlation function p with lag k degrades as a power law with exponent α such that:

$$p(k) = xK^{\alpha} \tag{3.9}$$

Then the Hurst exponent can be defined as:

$$H = 1 - \frac{\alpha}{2} \tag{3.10}$$

Regardless of the method to estimate the Hurst exponent, the values will vary between 0 and 1. In the financial literature it is generally assumed that a Hurst exponent below 0.5 is a sign of negatively correlated time processes, a Hurst exponent above 0.5, can show positively correlated processes, and the exponent is used to detect long-term "memory" trends in economic time series [Weron, 2002, Tom and Andrew, 1991]. As with the autocorrelation, the Hurst exponent might help unravel hidden relationships in biological time-series. Furthermore, in contrast to financial time-series, biological experiments can be reproduced, and modified, so as to be used to test the validity of these widely used methods in finance.

3.4 Positive Feedbacks in Biology

3.4.1 Naturally occurring feedbacks

I have briefly mentioned the importance of feedbacks in the field of synthetic biology to generate memory devices (section 1.2.1). Examples of naturally occurring feedback mechanisms also abound, and the functions are as expected very wide, and operate at different scales. Intracellularly in [Lahav et al., 2004] the authors show that mammalian cells are constantly checking for DNA damage and feeding back the information through the MDM2-P53 interaction. This interaction occurs at a nano meter scale, however feedbacks have been shown to work within bigger cells such as Xenopus Oocytes (diameter of $600\mu m$ compared to $10\mu m$ in epithelial cells used in the experiments of Lahav and collegues) to coordinate mitotic entry and thereby achieve supra diffusion velocities [Chang and Ferrell, 2013].

Feedbacks operate at multicellular levels too, to ensure tissue homeostasis, relaying the positional information ([Dubuis et al., 2013]) to cells, and stabilizing the chemical gradient patterning cues given by chemical molecules during development. Surprisingly, in biological systems the receptor and ligand are rarely produced in the same cell; the case is quite different in synthetic biology.

3.4.2 Positive feedbacks in synthetic biology

Positive feedbacks have been synthetically generated in biological systems. Many of these feedbacks have been generated using well known trans-activators and promoter sequences. The goals or utility to generate these feedbacks vary from group to group, for example in [Daniel et al., 2013], the authors generated a feedback to significantly increase the dynamic range of response in a bacterial strain by computing an input signal in an analog as opposed to digital manner. Similar conclusions were reached in [Becskei et al., 2001], where the authors introduced stability into a noisy gene network by introduction of positive and negative feedbacks. Others such as [Ajo-Franklin et al., 2007], have exploited the intrinsic bimodality of positive feedbacks, to construct a memory device in *Saccharomyces Cerevisae*. This approach is particularly interesting, as it can present a non-genetic mechanism for the transmission of information across generations. A more recent study has used S. Cerevisae, to design and characterize the strengths and signs of synthetically designed feedbacks by using a feedback factor "F" that accurately describes the sign and strength of the feedback [Schikora-Tamarit et al., 2016].

Only a few authors have ventured into generating positive feedbacks in mammalian cell cultures (for example [Siciliano et al., 2011]) and none to my knowledge have explored the effect of feedbacks in organoid cultures. The reasons for this lack of mammalian cell studies might be varied. The most likely contributor is the technical laboriousness and cost of doing genetic studies in mammalian cells, which might hamper the reproducibility of experiments across different labs.

Interestingly, no mention of a general measure for feedback contribution to the systems is made in any of the works presented above and others. Such a dimensionless measure could be of utmost relevance to compare feedbacks across organisms and scales. In Part III, I will suggest a general measure to describe feedbacks.

Chapter 4

SCOPE OF THIS WORK

In this Introduction I have written mostly about mathematics and relatively little about biology. This has been a conscious decision because, although I have learned from both approaches and later chapters focus heavily on biology, it is important to put the basis of this thesis in a broader philosophical context. This basis is to further our understanding into what the basic elements of spontaneous order are.

The work presented here is synthetic in nature, and thus should fit in the broad field of synthetic biology (**Chapter 2**). In this sense, it was firstly an engineering approach, whereby the biological questions would come once the system was engineered. Although, perhaps unconventional for the investigative biologist, sometimes building something can help us better understand this process. For example, the laws of aerodynamics were not established by watching birds fly, but by building airplanes.

In this thesis I have focused on studying and using the Hepatocyte Growth Factor (HGF) [Montesano et al., 1991] as a transmitter of information. As with other growth factors, HGF is produced and secreted by mesenchymal cells and is sensed by a sole known membrane bound receptor c-Met ([Park et al., 1986]) in epithelial cells. The HGF <-> c-Met interaction leads to a signaling cascade within cells, and triggers diverse phenotypes mostly associated with disruption of tissue architecture and increased cell motility. During development, this signaling pathway is essential in the coordinated transition of epithelial cells to mesenchymal cells. For example, the knock-out of HGF in mice embryos renders them non-viable after days 13.5 and 15.5 [Uehara et al., 2000].). In cancer, it has been shown that aberrant increase in HGF-c-Met signaling can substantially increase the metastatic potential of cells (reviewed in [Mizuno and Nakamura, 2013]). This phenotypic consequence has also been exploited in an example of inter-kingdom "hack" by the bacterium *H. Pylori* through the Cag-A to c-Met intracellular interaction which leads to morphological changes in host cells that are beneficial for the bacterial

infection to progress ([Tan et al., 2009, Bagnoli et al., 2005]).

In the work presented here this HGF-c-Met signaling was exploited to synthesize a sender-receiver cell line with a measurable HGF-dependent output, in the form of GFP driven by a matrix metalloproteinase promoter (pMMP1). MMPs have been shown by ourselves in **Part II** of this work and others to be differentially upregulated by an HGF stimulus in MDCK cells [Chacon-Heszele et al., 2014, Hellman et al., 2008]. To obtain the HGF-receiver cell line presented in chapter II I generated a cell line with a genomically integrated pMMP1-d2EGFP DNA sequence (henceforth called the **ACS1** cell line).

The goal of this thesis is to investigate whether the creation of an auto-catalytic gene signaling network can generate macroscopic spatio-temporal patterning. For this, in **Part II** I first engineered a mammalian cell line to read-out the spatiotemporal diffusion of the morphogenic Hepatocytic Growth Factor (HGF) protein. Once this cell line had been established, using the same genetic components, in **Part III** I introduced an inducible autocatalytic feedback loop into this cell line. To confirm that the feedback loop was functioning, I applied some of the time-series analysis methods described above. As I will show, some of these methods succeed in identifying a positive feedback in the data whereas others do not.

The results of **Part II** have been published in a peer-reviewed journal (see [Carvalho et al., 2014]) and have been included here. Further discussion of some of the topics presented in this work, have also been published in form of a review in [Barcena Menendez et al., 2015].

Part II

A genetically-encoded sender-receiver system in 3D mammalian cell culture. ACS Synthetic Biology 3 (5), 264–272 (2014).

Carvalho A, Bárcena D, Senthivel VR, Zimmermann T, Diambra L, Isalan M. Genetically encoded sender-receiver system in 3D mammalian cell culture. ACS Synth Biol. 2014 May 16;3(5):264–72. DOI: 10.1021/sb400053b

Part III

Generation and characterization of a positive feedback cell line

Chapter 5

INTRODUCTION

In vitro tissue engineering has been recognized to hold the potential to revolutionize medicine. Still, a fundamental unresolved challenge remains: how to deliver nutrients and signaling molecules to masses of tissues generated *in vitro* in the absence of vascularization? During development, an active vascular network only starts to develop between the 5th and 6th week after gestation in humans, and roughly at days 7-8.5 in mice, leading to a fully functional murine heart at day 13 [Kaufman and Navaratnam, 1981] and [Kau, 1994]. At this stage, the embryo consists already of billions of cells which act in coordinated manner to position themselves both in space and time to form tissues.

For synthetic biologists, there are currently two approaches to engineering tissues, both of which I described in detail and presented examples in **Chapter 2.1.3**. The first, top-down approach uses biomimetic scaffolds designed specifically so that cells will follow physical cues and assume their biological potential. The second approach is a bottom-up approach, in which a starter colony of cells is led to develop and construct all necessary structures without the need for a "pre-printed" scaffold.

It is unclear which approach is closer to nature, regardless, robust signaling mechanisms which transform disordered energy from the environment into ordered forms and structures must exist (see **Section: 1.1.1**). Mammalian cells, for example have a vast repertoire of pathways to signal within themselves, with other cells, and with the environment. The communication agents are varied and include structures such as exosomes, RNA molecules, small chemicals, proteins, or through direct sensing of physical cues.

In the following section (6.1) I will present experimental results expanding on my findings of Chapter II. The goal was now to expand the system by introducing a positive feedback. To do this, I introduced into the ACS1 cell line HGF itself and a red fluorescent protein (tdTomato) under the control of the pMMP1 promoter, thus generating a synthetic autocrine feedback loop. There were two main moti-

vations to generate this system: first, it is biologically relevant to study the synthetic cell line, due to the aforementioned developmental and carcinogenic effects of HGF; to my knowledge there are no reported cases in which HGF and c-Met are expressed in the same cell. Secondly, from a dynamical systems perspective, building such a system could give insights into the consequences of autocatalytic reactions in naive systems, putting to the test their entropy-generating capacity.

Chapter 6

RESULTS

6.1 Synthesis of a positive feedback cell line

6.1.1 Design

The promoter for MMP1 is activated by HGF in MDCK cells grown in 3D collagen cultures (**Part II**). I now set out to introduce the signaling molecule HGF itself under the control of this promoter, to generate a "positive feedback" cell line (**PF Cell line**). To abrogate the expected strong mitogenic properties of HGF - hypothesizing that a constant stress could be detrimental to cellular health - I introduced a switch-like control via tetracycline-dependent repression (**Figure 6.1**). Furthermore, to be able to measure directly the positive feedback contribution to the signal, a second PEST destabilized fluorescent reporter (d2tdTomato) was introduced and linked to the HGF cDNA sequence via a self-cleaving P2A peptide. This module should allow me to monitor the dynamical nature production of HGF by cells. Lastly to monitor the influence of the positive feedback, this module was inserted into the ACS1 line, allowing me to compare results in a single monoclonal cell line. This strategy would allow me to eliminate potential genetic or epigenetic sources of uncertainty from the system. The added caveat is that there could be a potential competition for transcription factors between the two pMMP1 promoters.

Thus, there will be two sources of HGF throughout this work with corresponding nomenclature: HGF applied exogenously will be referred to as HGF_{ext} and was primarily purchased from Sigma-Aldrich® unless specified otherwise. On the other hand, HGF produced by cells from the positive feedback module plasmid will be referred to as HGF_{int}.

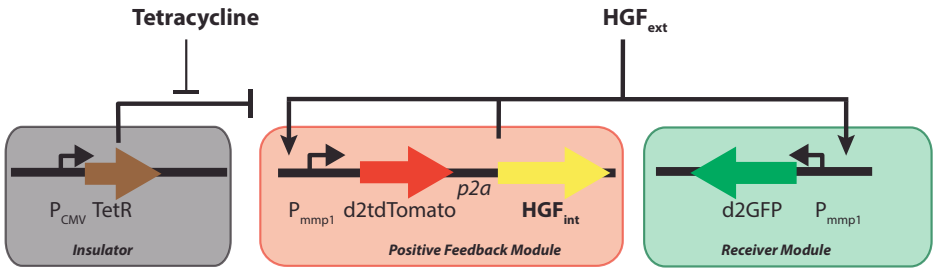


Figure 6.1 – Schematic representation of the plasmid modules used to generate a cell line with both an HGF receiver module (green) and a positive feedback *(PF)* module (red). The PF module has a tet-operator sequence, and is kept in an *OFF* state by the insulator module (gray) which produces Tet Repressor. Addition of tetracycline should switch the PF module to an *ON* state. HGF and tdTomato in the PF module are transcribed as a single transcript linked by a P2A peptide and translated from a single ATG.

6.1.2 Cell line creation

I introduced the plasmids containing the modules shown in **Figure 6.1** into the receiver cell line used in [Carvalho et al., 2014] by Lipofectamine® transfection. I then submitted cells to several rounds of antibiotic and fluorescence selection to generate a monoclonal cell line (details are shown in the materials and methods **section 8.2.2**). I was specifically looking to generate a cell line which displayed high GFP/Tomato expression, upon addition of tetracycline. I observed that immediately after transfection, a high percentage of cells had this phenotype (**figure 8.1a**). I initially attempted to sort these freshly transfected cells into single wells of a 96-well plate. Using this method, I found that 1) the survival of these colonies was very low, and 2) colonies that did survive, lost the desired phenotype after culturing them for prolonged periods of time. To overcome these issues, I performed a two-step sorting strategy by first sorting cells in bulk, and selecting these with an antibiotic. After cells were selected for around 7 days, I subjected them again to treatment with tetracycline and sorted this time single cells into 96-well plates. This approach allowed me to generate positive clones.

In accordance to what I did with the receiver cell (**Part II**), clonal colonies were tested in 2D for activation of the PF module, and further tested in 3D-Collagen cysts. This was necessary, as I had previously observed that even though cells might respond to HGF_{ext} when grown in 2D-monolayers, this did not immediately translate to a response to HGF_{ext} in 3D cultures. Out of >200 tested clones, I found only one which responded to the HGF_{ext} cue. In **Figure 6.2** I show fluorescent imaging for this positive clone grown in 3D collagen, observed under a fluorescence microscope. Positive traits which are consistent with the fact

that this clone has autocrine HGF signaling:

–Low GFP and tdTomato expression in the absence of HGF_{ext} (6.2 Row 1)

–GFP translation and tubulation in response to HGF_{ext} (6.2 Row 2)

–Production of GFP and tdTomato in the presence of tetracycline (6.2 Row 3)

–Activation of the GFP and morphological response to HGF_{int} (6.2 Row 3)

–GFP and tdTomato expression in response to HGF_{ext} in the presence of tetracycline (6.2 Row 4)

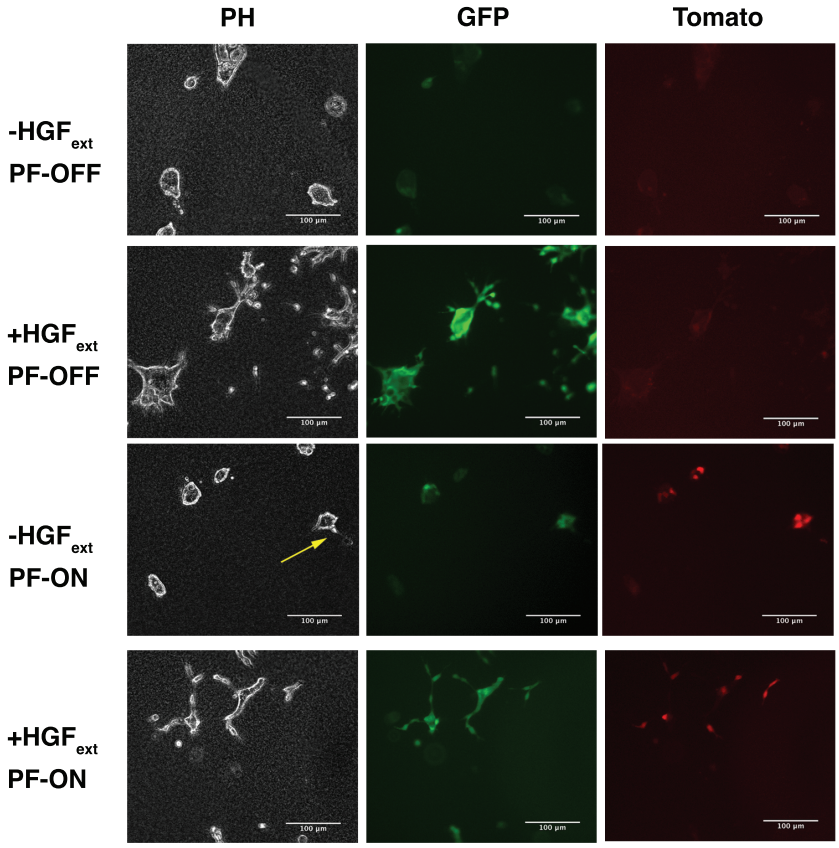


Figure 6.2 – The positive feedback (**PF**) cell line responds to extrinsic and intrinsic HGF. Images of a clone generated through the FACS sorting method described in **Figure 8.1a**. PF cells were grown as cysts in collagen for 7 days and were then stimulated for 24 hours with a dose of 50ng/ml HGF ($+HGF_{ext}$) and imaged in both GFP and tdTomato channels. Additionally, cysts were also imaged in the absence (**PF-OFF**) and presence (**PF-on**) of 1 μg/ml tetracycline. The yellow arrow indicates cysts tubulating in the absence of HGF_{ext} but presence of tetracycline indicating the production of HGF_{int}.

Of these traits, the observation of tubulation in the absence of HGF_{ext} is the clearest indication that functional signaling is occurring **Figure 6.2 3rd row arrow**. Furthermore, this was a confirmation that the 'ground state' of the system without repression is leaky and on. The elevated GFP levels observed in this condition further indicates that the receiver module is also sensing the produced HGF_{int}.

To quantitatively assess the activity of this cell line, I analyzed the images with a modified version of the custom Matlab® script used in **Part II** (see Materials and Methods section 8.3.2 for detailed explanation). This code allowed me to analyze the images in high throughput and thus estimate the percentage of cysts expressing GFP and tdTomato for a given experimental condition. In **Figure 6.3** Comparative results of the PF cell line grown as cysts versus cells grown as 2D monolayers and stimulated to HGF_{ext} or tetracycline. I can confirm the observation from the ACS1 receiver line (**Part II**) that activation of the pMMP1 promoter by HGF_{ext} is stronger in 3D cysts than in monolayers (**Figure 6.3** top row). I can also deduce that de-repression of the positive feedback induced primarily tdTomato expression, while GFP signaling could be detected in only about 4% of cysts under these conditions (**Figure 6.3** third row). In the presence of HGF_{ext} and tetracycline (**PF-ON**), GFP and tdTomato were both upregulated, indicating that although the promoters are independent of each other, both respond to the external HGF stimulus.

One of the main objectives of this thesis is to generate and study emergent "macroscopic" intercellular patterning using intracellular microscopic "components". For such patterns to arise in the presented MDCK system, it is indispensable that cells secrete the communication signal HGF_{int} into the medium. In **Figure 6** of **Part II**, I showed that the ACS cell line can be used to measure a diffusible HGF_{ext} signal and that this signal could be inhibited through the NK4 inhibitor. In these experiments, both HGF and NK4 were in fact produced in MDCK cells, albeit these cells had been transfected with a constitutively expressed HGF plasmid. In the PF-cell line, I observed phenotypically that HGF_{int} was being produced and sensed by at least a fraction of MDCK cysts, as was evident from the tubulation of some of the PF-cysts in the presence of tetracycline, and the activation of the GFP reporter. It was however unclear whether HGF produced under the weak mmp1 promoter would be secreted into the medium, and given the low levels of activation observed in **Figure 6.3**, I decided to pursue this further.

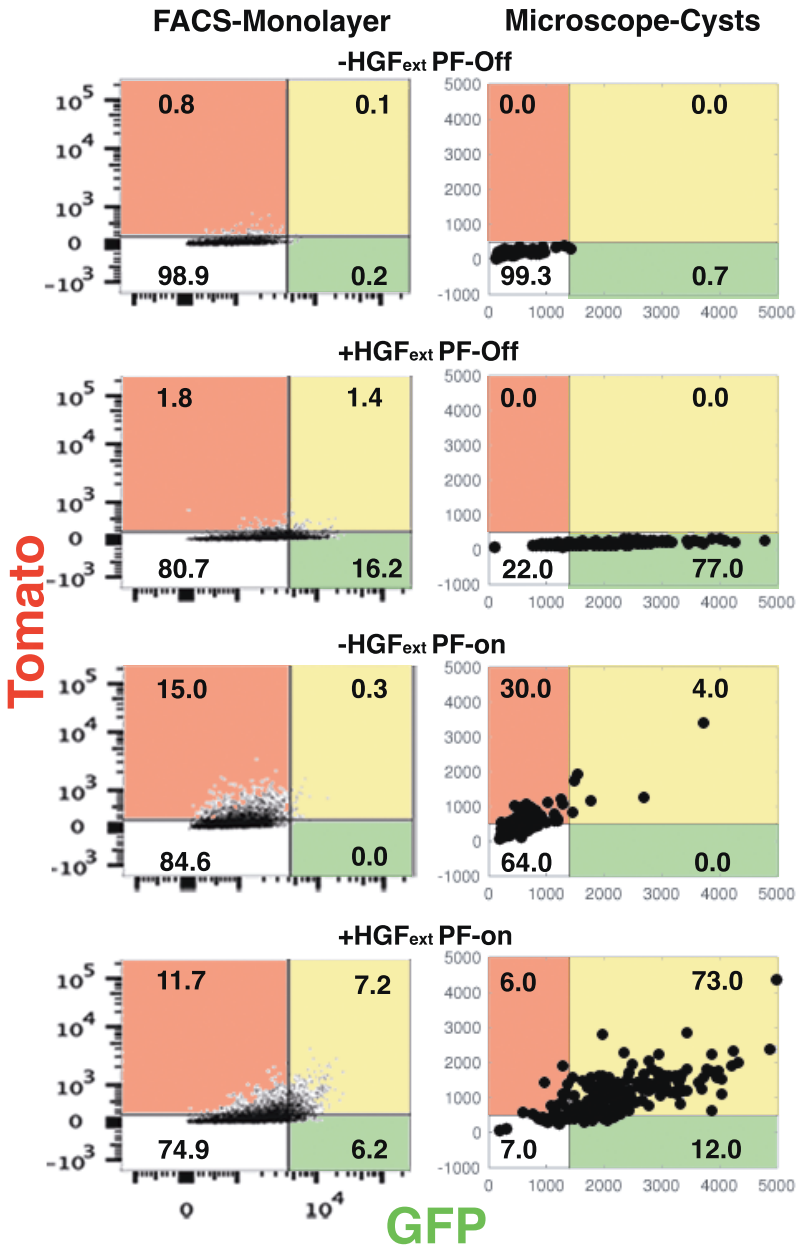


Figure 6.3 – Cysts Respond To HGF with higher fluorescence than monolayers, in a comparison of Receiver (second row) and Positive feedback (third and fourth rows) module activities. The PF-cell line was grown as monolayers or cysts and in both experiments cells were treated 24 hours prior to analysis with stimuli as in (6.2). The left row of plots shows the response in monolayers as assessed by FACS sorting (n=10,000 cells). The right row shows the response of cysts as assessed by quantitative microscopy (n=100 cysts). Numbers are the percentage of cells in either of the four gates.

To prove that HGF_{int} was being secreted into the medium, and not merely activating cells locally, I used the Quantikine Elisa kit to quantify the amount of HGF in supernatant of cells treated with tetracycline which were grown on filter inserts. MDCK cells are known to polarize, forming an apical and a basal side when grown in confluent adherent 2D cultures. Thus, this experiment would give me information as to whether HGF_{int} was being secreted and whether the secretion displayed a polarization bias. In **Figure 6.4** I show that I could detect HGF from the apical supernatant after cells were treated with tetracycline at a concentration of 3ng/ml. In contrast, I could not detect HGF from the basal side of the insert. It was not immediately clear whether this was due to the lack of permeability of the membranes to HGF, as I could not detect any significant HGF from a cell-less control were HGF was applied on the apical side, and apical supernatant was analyzed or vice-versa (data not shown).

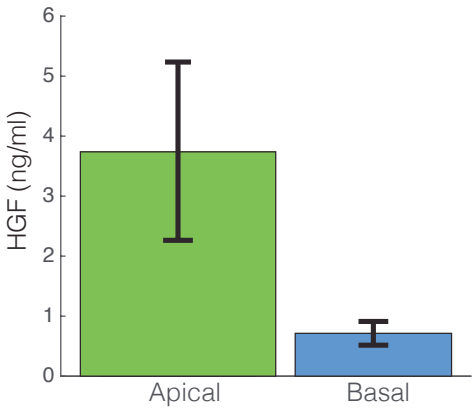


Figure 6.4 – MDCK monolayers produce HGF in the presence of tetracycline. Cells were grown on a 0.4μm filter insert until confluent and treated with 1 μg/ml tetracycline for 24 hours. 1ml supernatant from either the apical (top) or basal (bottom) side was then analyzed and with the HGF Quantikine Elisa kit. Samples were taken from three different wells (n=3). Experiments were performed by Natalie Scholes Sian. Data analysis and interpretation was performed by me and N.S.S.

Taken together, there was preliminary evidence that I had indeed generated a cell line with inducible autocrinic HGF signaling. Throughout this Chapter, the cell line will be referred to as the PF cell line. From the beginning, we anticipated that one of the biggest limitations to generating the PF cell line would be that the de-repression of the feedback module could lead to runaway feedback by means of the leaky expression observed with the pMMP1 promoter in the receiver cell. This does not seem to be the case, if anything, the feedback seems to be rather weak, as can be seen in row three of **Figure 6.3**, only a small fraction of cells express both tdTomato and GFP when the feedback is de-repressed by addition of tetracycline. In the following sections I shall describe analysis I carried out on this cell line in an attempt to quantify the strength of this feedback. I will then go on to describe attempts to generate spatiotemporal signalling experiments, to assess whether the signal could be propagated spatially and temporally between cysts.

6.2 Characterization of a positive feedback cell line

6.2.1 Feedback induced changes in Dose-Response

Receiver module (via GFP fluorescence)

I first performed a dose-response experiment using the same methodology for the analysis as in **Figure 3a and 3b in Part II**. To do this, I initially imaged cysts on two different days (n=2) and then pooled the experiments for the analysis, as the raw values did not differ significantly (See **Figure 6.5** upper left panel). In **Figure 6.5** I show the results of this experiment. The left panels show the raw data after background substraction and area correction. The right panels show the normalized data using the same normalization as in **Part II**. The dose dependent response of the PF cell line in the absence of tetracycline (**PF-off**) did indeed resemble that of the ACS1 cell line presented in **Part II**, with one notable exception: the mean maximum fold-activation was almost double at 8-fold (**Figure 6.5 upper panel)**. This higher activation could be attributed to technical measurement variability as the measurements were carried out on a different, more sensitive microscope, as outlined below.

For example, the maximum fluorescence measured in [Carvalho et al., 2014] (measurements taken at the CRG in Barcelona) hovered around 300 arbitrary fluorescence units (a.u.) with a 3 second exposure. In contrast, the measurements presented in this section (taken at Imperial College London), were carried out with a 300ms exposure time, and the measured a.u. reached 5000. This significant increase is expected to alter the dynamic range of signal detected and thus can alter the fold-activation results. Evidence pointing to the setup as a cause of significant contribution to the deviation in the observations included a different light sources, different camera module and different objectives (listed in **Chapter 8**, Materials and Methods).

I designed the PF-cell line specifically to act as its own control in the presence or absence of a positive feedback, and thus be able to draw conclusions from it in spite of technical variability. This setup helps abrogate the technical noise mentioned above. In **Figure 6.5**, I show that as expected in a system with positive feedback, the mean fold activation value increases when the feedback is on, both in the raw and the normalized data. Given that the differences observed were minimal, I resorted to statistically assess whether these differences in activation were significant. To do this, I computed the power of the data to differentiate the two conditions using Student's two sample t-test. The power test compares the distributions between **PF-ON** and **PF-OFF** at each timepoint and tests the null-hypothesis that PF-ON was higher than PF-OFF at a confidence greater than 95% given the number of observations, and the standard deviations (for details

of implementation: see *sampsizepwr* formula in Matlab). In (**Figure 6.5 lower panel**), I have plotted the power of the test in time. The experimental data was observed to prove the null-hypothesis only at the specific HGF_{ext} concentrations of 0, 1.0, and 4.2 ng/ml. Interestingly, the normalization of the data to the initial time value qualitatively increases the power of the test for all conditions.

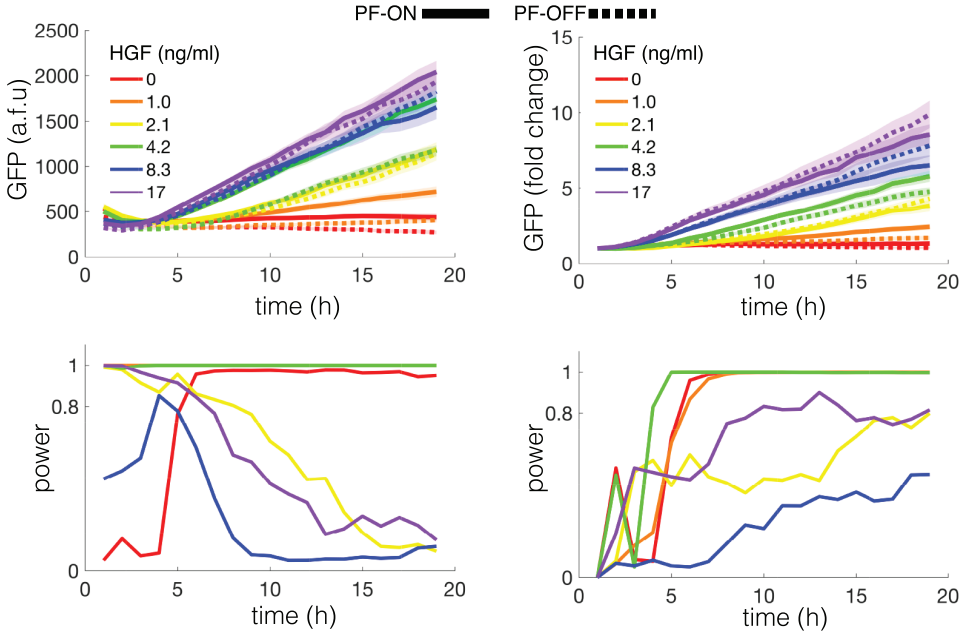


Figure 6.5 – Quantitative GFP response of the positive feedback **PF** cell line were grown in Collagen and stimulated with different concentrations of HGF_{ext} with PF_{on} (straight lines) or PF_{off} (dotted lines) of tetracycline. The upper panels are the medians over all cysts (arbitrary fluorescence units. n=100-192 depending on the condition) for either raw fluorescence data or data normalized to timepoint 1 (as in [Carvalho et al., 2014]), the shaded region represents the upper and lower bounds of the 95% confidence interval. The lower panels show the statistical power that the data for each concentration of HGF_{ext} with and without PF are from different distributions.

This initial observation seemed to indicate that the Feedback in the PF-Cell line is only acting at lower HGF_{ext} concentrations. Deviating from this trend however, the distributions for the 2.1 ng/ml HGF_{ext} could not be differentiated. The raw fluorescence values deviate substantially from cyst to cyst, and it is non-trivial to make an estimate of protein amount from this number. In **Part II**, I showed that the fold-increase normalization relative to the first timepoint can be a good way of making the different experimental conditions comparable to each

other. Indeed, as can be seen in (**Figure 6.5** right panel) the power of the data qualitatively improves in most of the conditions, however it only remains significant for the aforementioned concentrations.

Positive Feedback module (via tdTomato fluorescence)

The tdTomato response and in consequence that of HGF_{int} was designed to be dose sensitive to HGF_{ext}. The insulator module (TetR **Figure 6.1**) should ensure that in the absence of tetracycline (**PF-Off**), there will be no detectable tdTomato signal, even at higher concentrations of HGF_{ext}. This aspect of the cell line design is essential so as to minimize unwanted runaway positive feedback activation. In the left panel of **Figure 6.6**, I show that the tdTomato signal responds differentialy to the range of HGF_{ext} stimuli similar but not exactly to the response curves of GFP. There could be a number of reasons for this observation, for example, the TetO sequence could have altered the pMMP1 promoter dynamics making it more sensitive to HGF. Alternatively, the genomic integration of both the Positive Feedback module and the ACS1 receiver module are random. Thus, it is plausible that these modules reside in different locations of the MDCK genome, or even at a different copy number. These two hypothesis correspond to the activation parameters of the proteins. Unfortunately, these traits are both non-trivial to validate experimentally in mammalian systems. Attempts to validate the integration site by reverse PCR or the integration copy number by southern blot, were not successful.

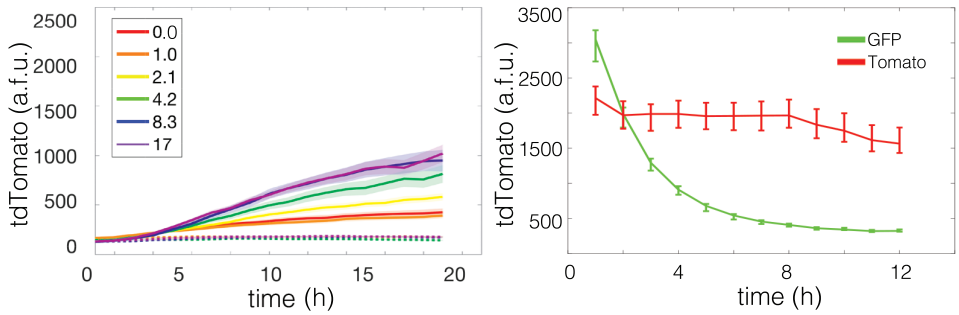


Figure 6.6 – Fluorescence activation of tdTomato in response to different doses of HGF_{ext} in the presence and abscence of tetracycline (solid vs dotted lines respectively)(left panel). Cysts (n=121) were also treated with cycloheximide after 24 hours of 17ng/ml HGF_{ext} stimulus and imaged for 12 hours (right panel). a.f.u = arbitrary fluorescence units

A third hypothesis to account for the different dynamics of tdTomato activation could be that the rate of degradation of the tdTomato fluorophore was altered in relation to that of GFP. If this were the case, the accumulated tdTomato would saturate the signal at much lower concentrations of stimulus. In (**Figure 6.6** right panel) I show results of cysts imaged after a dose of 1μg/ml cycloheximide to inhibit protein biosynthesis 24h after HGF_{ext} and tetracycline treatment. Evidently, there is a clear difference in the degradation rates of the two fluorophores. The tdTomato signal does not degrade at all whereas the GFP signal is completely gone after 12 hours. This discrepancy alone could account for the apparent difference in response dynamics between the PF and Receiver modules.

I did not attempt to characterize this difference in degradation. There are two plausible explanations on which I will briefly elaborate. On the one hand, the self-cleaving P2A peptide sequence between the HGF and the tdTomato cDNA is known to leave peptide residues both in the c terminal and n-terminal proteins after being cleaved ([Rothwell et al., 2010]). Although not reported, it is possible that the carboxy terminal scar might interfere with the PEST domains' targeting to the proteosome. A second possibility could be the conformation of the homodimeric tdTomato protein might obstruct the PEST domain, as I have found no reference in the literature of a d2tdTomato protein. Further experiments to resolve this issue will be needed; to date, attempts to generate a degradable red fluorescent protein which would display no fluorescence spectral overlap with our microscope configurations were unsuccessful. Thus for most quantitative analysis, I have used only the GFP channel where the tdTomato channel served as a qualitative guide on whether HGF was being expressed in cells.

Correlation between HGF_{int} production and GFP

The PF cell line was designed to monitor the amount of HGF_{int} that is dynamically being produced at any given moment by means of the intensity of tdTomato fluorescence. The rationale behind this was to provide a tool to directly measure the contribution of the positive feedback to the system, and establish how this contribution behaves in time. As shown in **Figure 6.6**, and discussed above, the difference in degradation rates between GFP and tdTomato abrogates this functional design of the PF cell line. Notwithstanding, some measures could be extracted from the available data. For example, I found particularly intriguing to explore how the activation of receiver and feedback modules correlate within each cyst. In **Figure 6.7** the estimated Pearson correlations of GFP and Tomato for individual cysts from the experiment in **Figure 6.5** are shown. As expected, in the absence of feedback (no tetracycline) the correlation was negligible. Interestingly, at higher HGF_{ext} doses a sub-population of cysts displayed correlation which could hint to leakiness in the repression by TetR of the positive feedback module.

In the presence of feedback (with tetracycline), there is a small increase in the correlation in the absence of HGF_{ext} which confirms that there is communication between the two modules, at least in a subset of cysts. Interestingly under these conditions, there is also a small percentage of cysts (10%) which are apparently anticorrelated. This could be an indication but does not prove promoter competition within these cysts. Above a concentration of 4.2ng/ml of HGF_{ext}, the correlations in cysts is maximized (more than 90% of cysts show a correlation above 0.9). This dose dependence of the correlations, was further proof that both promoters are being dynamically regulated by different amounts of HGF_{ext}.

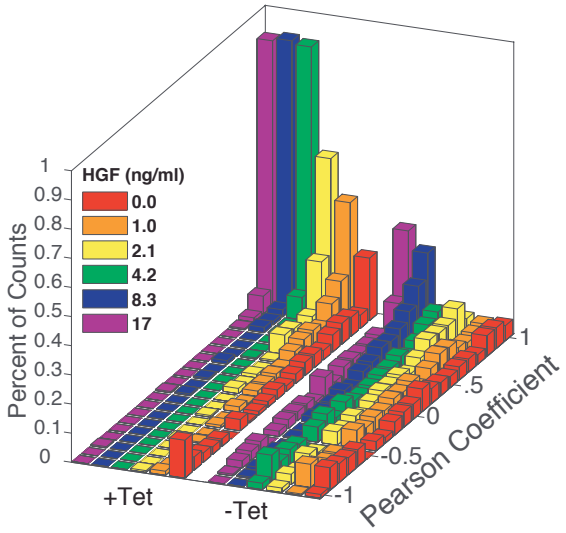


Figure 6.7 – Pearson correlation of GFP and td-Tomato trajectories in time at increasing concentrations of HGF (colors), with and without positive feedback.

Analysis of the Steady State

The time window of observation of the experiment presented and discussed in the previous section (**Figure 6.5**) was not sufficient to observe the saturation of the response signal. The differences in mean fold activation between the two states - **PF-on** and **PF-off** - were barely significant. I hypothesized that perhaps some effects of the positive feedback could manifest while at steady state. In **Figure 6.8 upper left panel** I show the behaviour of PF cells under similar conditions as in the previous experiment, however these cysts were imaged starting 24 hours post stimulus. In this case, the fold changes are not in regard to the first timepoint, but in regard to the mean of the **PF-off** and $0ng/mlHGF_{ext}$ condition (red dotted line). Under these conditions, the positive feedback causes a shift upwards in the steady state levels in all conditions except for the highest (17ng/ml purple lines) HGF_{ext}. This observation can be seen as further evidence of a functioning autocrinic signalling.

In agreement with a the observations made in the previous section (**Figure 6.6**) of deficient tdTomato degradation, the mean fluorescence for the cysts continues to increase and seems to converge to a single point (**Figure 6.8** upper right panel). This is shown in clearer format in **Figure 6.8** lower left panel where I am showing

the average velocities over the time-series for every cyst. The GFP fluorescence displays a net decrease over the time periods observed except for the higher concentration which reaches the maximum activation at roughly 24 hours. In contrast the net difference in time of the tdTomato fluorescence is mostly positive except at the higher concentrations where the saturation level of roughly 6-fold activation have been reached earlier.

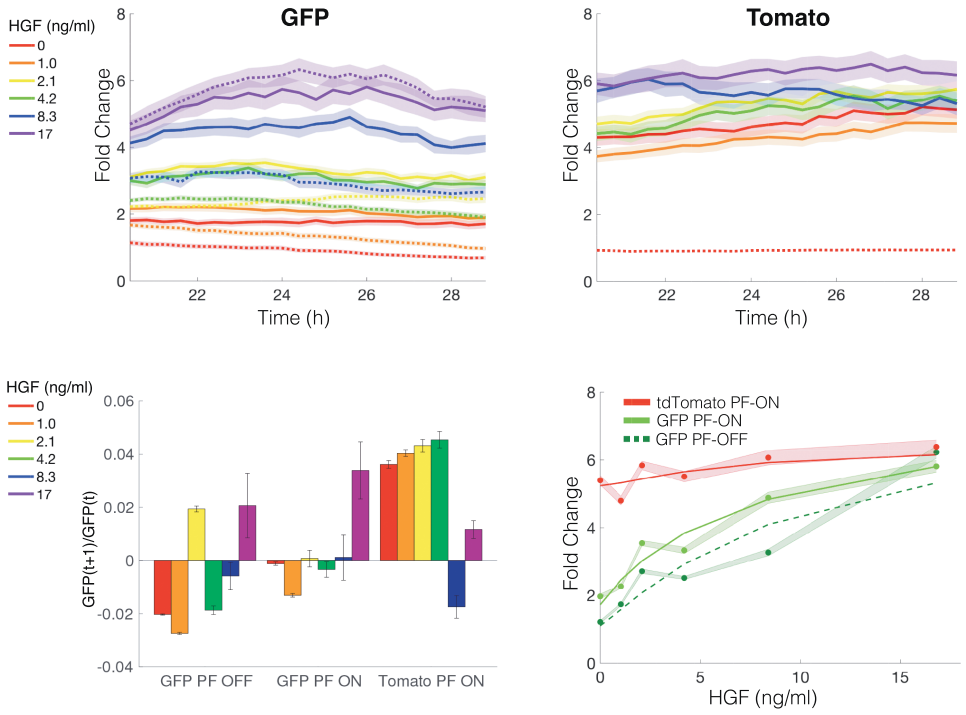


Figure 6.8 – Quantitative GFP response of the PF cell line were grown in Collagen and stimulated with different concentrations of HGF_{ext} with PF_{on} (straight lines) or PF_{off} (dotted lines) of tetracycline. The upper panels are the medians over all cysts (n=100-192 depending on the condition) for either raw fluorescence data or data normalized to timepoint 1 (as in [Carvalho et al., 2014]), the shaded region represents the upper and lower bounds of the 95% confidence interval. The lower panels show the statistical power that the data for each concentration of HGF_{ext} with and without PF are from different distributions.

To quantify HGF_{ext} induction, and in particular the contribution of HGF_{int} to this quantitative measure, I analyzed the response dynamics at the steady states.

Table 6.1 – Best fit parameters using Markov Chain Monte Carlo exploration.

Parameter	GFP ACS	GFP Thesis	GFP PF Thesis	TOM Thesis
b	0.9	1.1	1.7	5.2
V	3.65	6.3	5.9	1.14
K_d (ng/ml)	12.6	9.1	7.6	6.32
n	1.19	1.16	0.99	1.4
R_{max}	4.55	7.41	7.6	6.2

I use the same approach as in **Part II** section 8.1, by fitting the data points to the sigmoidal response given by:

$$b + \frac{V * a^n}{K_d^n + a^n} \tag{6.1}$$

The correct parameters were searched by fitting the data shown in (**Figure 6.8** lower right panel) using the Markov Chain Monte Carlo (MCMC) algorithm, varying the initial baseline response b, the half activation V, the dissociation constant k_d and the hill coefficient n. Given the microscopy technical measurement difference discussed above, the parameter values were not expected to be those presented in **Part II** and are listed in **Table 6.1**.

A clear conclusion from these initial experiments is that the "strength" of the positive feedback, did not seem to be enough to elicit full activation of the MMP1 promoter in the time-frame of the observations, causing runaway "feedback". This could hint to the fact that the feedback is either globally thresholded across cysts, or that the HGF_{int} signal is not reaching cells unimpeded. The observation that only a subset of cysts respond strongly to the de-repression of the HGF module **Figure 6.3**, lends support to the hypothesis that there could also be a stochastic element whereby, a sub-population of cysts is hypersensitive to HGF. If this is the case, at higher doses of HGF_{ext} this population would be masked by the response of a high percentage of cysts, rendering the hyperactive outliers of lower concentrations practically invisible.

Data from **Figure 6.3** seems to lend credence to the outlier hypothesis. It is unclear what the underlying cause for this observation could be. I can at this point only speculate about the reasons for this observation. For example, the lack of signal propagation in a cyst which produces HGF_{int} could indicate that the receptor is not perceiving this signal. One tantalizing explanation for this is that HGF_{int} secretion occurs towards the opposite direction of where the HGF receptor *c-Met* is localized. Given the heterogeneous composition of the collagen I fibrils upon which the cells grow, and the sensitivity of the cells to their substrate (see further down in section 6.3 for evidence supporting this hypothesis) it is fathomable that

a subset of cysts mis-localize the receptor or the morphogen or both, and are thus more sensitive.

The above conclusions made it particularly interesting to further analyze the effects of the positive feedback at lower HGF concentrations, to discern what the contribution if any it makes to the response dynamics to HGF. Furthermore, it would be interesting to explore other measures beyond the mean fold activation differences. In the next sections I will elaborate on potential methods to discern and quantify the feedback in the PF-cell line.

6.2.2 Estimating the feedback gain and factor

As discussed in the previous section, at certain HGF_{ext} concentrations, I observed a positive feedback effect on the mean of a population of cysts. The feedback seemed to be rather "weak", in this section I will demonstrate some approaches to estimate what, if any measures can be attributed to the feedback contribution to the PF-cell line. In the first part of this section I repeated the dose-activation for three independent HGF_{ext} concentrations. Firstly, 0ng/ml HGF_{ext} which will serve as a control on how the positive feedback influences the system without external inputs. Secondly 4.2ng/ml of HGF_{ext} as this was the concentration at which the strongest difference between PF and non PF was observed. And lastly I also chose the highest concentration of HGF 17.0ng/ml, as no PF was observed and this can serve as a negative control for feedback.

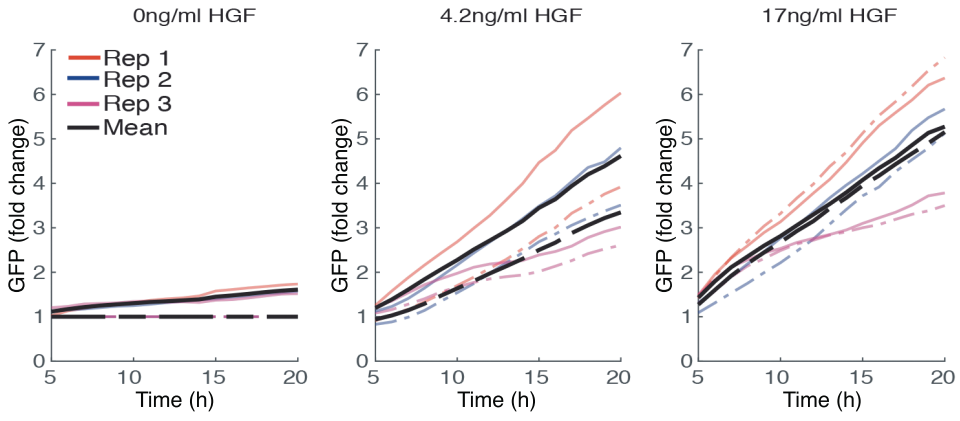


Figure 6.9 – Normalized means of three different experiments at three different concentrations of HGF_{ext} in the presence (straight line) or absence (dotted lines) of tetracycline. The mean of the means are shown as black lines.For each individual Replicate (Rep 1-3) n>100 cysts.

In **Figure 6.9**, the colored lines labeled as Rep 1-3 I represent each the mean of individual biological replicates. This representation is important to illustrate the high variability in these experiments comes both from intrinsic (each cysts variability within a given day) and extrinsic (batch effects such as temperature, basal membrane composition, etc..) sources. Furthermore, it shows that both variabilities increase at higher concentrations of HGF_{ext}. Again, cysts with the de-repressed positive feedback were distinguishable only at lower concentrations of HGF_{ext}, confirming the previous observations.

Positive feedback strength

I first explored the hypothesis that HGF_{int} production was acting as a dynamical non-linear amplifier of the HGF_{ext} signal. I applied a methodology originally used by climatologists (who adapted it from electronics, see introduction section: 3.2) to estimate the feedback contribution to the system. In this case I assume that a concentration of HGF forces the system to express a response GFP. The total amount of HGF in the system for any given time-series of observation can be thus be described as:

$$HGF_{total} = HGF_{int} + HGF_{ext} \tag{6.2}$$

The presented data demonstrate that I can measure the fold-output response of the system via quantitative microscopy analysis of GFP fluorescence. The effect of the feedback in the system, can be discerned by comparing the response of the PF-cell line in time with and without a positive feedback. Therefore, the total amount of GFP can be decomposed in the parts coming from HGF_{ext} and a part coming from HGF_{int} so:

$$GFP_{total} = GFP_0 + GFP_{PF} \tag{6.3}$$

Where GFP_0 is the measured response of the cysts to HGF_{ext} and GFP_{pf} is the response of the cysts in the presence of the positive feedback. To estimate the dynamic contribution of the feedback to the system, the rate of activation per unit time can be analyzed, which allows me to calculate the feedback gain g as described in **section 3.2**:

$$g = \frac{\Delta GFP_{feedback}}{\Delta GFP_{feedbackOFF}} \tag{6.4}$$

For the analysis, I have assumed ΔGFP to be the difference in normalized fold GFP values for each time-point by the time preceding it. In order to estimate

ΔGFP for each condition, I calculated the mean of the differentials of GFP expression between two contiguous time-points for each individual cyst. These values remained relatively constant in the time frame of observations, thus I deemed it possible to compare between conditions.

By this measure, the rate of GFP production increases with higher HGF_{ext} concentrations (**Figure 6.10**). At lower HGF_{ext} concentrations, there was also a statistically significant increase in GFP rate in cells with the positive feedback. This difference can be solely attributed to the effect of HGF_{int} on the cysts. The net feedback contribution to the system can be calculated by $\Delta GFP_{feedback} - \Delta GFP_{feedbackOFF}$, and the gain of the feedback could be estimated by dividing this value by the equilibrium change of GFP.

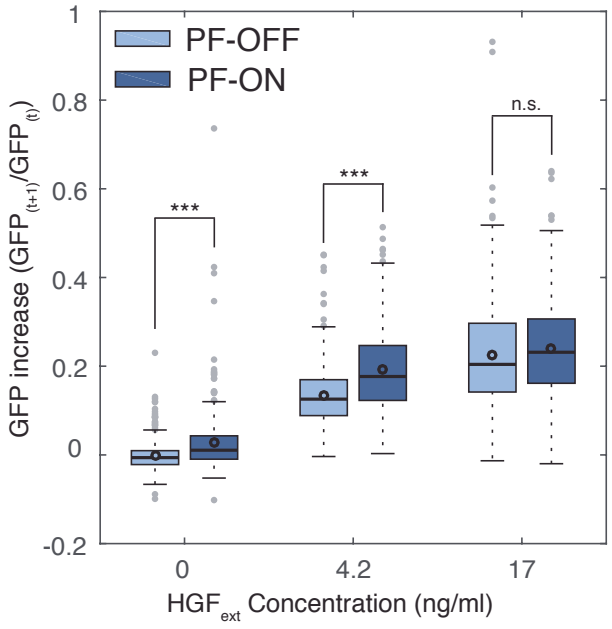


Figure 6.10 – The rate of GFP production increases with HGF concentration and with the positive feedback. Boxplots of the average change in GFP per unit time for each individual cysts. Boxes represent the 95 percentile of cysts, and contain lines for the median and a circle for the mean. Error bars represent 1 standard deviation from the mean, and single dots are outliers. The significance between positive feedback (dark blue) and no positive feedback cysts was calculated using the Kruskal Wallis test method and *** p < .001, n.s. not significant p > .05.

For the condition with 0ng/ml HGF_{ext} the ΔGFP is close to zero, thus the feedback gain tends towards infinity. For the higher concentrations however it was possible to estimate a feedback gain contribution to the system of 0.4679 for 4.2ng/ml HGF_{ext}. An almost undetectable gain of 0.0423 for the condition with 17ng/ml, was deemed non-significant.

The feedback factor f as per [Hansen et al., 1984] is related to the gain g by:

$$f = \frac{1}{1-g} \qquad (6.5)$$

and is an "intuitive quantification of the strength of feedbacks". For the above mentioned conditions, I found a feedback factor of 1.8793 for the 4.2ng/ml concentration and 1.0442 for the higher condition 17ng/ml. To my knowledge there have been no previous efforts in the experimental biology literature to quantify the strength of a feedback using the

above mentioned methodology, thus it is not possible to compare the feedback gain of my system to that of other biological systems, synthetic or natural.

Where possible, it would be interesting to compare other purported feedback systems in biology using this measure, as it is a dimensionless attribute. One advantage of the synthetically designed system presented here, is that the effect of the feedback can be turned on and off at will, thus making the detection of said feedback more straightforward.

6.2.3 Computational Feedback Model

Given the lack of standardized quantification methods for strengths of biological feedbacks in the literature, with the help of Marc Sturock I devised a simple kinetic model for a linear autocatalytic reaction. I hypothesized that such a model could serve as a guide to find relative quantities of the PF system. This model was not devised to fit or model the observations in the data, but rather to theoretically guide me to understand feedbacks, and help me devise a method to ascertain whether the PF cell line was indeed a positive feedback. The model consists of three simple reactions with respective rate constants:

$$\emptyset \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A \xrightarrow[k_3]{} A + A \tag{6.6}$$

I reasoned that with this simple model I could explore parameters for different birthrates k_1 and positive feedback strengths k_3 of an activator A. The parameter rate k_2 is akin to a degradation rate, and in this form is a negative feedback which serves to set a limit to growth. I have introduced this parameter in order for the concentrations of A not to tend towards infinity, and have not varied it throughout the exploration.

The rates and concentrations are given in arbitrary and dimensionless units, and the changes in concentration of A are calculated using Gillespie's method whereby each time-step of the simulation was akin to 1 simulated minute of the reaction ([Gillespie, 1977]). This model will serve as a guide, to test other measures for positive feedbacks. I have manually chosen the parameters so that levels similar to the absolute raw values of GFP in the data are reached, however the results observed here apply generally. The steady state concentration of A will be reached at: $\frac{k_1}{k_2 - k_3}$. A copy of the code for the model is found in the Materials and Methods section 8.3.1.

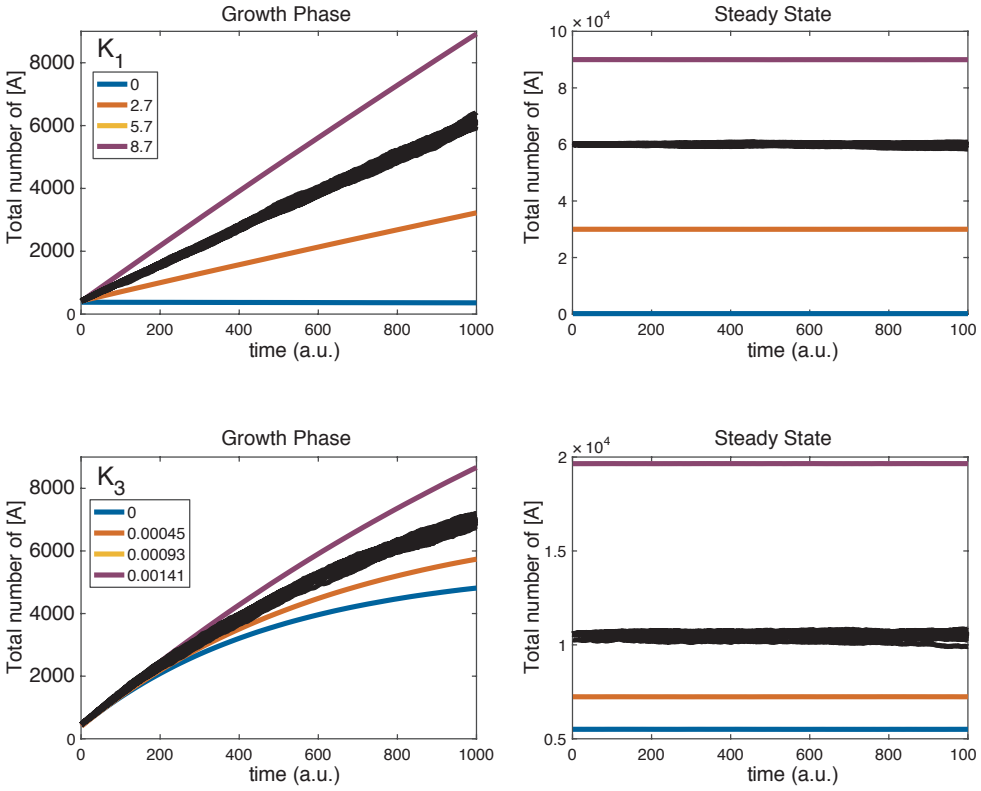


Figure 6.11 – Results from the stochastic model in equation 6.6, with variation of production rate k_1 (upper panels $k_1 = 0, 2.7, 5.7, 8.7$) and a fixed k_3 of 0.0019. I also varied the positive feedback rate k_3 (lower panels $k_3 = 0, 0.00045, 0.00093, 0.00141$) and fixed k_1 to 11. 1000 timesteps are plotted for either the growth phase (left column) or the steady state phase (right column). The color lines are the averages over 1000 runs of the stochastic model. Black lines show an example of 100 model runs.

In **Figure 6.11** I show example traces for different values of K_3 and K_1 for and average of 100 simulations for each condition. In this case, I chose parameters so that they match the raw values observed in the experimental data. I have displayed the concentration of A for both the growth phase and the steady phase for different parameters of K_1 and K_3.

Running the stochastic model at such high copy number of molecules is computationally expensive, thus for the following analysis, I used different parameters for K_1 and K_2, which resulted in a equal range of A steady state means between 0 and 200. The result of runs of 10,000 steps are represented as a heatmap in **Figure 6.12**, and the mean of each trace is shown in **Figure 6.13**. Using these parameters, there are no obvious differences in the behaviors of the means between the model

runs. This was necessary, as the objective of this analysis was to differentiate between a positive feedback and no positive feedback. In the following sections I will elaborate on the measures we explored to understand the effects of positive feedback.

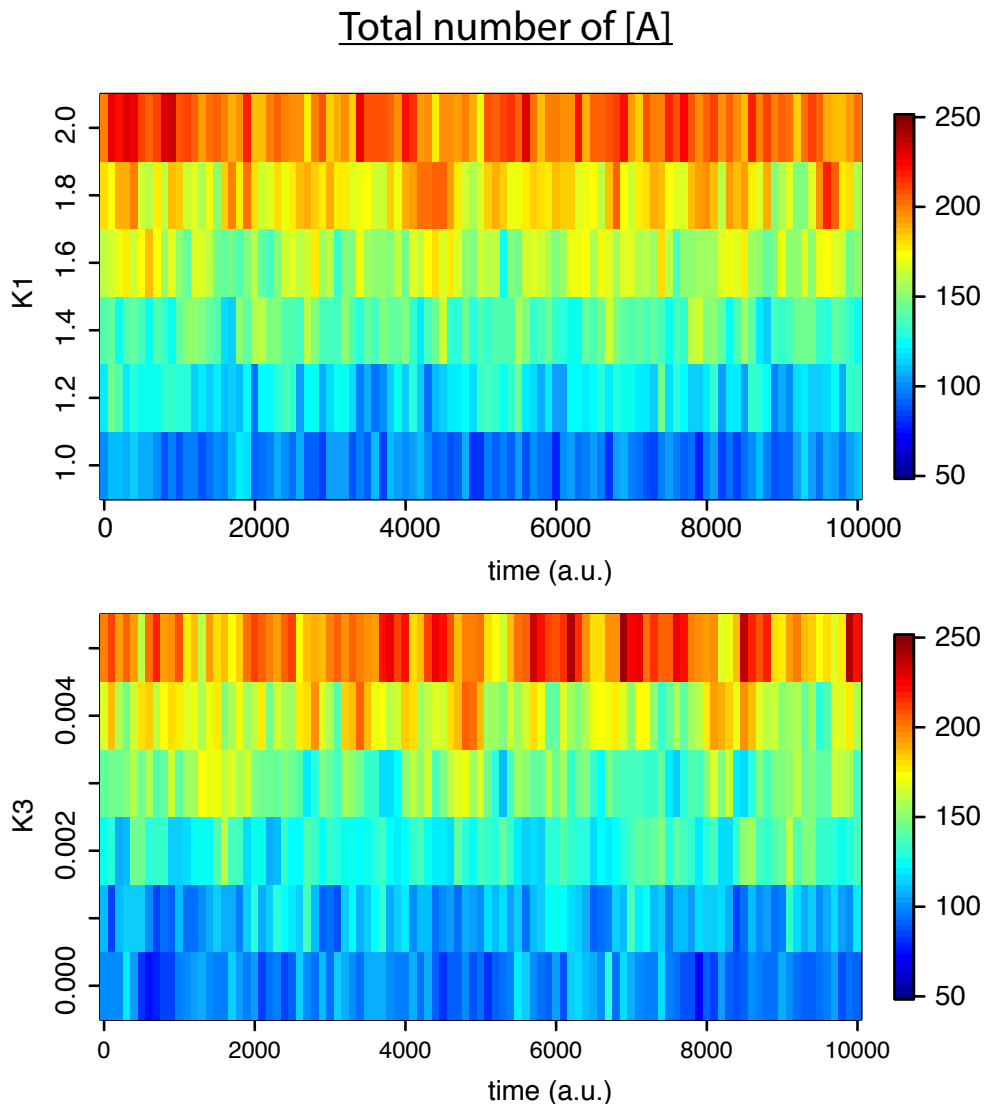


Figure 6.12 – Stochastic model shows similar mean activities at the steady state levels with and without positive feedback. Results are shown from the stochastic model in equation 6.6, with fixed $K_2 = 0.01$ and variation of production rate $k_1 = 1.0, 1.2, 1.4, 1.6, 1.8, 2.0$ (upper panel) with $k_3 = 0$, and variation of $k_3 = 0, 0.0016, 0.0028, 0.0037, 0.0045, 0.005$ with $K_1 = 1.0$ (lower panel). 1e5 timesteps are displayed.

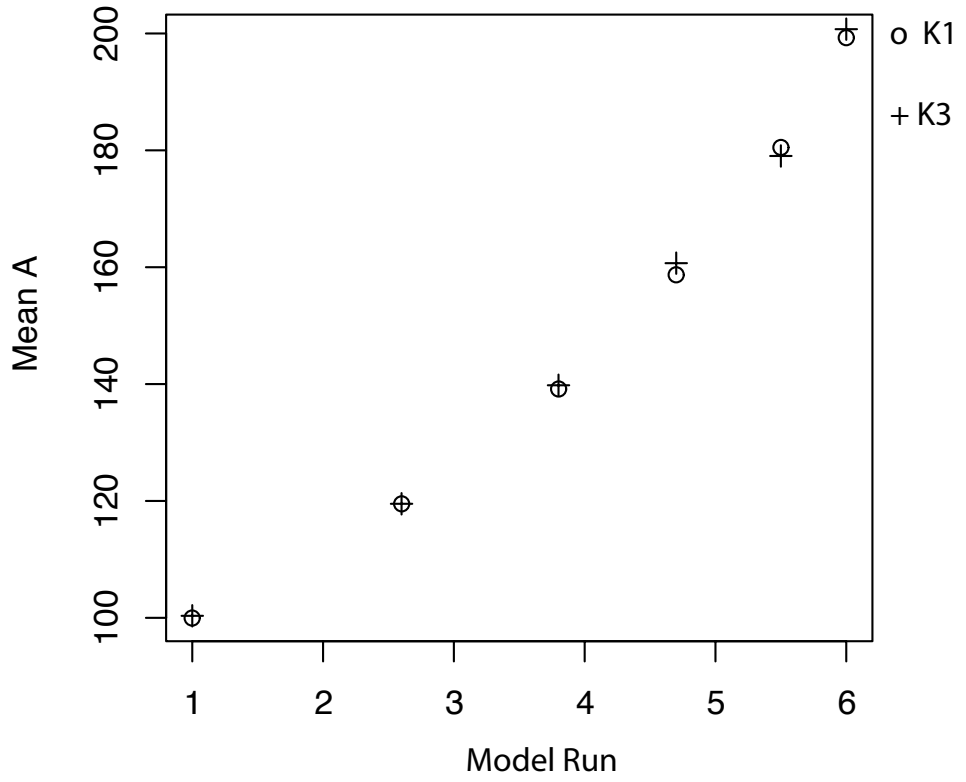


Figure 6.13 – Stochastic model shows similar mean activities at the steady state levels with and without positive feedback. Results are shown of the mean of the stochastic model in equation 6.6, with fixed $K_2 = 0.01$ and variation of production rate $k_1 = 1.0, 1.2, 1.4, 1.6, 1.8, 2.0$ and $k_3 = 0$ (circles). The mean of the runs with $k_3 = 0, 0.0016, 0.0028, 0.0037, 0.0045, 0.005$ and $K_1 = 1.0$ are displayed as crosses.

Autocorrelation

An autocatalytic process can induce a memory effect by means of its concentration. Intuitively, this explains the increase in "GFP velocity" observed above in the data in the exponential phase (**Figure 6.10**), as more HGF will lead to more HGF, etc... I asked myself, whether it was possible to observe this self-similarity in the steady state too. An unbiased indicator of this effect used widely in economics is the autocorrelation function (see equation 3.7 in the Introduction). This function serves to measure how a series of data correlates to itself at different intervals (τ). A long-term memory process will show an increase in autocorrelation if feedback is present, i.e. the parameter K_3 is higher.

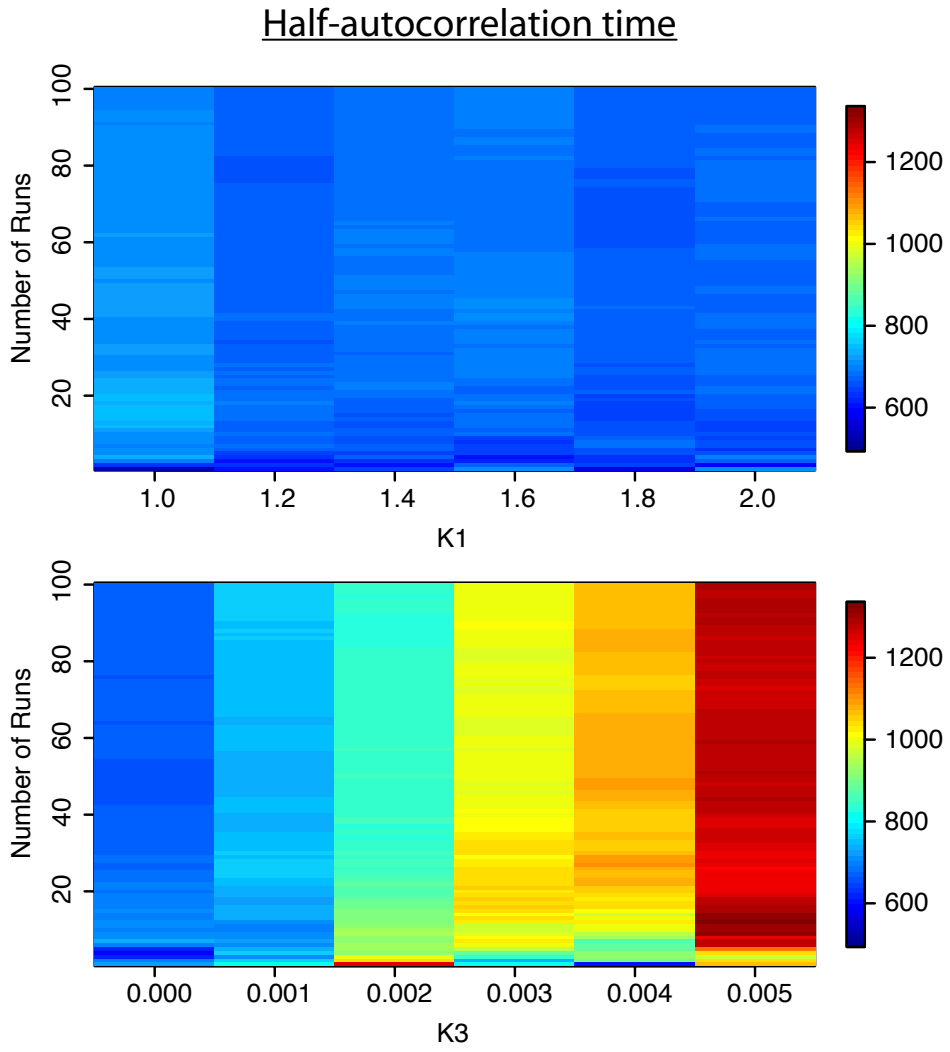


Figure 6.14 – Effects on the autocorrelation time for the stochastic positive feedback model. The model was run for 100,000 time steps for each value of either K_1 or K_3 (x-axis). The auto-correlation of each run was calculated for a maximum lag of 1000 using the acf function in R. The average autocorrelations are plotted for increasing number of runs (y-axis).

I looked at how the autocorrelation time for the concentration of A behaves in the model presented above. When I initially ran the model for short periods of time (1000 minutes), or few stochastic repetitions (1000 runs) I could observe no apparent difference in autocorrelation times when increasing either the K_3 or the K_1 parameters. As I increased the number of runs, and averaged the autocor-

relation times, indeed a trend emerged of an increase in autocorrelation time for higher K_3 parameter values. In contrast, variation of the K_1 parameter had no effect on this parameter under the same conditions (**Figure 6.14**).

This observation indicates that positive feedback can in fact induce long-term memory in a system, albeit mathematically, this can only be observed closer to the deterministic limits of the system, or when making high numbers of observations. For example, the model uses timesteps of 1 minute, thus the total time of "observation" in the above run ($1e5$ timesteps) was of 69 days, this is an unreasonably long time in biological setups. Furthermore, as is evident in **Figure 6.14** even at these high number of observations, the autocorrelation only increases significantly above 10 observations.

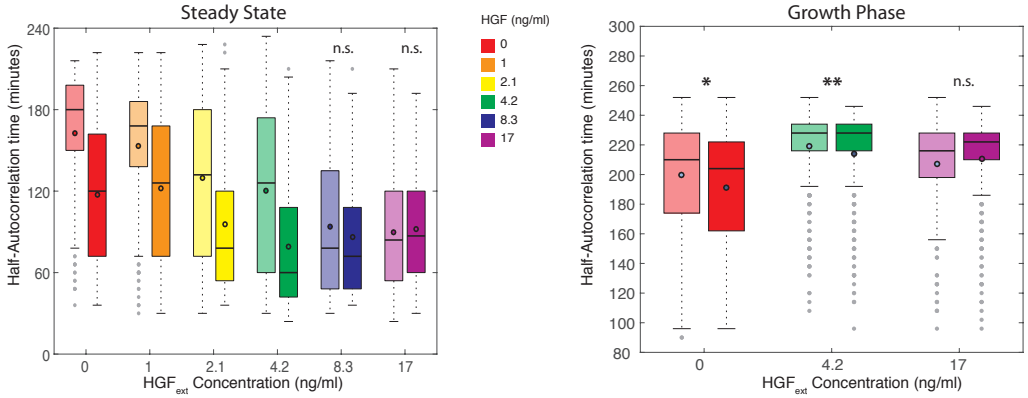


Figure 6.15 – Positive feedback decreases half auto-correlation time in cysts in steady state. Boxplots of half-autocorrelation times (lag=5) of individual cysts for the different HGF concentrations. Results are of observations at steady state (start at 24h) and growth phase (start at 2h). * $p < 0.005$, ** $p < 0.001$, n.s. not significant $p > 0.05$. HGF amounts are shown in colors, whereby the shaded color always indicates the condition without positive feedback.

I next proceeded to measure the auto-correlation times of the individual cysts in terms of GFP. To do this, I used the autocorr function in the Matlab statistics toolbox to analyze the time series of cysts with and without positive feedback for individual cysts using the raw GFP values, and then averaged these over time for the different HGF_{ext} concentrations at steady state from **Figure 6.8**.

Interestingly, as I show in **Figure 6.15**, the half-autocorrelation time behaved opposite of what the model predicted. Under all HGF_{ext} concentrations, the autocorrelation time decreased when the positive feedback was present. There was no immediate explanation for this observation, however it is possible that the extrinsic noise inherent to the system could influence this measure. It could be interesting for future explorations to include extrinsic noise in the model.

Thus, results from the model indicates that there is indeed a measurable difference in the auto-correlation time depending on the strength of a positive feedback. This effect however, is only visible when comparing different strengths of feedbacks, which might be difficult to observe in nature. On the other hand, even if a feedback process could be differentiated by the autocorrelation time, I have shown that the observation time or data required to observe any effect is high and not easily attainable in biological experiments. The reasons for this are mainly technical: keeping cells alive and under observation for such a long period is not trivial. Furthermore, the vast amount of imaging data such a set of experiments could generate, would be overwhelming.

Hurst Exponent

As discussed in the introduction the Hurst exponent is another measure of long memory effects in systems related to the autocorrelation function of a time series. The exponent is a measure from 0 to 1 whereby 0.5 is a brownian random process. As with the half autocorrelation time above, the Hurst exponent is expected to increase with the positive feedback ([Lillo and Farmer, 2004]).To calculate the Hurst exponent of A in the model, I used the corrected Hurst function of the pracma package in R ([Weron, 2002]). The results of this analysis are shown in **Figure 6.16**. The Hurst exponent indeed increases in the longer runs of the model simulations with higher K_3 values, albeit this trend is only visible again with increasing number of repeats. Interestingly, the Hurst exponent was above 0.8 regardless of the strength of the K_1 parameter. This fact seems to indicate that the process analyzed was not completely random.

An attempt to estimate the Hurst exponent from the steady state experimental data, was not technically possible. The time measurements were not sufficient to correctly fit the rescaled range time series. Thus, as with the autocorrelation, it would seem that with the Hurst exponent I was again met with the paradigm of having a predictor of feedback but too low number of experimental observations to be able to use this measure with precision.

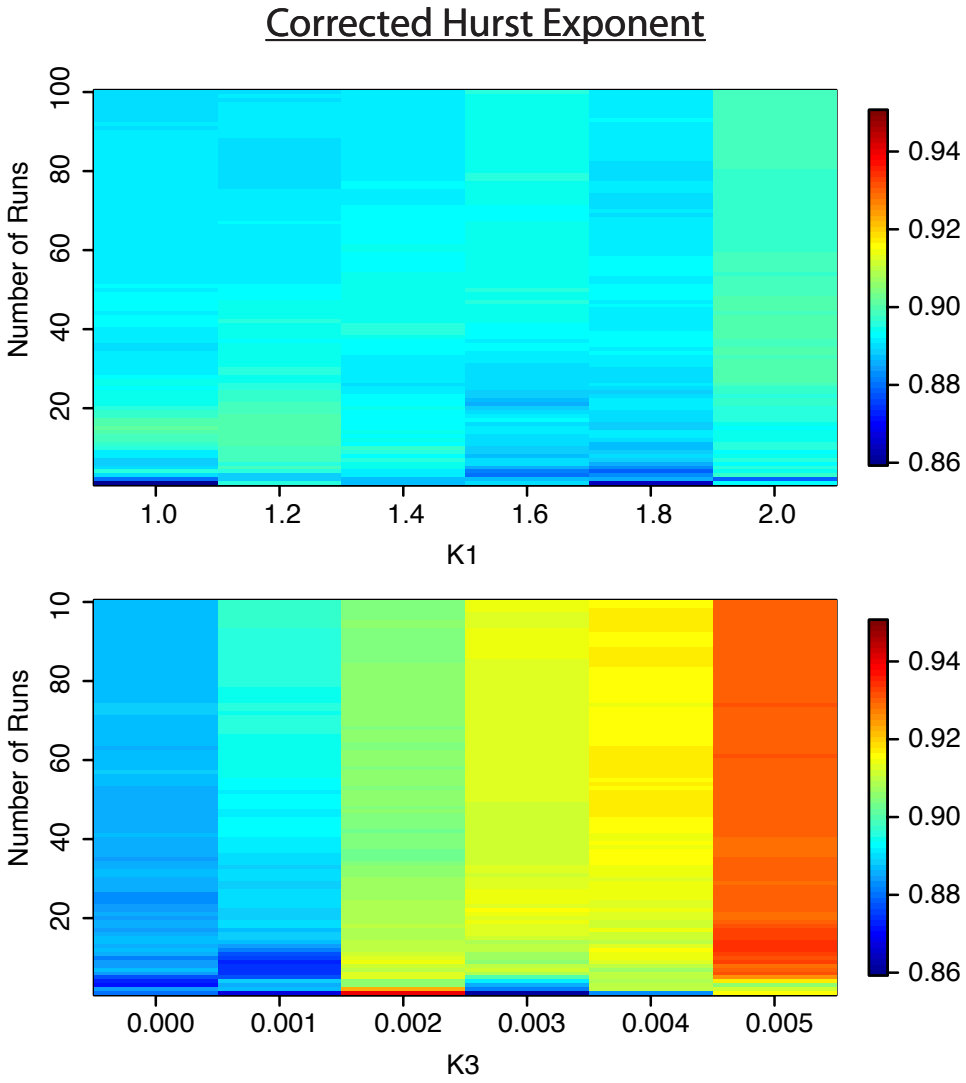


Figure 6.16 – Effects on the corrected Hurst exponent ([Tom and Andrew, 1991] and Physica A 312, 285-299) of variation of parameters in the stochastic positive feedback model. The Hurst exponent was calculated using the hurst function in the pracma R package using the default boxsize of 50. The model was run for 100,000 time steps for each value of either K_1 or K_3 (x-axis). The Hurst exponent was calculated for different time lengths (y-axis).

Measures of noise

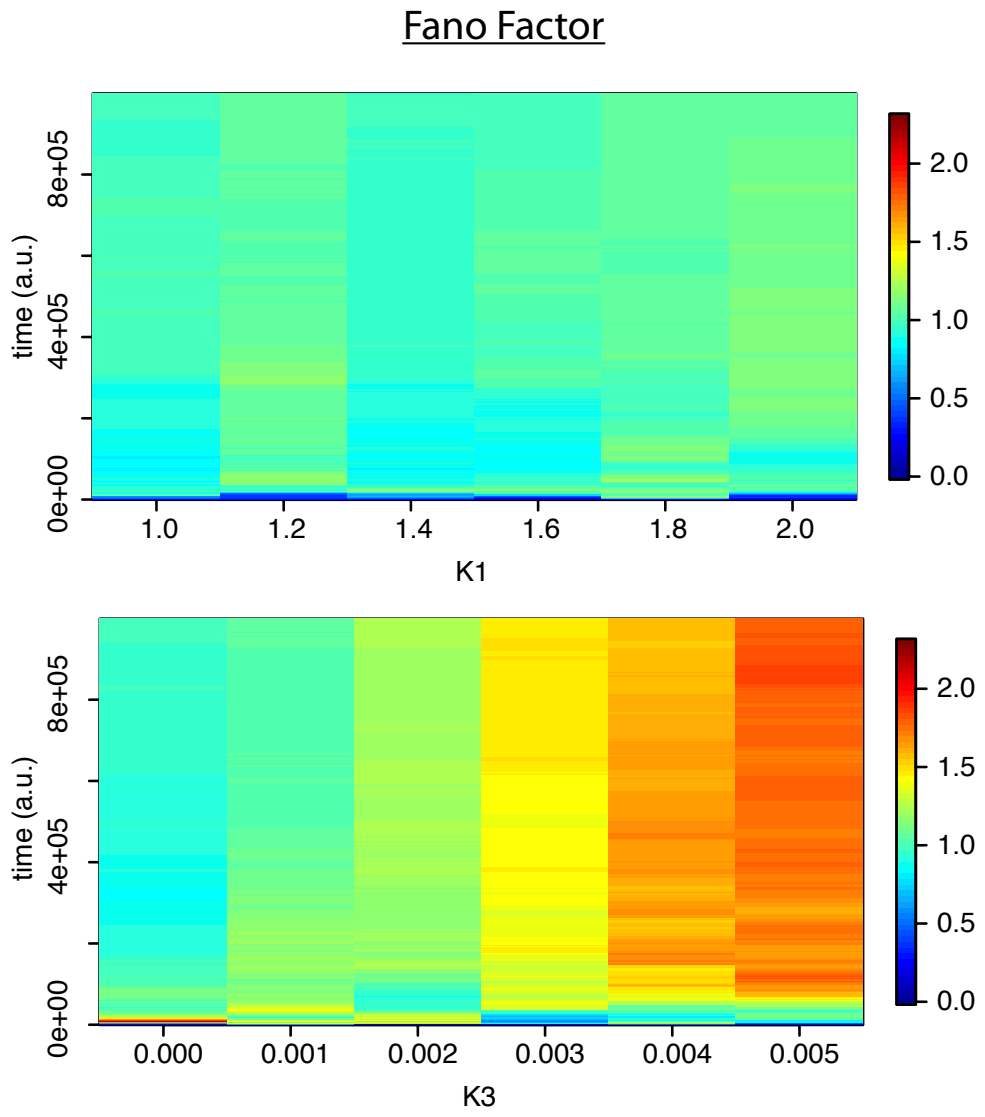


Figure 6.17 – Effects on the noise of the stochastic positive feedback traces. The model was run for 100,000 time steps for each value of either K_1 or K_3 (x-axis). The Fano factor of each run was calculated by dividing the variance by the mean for a maximum time length of 100,000 at increasing time windows (y-axis).

Another measure that is expected to increase in the presence of feedback, is the variance in activity. Whereas negative feedbacks are usually applied to modulate

noise in systems, positive feedback is projected to increase noise. To measure the noise in the model I used the measure for noise known as the Fano factor: $\frac{\sigma^2}{\mu}$ for different time windows in a simulation run for $1e6$ time steps. In contrast to the coefficient of variation, which is a ratio between noise and signal, the Fano factor is a measure of noise to signal. As with the previous measures, the Fano factor increased with higher positive feedback parameters (**Figure 6.17**). Again, this observation was only significant at higher number of observations. In contrast to the increase in autocorrelation observed previously, which became evident as more samples were observed, to calculate the fano factor I looked at different time window lengths within a longer run of the model data.

I next estimated the Fano factor of the data presented in **Figure 6.8**. Assuming that the cyst expression of GFP is truly in steady state, I concatenated the individual cyst traces of 22 hours to construct a long time series. I then calculated the fano factor for each of the long series of HGF_{ext} conditions (**Figure 6.18**).

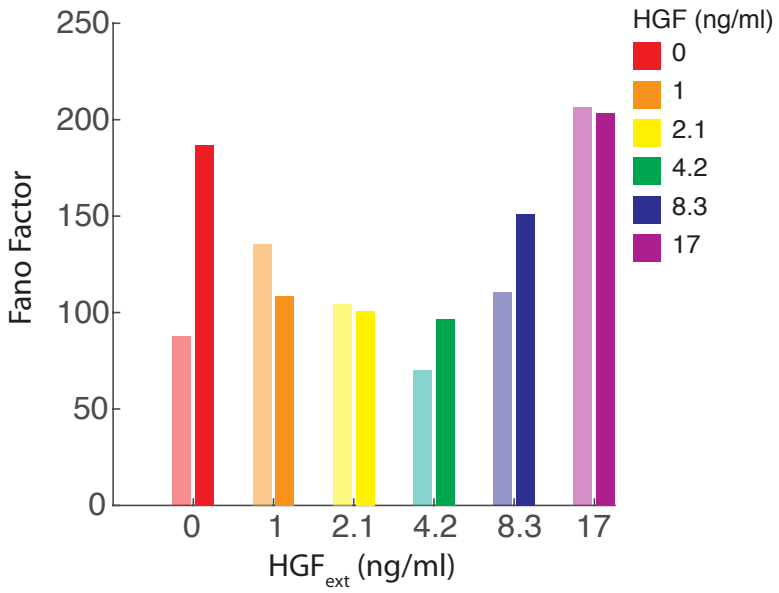


Figure 6.18 – Fano factor changes on steady state cysts in the presence or absence of feedback. The timeseries of the individual cysts for each condition was concatenated to create a long trace. The fano factor was calculated as the variance over the mean of the concatenated trace. HGF amounts are shown in colors, whereby the shaded color always indicates the condition without positive feedback.

Using the Fano factor obtained from the different HGF_{ext} conditions, it becomes apparent that the noise in the PF cell line only increases when no external

HGF is applied. This is an indication of the hypersensitive outliers mentioned above and seen clearly in **Figure 6.3**. The noisiest state of the system is seen at the maximum of HGF_{ext}.

Most effects I looked at in the model, became apparent only after looking at a high number of repetitions. Alternatively, I could also detect differences when I observed longer time-frames. For this reason, for most of the measures discussed below I have run the model starting from the steady state for 100,000 time steps.

Measures from Information Theory

The mathematical theory of communication as postulated by Shannon, is a measure which has found applications in many fields, and is slowly gaining credence in biology. This theory provides with two measures, firstly it can serve as an indicator of the uncertainty of a variable given a set of observations via the entropy. And secondly, it provides for a framework of comparing two uncertainties and establish whether information is shared between them via the Mutual Information (see section 1.2.1 formula 1.14).

In terms of the entropy, I used similar methodology as above to first extract this measure from the stochastic model at different parameters. The entropies of the variations of the K_3 parameters, are constantly higher than the entropies of the K_1 variants. This could be due to slight variations in the means, however, in contrast to the Fano Factor, increasing either parameters induces a change in entropy. At the highest observed parameter for both K_1 and K_3, the mean of the time series in both cases is equal to 200 (see **Figure 6.12**). In this case, the system contains slightly more entropy if it has a positive feedback.

An advantage of using information theoretical measure of entropy, is that it is relatively straightforward to compare the entropies of two variables, via the Mutual information. In my case, I computed the Mutual Information of the two entropies for different K parameter levels with same means. In this case, I was looking for regions where information was maximized between the two graphs, which intuitively translates to the biggest observed differences. Using this analysis it becomes apparent, that as the length of the data analyzed increases, the differences between the two parameters decreases, and eventually stabilizes at roughly 0.2 bits independent of the run (**Figure 6.20**). At lower repetitions, the differences between the two parameter runs becomes more apparent, this is due to the stochastic nature of the model. In conclusion, although the entropy of the system increases slightly with the positive feedback strength, the information shared between the two parameters is very similar, thus this might not be a good measure to capture differences under these circumstances.

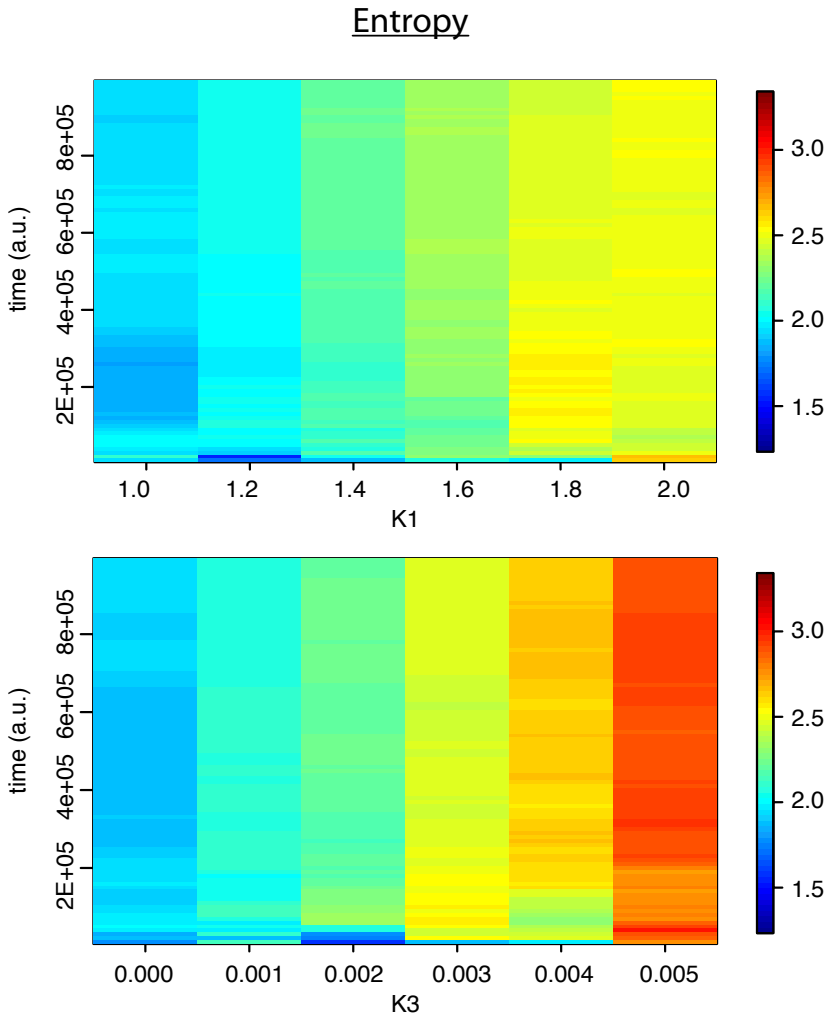


Figure 6.19 – Effects on the Entropy of A of variation of parameters in the stochastic positive feedback model. The entropy was calculated using the entropy and discretize functions in the entropy R package using the default parameters. Data was discretized into 20 bins of size roughly 10. The model was run for 100,000 time steps at increasing time windows (y-axis), for each value of either K_1 or K_3 (x-axis).

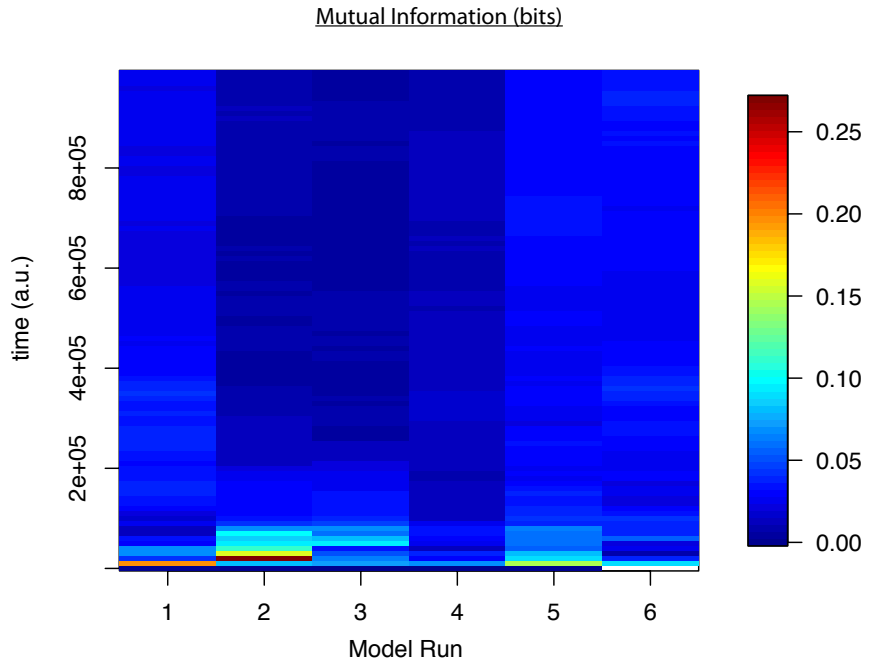


Figure 6.20 – Mutual information between stochastic model runs with similar means. The mutual information was calculated using the mi.empirical function in the entropy R package using the default parameters. Data was discretized into 20 bins of size roughly 10.

Seeing the results from the model, I did not analyze the data using the entropy function, as it was not obvious what the differences would be. Furthermore, there might be another reason as to why the entropy could be a misleading quantity for the presented data set. The measure of entropy was originally postulated by Shannon as a solution for a continuous variable, as opposed to the discrete data usually acquired in experiments. As such, outliers which play a hugely important role in biology might be overlooked when analyzing just the mean. In physical sciences, on the contrary, observations of variables very often obey the mathematical rules and behave as gaussians.

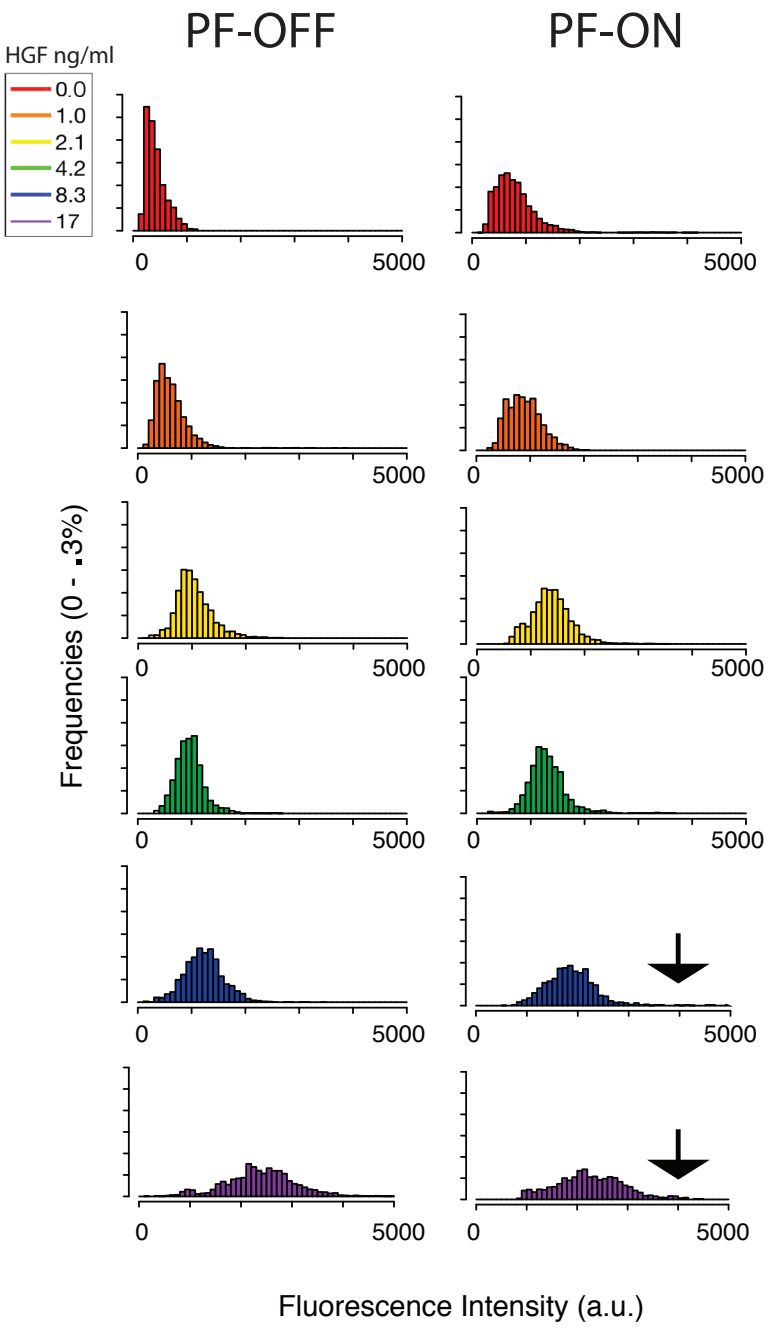


Figure 6.21 – Histograms of the experiment shown in figure 6.8. All histograms are divided into 800 bins of size 100, and normalized to frequency of counts. HGF_{ext} concentration increases downwards. Arrowheads point towards "outlier" cysts.

The design of the PF cell line was meant to serve as a way of measuring the

input and output of the system by means of the fluorescent molecules. However, given the differences in degradation of the two fluorophores (**Figure 6.6**) it was not possible to assess the question of mutual information experimentally. In this section will first show as above, how the Entropy of the theoretical model behaves in comparison with the non positive feedback. And later on I will briefly touch upon the theoretical framework I envisioned to address the question of mutual information, which, although not applied to the data, might serve as a guide to biologists working on sender and receiver systems.

The basic idea was to address the question of whether HGF_{int} is received by the receiver module, and if there is a communication channel, then what the rate of information transfer is, and whether the positive feedback influences this rate as time progresses. The framework to address this question, again comes from information theory (**section 1.2.1**). In the PF cell, HGF_{int} is coupled to a tdTomato fluorophore, I hypothesized I could use the measured fluorescence as a proxy of how much HGF_{int} was being produced at any given time. As discussed above, one method to estimate the relationship of two variables is through their shared or mutual information (MI). In the context of the PF cell line, the mutual information and entropies could be calculated using **equation 1.15** so that:

$$\begin{aligned} MI(GFP;TOM) &= H[GFP] + H[TOM] - H[GFP,TOM] \\ H[GFP] &= -\sum_{GFP} P(GFP) log P(GFP) \\ H[TOM] &= -\sum_{TOM} P(TOM) log P(TOM) \\ H[GFP;TOM] &= -\sum_{GFP}\sum_{TOM} P(GFP,TOM) log P(GFP,TOM) \end{aligned} \quad (6.7)$$

In order for this approach to provide robust insight, the data collected must be sufficiently dense to generate a good approximation of the theoretical probability distribution function (PDF) of the system for each variable. From what I have seen, the PDF should be composed of a minimum number of roughly 150 cysts per meassurement, although as we have seen, this might vary depending on the condition, and one must decide whether outliers are important in this context.

6.3 Morphogenic engineering of MDCK cells

One of the long-term goals of biologists, is to understand and perhaps recreate both temporal and spatial biological patterning mechanisms observed during the process of embryonic development. Biological feedbacks have been shown to significantly enhance the speed of propagation of spatial signals by generating signaling waves [Chang and Ferrell, 2013]. In this section, I set out to explore whether the weak autocatalytic feedback presented in the previous section affects signaling transmission in the spatial dimension. I will also describe a novel form achieved by the morphogenetic engineering of MDCK cells.

I initially set out to mimic the HGF spatial diffusion experiments presented in **Part II** this time with the PF cell line. Cells were observed for up to 48 hours, and no significant difference in signal propagation in the presence or absence of the positive feedback could be detected(data not shown). Evidently the low observed strength of the positive feedback discussed in the previous section was not sufficient to transmit a signal in space. This again, could be due to the small number of cysts which are hyper-sensitive to the feedback. This weak production and secretion of HGF_{int} from the PF cell line is not enough to create a diffusible signal that covers the distance between cysts, which can range from a few microns to a few hundred microns. This problem could possibly be overcome, by decreasing the space between cysts, or by creating a continuous cell epithelium.

6.3.1 Matrigel reduces HGF mitogenic properties preserving pMMP1-Response

One of the most prominent effects of HGF on MDCK cysts is the formation of tubule like structures. Ultimately at higher HGF concentrations, the cysts dissociate into single very motile cells (**section 6.2**). When cells are grown at close proximity to each other, then this motility interferes with a signal propagation mechanism driven solely by diffusion, as the cells would come in contact with each other. To abrogate this phenotype, I explored growing cysts in a Matrigel substrate as it has previously been reported that MDCK cells grown in Matrigel do not tubulate under high doses of HGF stimulus [Santos and Nigam, 1993, Kwon et al., 2014]. These results could be recapitulated, and furthermore, I could show that the HGF to GFP signaling channel remained functional in cells grown in a Matrigel substrate (**Figure 6.22**).

In order to study temporal or spatial signal propagation, the protocols so far discussed for quantitative fluorescence imaging require that cells lie on a single Z-plane. This is necessary for the quantitative analysis, so that there is no spectral overlap from point spread functions of cysts lying directly above or underneath

the focal plane. In this regard, Collagen I substrate was ideal, as in a two layer system, the top layer could easily be pealed off with forceps, leaving a single layer of cysts at the interface. Matrigel however, is somewhat less rigid, as it is mainly composed of Collagen IV fibrils which in contrast to Collagen I, form more heterogeneous molecular structures. This made the pealing off of the top Matrigel layer almost impossible.

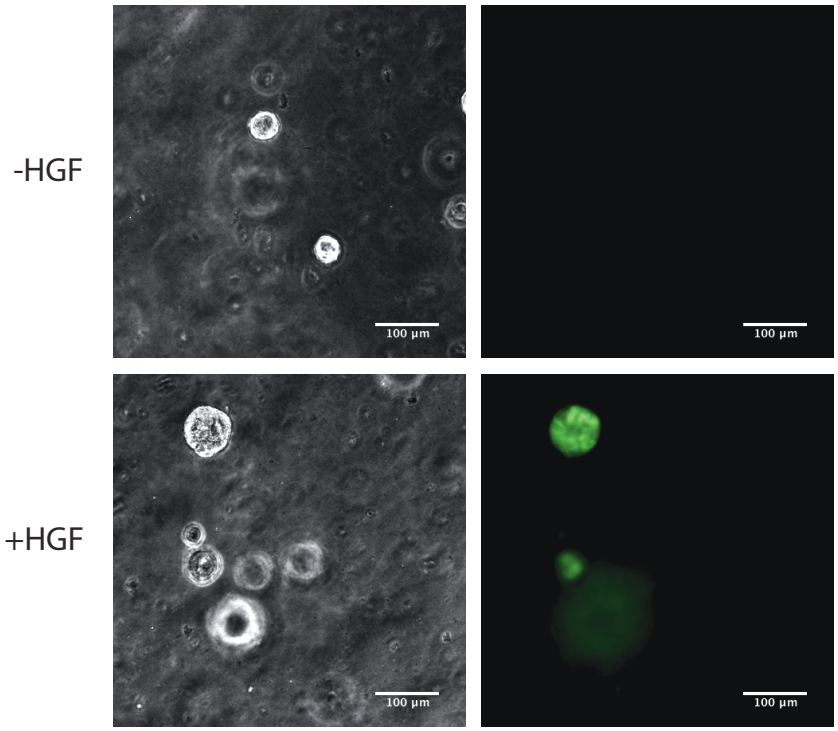


Figure 6.22 – The positive feedback cell line responds to HGF when cells are grown in Matrigel. 50,000 cells were diluted in 1ml Matrigel (3mg/ml final concentration in MEM) and grown for 5 days. After which 50ng/ml HGF was added to the medium, and then cells were imaged after 24 hours.

I reasoned that introducing further collagen I to the matrigel mix would confer the right balance to the matrix between rigidity and non tubulation properties. In **Figure 6.23** I show the results of growing the PF cell line under different ratios of Collagen I to Matrigel. I first coated wells with a layer of the mix at the indicated ratios (bottom), and let it solidify for 60 minutes. I then mixed cells into a second mix at the indicated ratio (top) and carefully deposited this onto the solid bottom layer. The results of **Figure 6.23** could not be technically reproduced for the higher Matrigel concentrations.

Generally in **Figure 6.23**, I observed a difference in phenotypic response to

HGF depending on the ratio of Matrigel:Collagen used. For example, at higher concentrations of Matrigel in the scaffold membrane, tubulation subsided substantially in response to an HGF stimulus. Notwithstanding this observation, the problems with rigidity remained, as it was impossible to remove the top layer reproducibly even with the lower Matrigel ratios making it impossible to image the cells quantitatively under these conditions. The most reproducible condition with the least tubulation was found to be the one where both the bottom and the top layer consisted of 40μl Matrigel and 120μl Collagen I.

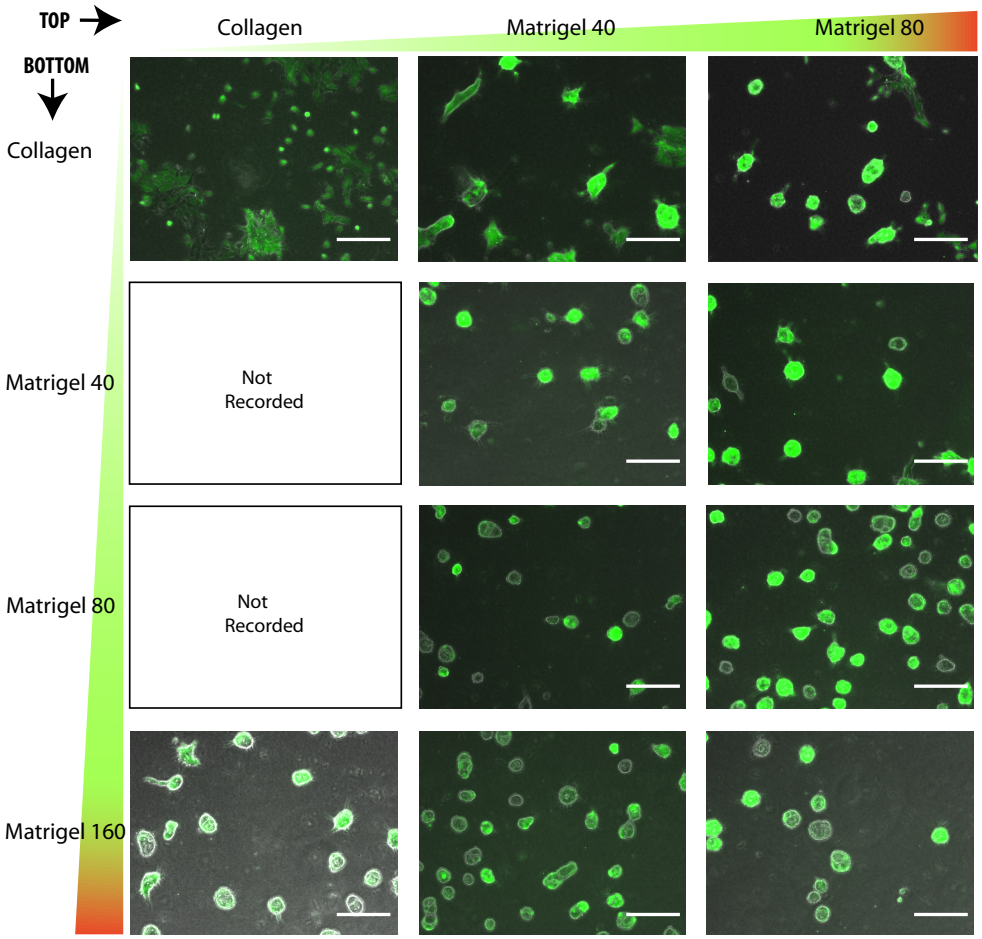


Figure 6.23 – Different ratios of Matrigel:Collagen have phenotypic effects on MDCK morphogeny. Cells were seeded on a precoated well with indicated amount of Matrigel:Collagen ratio. The total volume was kept constant at 0.2 ml and the numbers indicate the volume of 16mg/ml Matrigel which was mixed with collagen at both bottom and top layers. Cells were stimulated with 50ng/ml HGF.

2D micromass

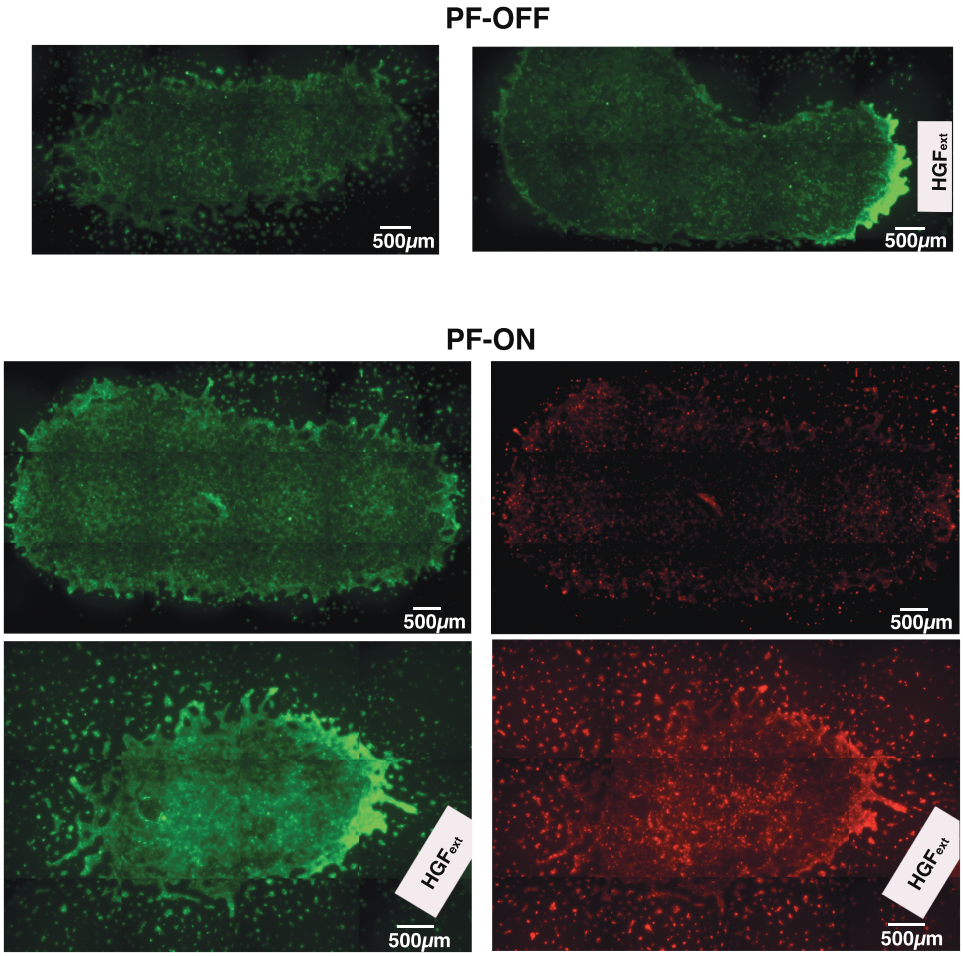


Figure 6.24 – Micromass cultures respond to HGF on the edges of the colonies. 40,000 cells were "sandwiched" between two collagen layers. Only the GFP channel is shown for the condition without positive feedback (PF-OFF). Scale bars are slightly different in each image, as the colony shapes and sizes are very variable. A 2μl drop of 10ng/μl HGF was added to a rectangular Whatman paper (white rectangle on figures), and cells were imaged after 24 hours.

In the process of experimenting with the different substrates for MDCK cells, I made the observation that occasionally when cells were plated on a matrigel plate and let to settle, after which a second layer was added, groups of cells would form cyst-like structures. I observed that cells seeded this way, would form cysts after 24 hours of addition of Matrigel/Collagen mix. Interestingly, when cells were seeded this way only with collagen 1, the cells grew as 2D monolayers, albeit at a

slower rate. This observation led me to attempt to explore whether I could make bigger continuous cyst like structures which would help me reduce the space in between the cysts, ideally to 0.

I cultured cells in micromass cultures such as is done routinely in mouse embryonic fibroblasts [Raspopovic et al., 2014]. I seeded PF cells at a concentration of 10,000 cells/μl on a collagen substrate, and then added a local source of HGF_{ext} impregnated on a sterile Whatman paper. Cells grew as a single large colony and upon addition of a top layer of collagen formed tubule-like structures mainly at the edges of the cultures. Upon addition of the localized HGF_{ext} source, cells in fact responded by expressing GFP at the edge of the colony **Figure 6.24**. Furthermore, in the presence of tetracycline, the tubule like structures at the edges of the colonies, activated both the GFP and the tdTomato reporter.

Taken together these results support previous observations of baso-lateral localization of c-met. It seems reasonable to suggest that, since MDCK cells express the c-MET receptor on the baso-lateral membrane, cells at the edge of such a micromass would have more exposed receptors if an HGF stimulus was applied locally. This also supports the theory mentioned previously in **section 6.2**, where I postulated that the presence of outliers in **Figure 6.3** could be due to mislocalization of either c-met given that the internally produced HGF is mostly secreted apically (**Figure 6.4**).

The results mentioned above were encouraging, however there was still no clear evidence of spatial signal propagation within the micro-mass culture. Although the edges appeared to be responsive to HGF, It was unclear whether this difference was due to day to day experimental variation. Furthermore, as the edge of the micromass is very heterogeneous, the quantification of a spatial signal propagating is non-trivial. With this in mind, I reasoned that the signal might be able to propagate in a structure composed of "edge" only cells.

6.3.2 Cellular Cables

From the micromass culture experiment above, I concluded that the responding cells at the edges had the basolateral side exposed to the HGF_{ext} source. In order to this to happen, cells must mechanically fold onto each other, and thus, would form cyst like structures at the edges. Following up on this observation, I attempted to generate a cell structure composed of edges only, as this could possibly be the ideal system to study spatial signal transition in a continuous system. To do this, I generated a scaffold for the cells to grow into such structures. A schematic drawing of this and the other methods used in this work is shown in **Figure 6.25a**. The exact method used is described in **Chapter 8**. Representative images of the different stages in this protocol are shown in **Figure 6.25b**).

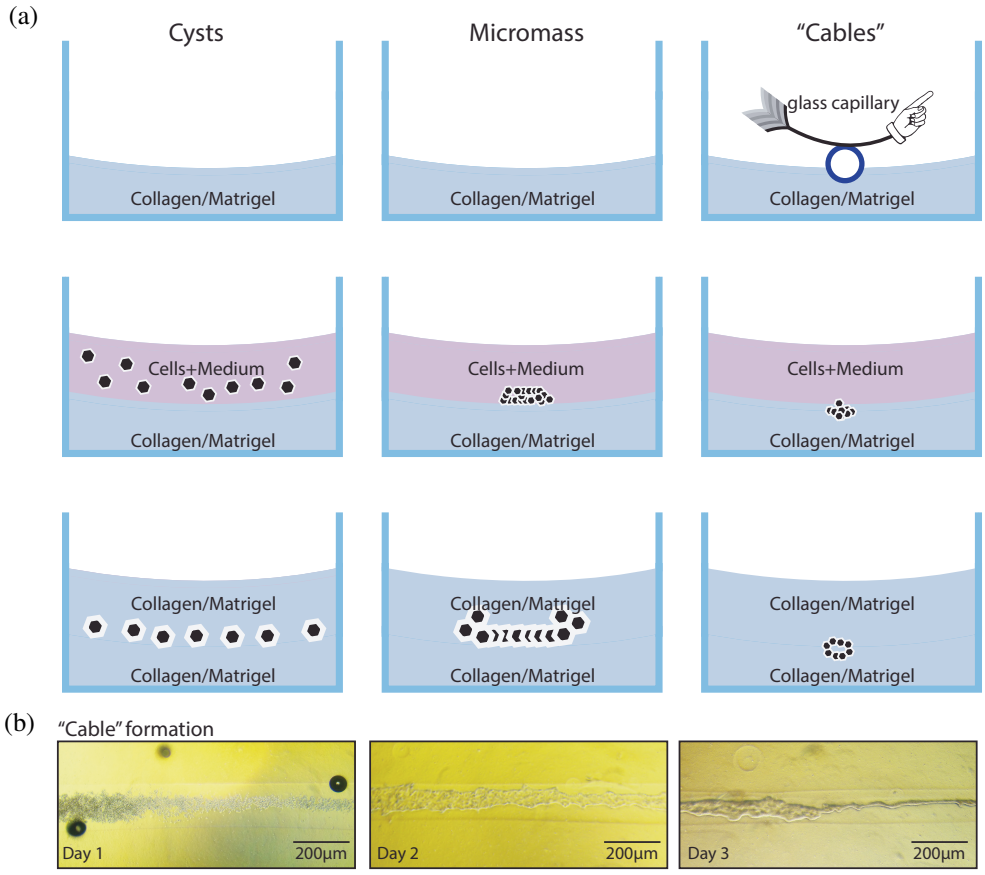


Figure 6.25 – Different morphogenic growth methods for MDCK cells. MDCK cells grow in collagen/matrigel matrices, form a polarized cell layer with a pseudo lumen. A schematic diagram depicting a cross section of a cell chamber is shown in **Figure 6.25a**. Depending on the initial growth conditions, these cells will grow into roughly 100cell diameter cysts left column, a micromass colony (middle column), or a cable (right column). In **Figure 6.25b**, the three stages of the formation of cables are shown.

I next proceeded to characterize these "cables" by immuno-labelling actin and cadherin proteins. I was interested in seeing whether the well characterized apical-basal polarity of MDCK cysts was maintained in these structures. In **Figure 6.26a lower panel** these "cables" are shown to retain the apico-basal polarity of cysts, with actin localized apically and cadherin junctions localized baso-laterally. Attempts to label the c-met receptor failed, however it is well established that the receptor co-localizes to the tight junctions with E-cadherin, thus, I also expected that the c-met receptor would be baso-laterally localized.

Another interesting conclusion from these experiments is, that some cables which had larger "diameters" were observed to form multiluminar structures, hint-

ing to a physical constraint of the "cables" diameter **Figure 6.26a upper panel**. All attempts to generate consistently sized "cable" failed using this protocol, thus, all subsequent experiments do not take this architectural constraint into account.

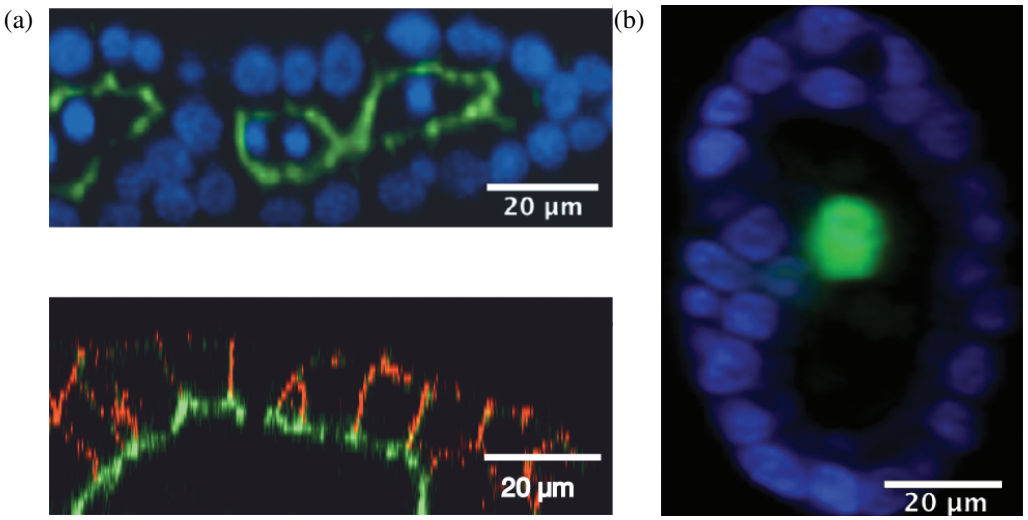


Figure 6.26 – Cables are diverse and show bipolarity as cysts. Representative confocal image sections show that actin (green) was localized at the apical side of the cables **6.26a top panel**, and E-Cadherin (red) was localized at tight junctions and apically **6.26a bottom panel**. PF cells with GFP are localized in the inside of the cables **6.26b**.

The experimental data shown in **Figure 6.26a** were performed with wildtype MDCK cells. I found that both the receiver and the PF cell line formed "cables" when grown under this protocol. I next imaged the the PF-Cell line under the confocal microscope. In this case, GFP positive cells were seen to localize mainly on the inside (apical side) of the cables even in the absence of HGF_{ext} (**Figure 6.26b**). This seems to indicate that the cables and possibly the cysts do not express the pMMP1 promoter homogenously.

HGF response in cables

I deemed this novel growth form intriguing enough to explore whether these structures would respond to HGF_{ext} and propagate the signal in the presence of a positive feedback (PF). The "cables" after all fulfilled the requirement for a quasi-continuous system. Moreover the observations from the micromass experiments in **Figure 6.24**, indicated that the morphological change of the MDCK cells, would expose their basolateral sides in a direction that would allow HGF to come in contact with c-Met.

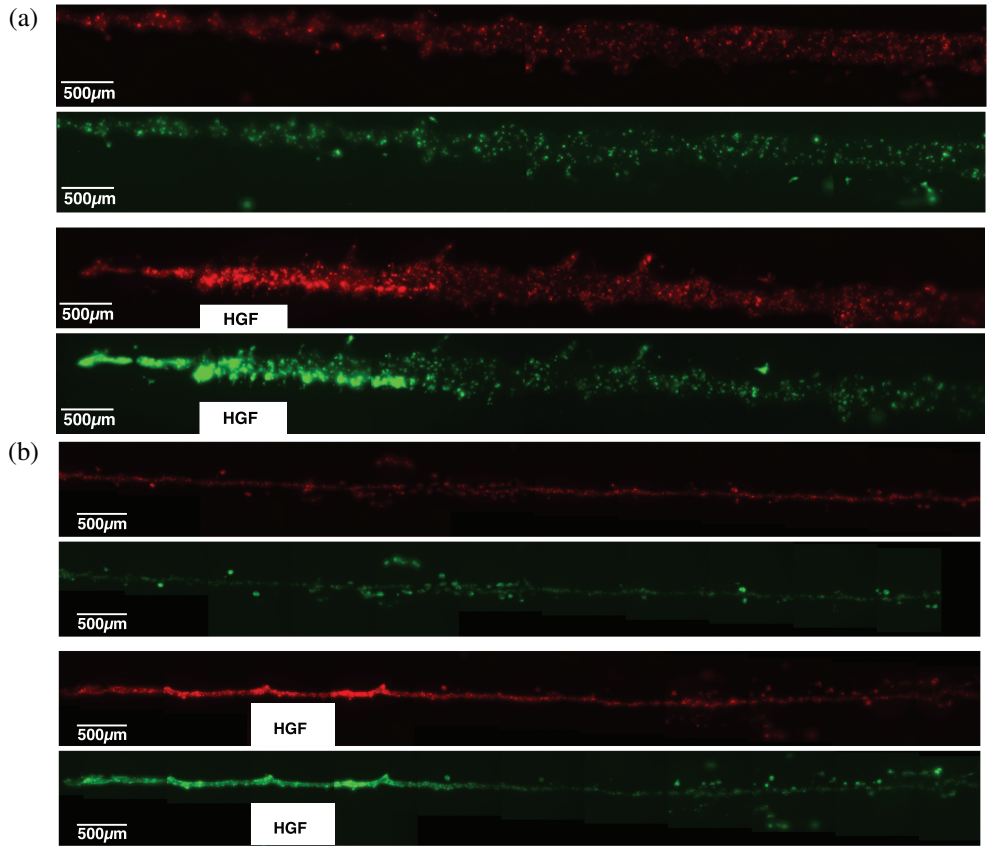


Figure 6.27 – Cable response to a localized HGF_{ext} source after 24 hours. PF cells were cultured as cables as described and in all experiments shown, tetracycline was added 4 hours prior to the HGF_{ext} (**PF-ON**). Thick cables (6.27a) contain roughly 40,000 cells whereas thin cables (6.27) contain 4,000 cells only.

I setup "cables" to test whether there was a response to a localized source of HGF_{ext}. As mentioned above, there were two possible configurations for the "cables", 1) a thicker multiluminar structure, and 2) a thinner cylindrical structure. It was generally easier to form the thicker structures, however as with the micromass experiments above, I found that the response to the HGF source would be somewhat localized to the edge closer to the source (figure **Figure 6.27a**). In contrast, when cables are mainly "thin" as is shown in **Figure 6.27** then as expected the response is homogeneously distributed.

In both the above experiments I have shown how the cables respond to HGF in the presence of tetracycline (positive feedback on) after 24 hours. In neither of these cases does the signal seem to propagate far from the source in the time window of observation. Regardless of the thickness of the cables, it would seem

that the distance traveled of the signal is similar as to what would be expected from diffusion only processes.

6.3.3 Spatiotemporal response in MDCK cables

Next I tested whether there were significant differences between the non-positive feedback and the positive feedback case in terms of spatial signal propagation. For the analysis, with the help of Yuriy Alexandrov at Imperial College, I modified the previous Matlab script used to analyze cysts, which can be found in the Materials and Methods section.The modifications of the code was to automate the detection of the cable structures. As the cables do not change shape in time, it was no longer necessary to detect and track multiple cysts in one image. Instead, it was now necessary to identify the source of HGF, and to establish a coordinate system to measure the GFP and tdTomato response as a function of distance to the source in time. This allowed me to extract quantitative data from fluorescent images and to get an initial overview on how the cables responded to an HGF stimulus.

I set up multiple cables as the ones described in **Figure 6.27**. Furthermore, as there was no apparent effect visible after 24 hours, I imaged the cells for prolonged periods of time after addition of HGF_{ext}. To display the spatio-temporal dynamics of the GFP signal, in **Figure 6.28** I represent the results of individual experiments in form of a kymograph observed over a period of 7 days. As expected, the system has a very high variability which stems both from the experimental setup (extrinsic variability), and the internal configuration of the cables (intrinsic variability). There does seem to be a trend towards temporal signal propagation by the positive feedback (**PF-On** of **Figure 6.28**). In all replicates of the PF-on condition, the GFP signal persists even after 7 days, in stark contrast to the PF-OFF, which seems to switch off after 4 days.

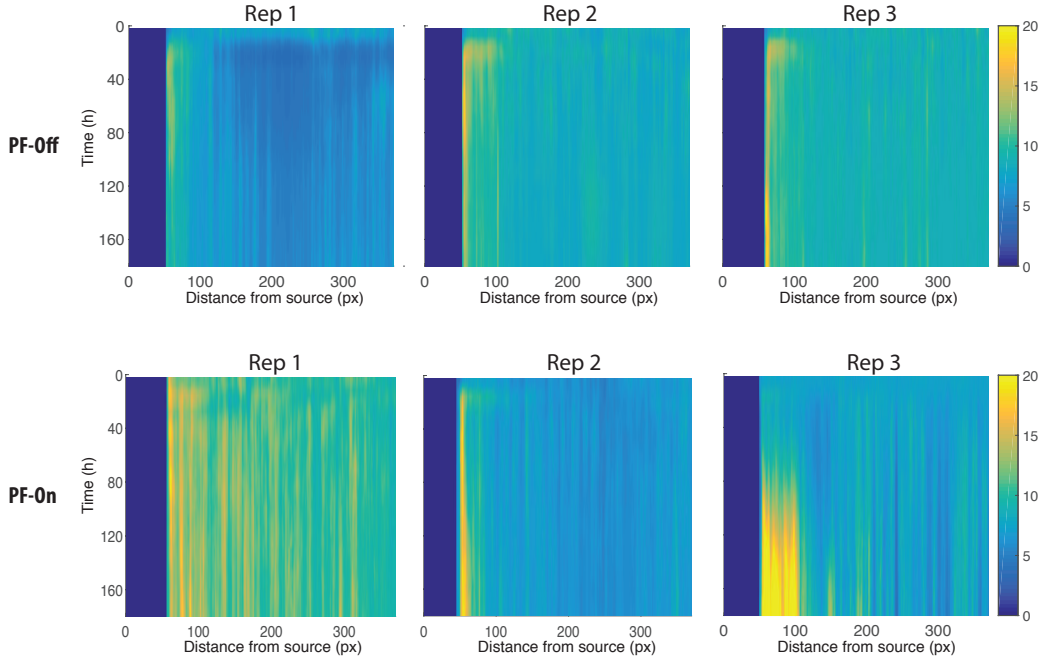


Figure 6.28 – Kymographs of three replicates showing variability between experiments. HGF_{ext} was applied on a sterile Whatman paper and, cables were imaged in 4 hour intervals for 170 hours (7days). For the normalization of the image data, I used the data from each cable from the 20% end furthest away from the source, if this area did not show any significant temporal change in GFP fluorescence. The cables used in this case were like the thicker ones in **figure: 6.27a**.Distances are given in "image pixels", with each pixel representing 1.5μm

In light of the persistent temporal GFP signal detected after 7 days of observation, I reasoned that HGF_{int} was being produced in the PF-ON cells as opposed to the PF-off cells. It was possible, that the produced HGF was accumulating close to the source, and that the diffusion thereof could be slower than expected.Hormonal signaling across tissues by diffusion only, and in the absence of vascularization might be a slow process, or require certain thresholds to occur. If this was the case, then spatial propagation would begin only at a later point in time. I explored how the system responded over extended time periods, the results are shown in **Figure 6.29**. In this case I am displaying the raw data of these longer observations. The GFP signal was visible throughout the cable structures, and as such I could not apply the same normalization as in **Figure: 6.28**. Once again, I could detect a signal in close proximity to the source of HGF_{ext} which persisted for up to 7 days in the PF-on but not in the PF-off condition.

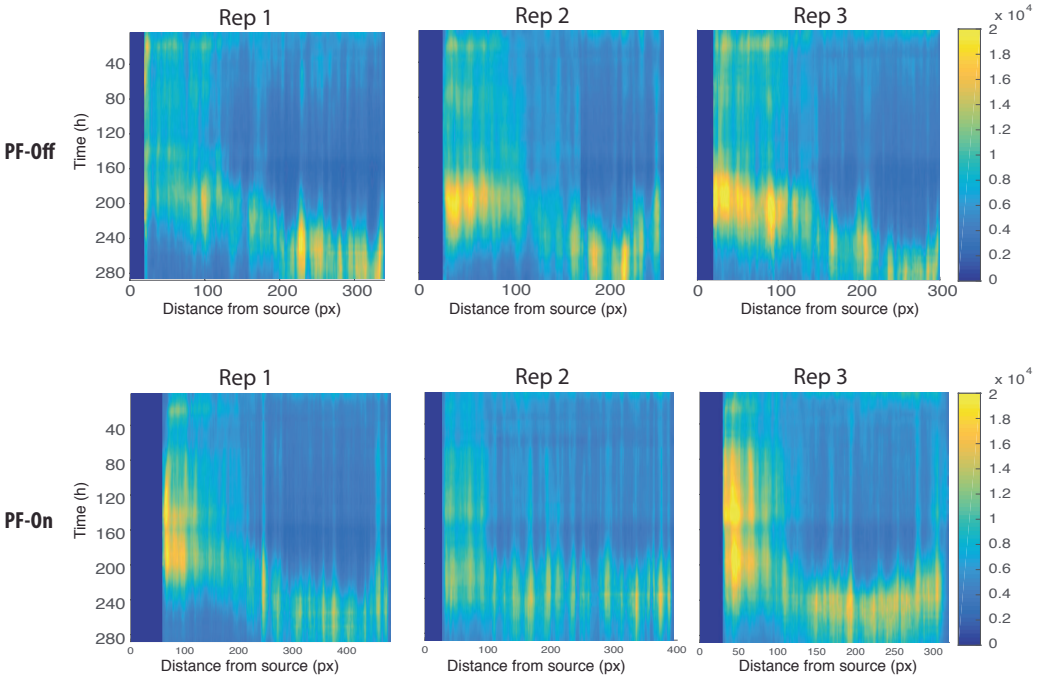


Figure 6.29 – Kymographs of three replicates showing variability between experiments. HGF_{ext} was applied on a sterile Whatman paper and, cables were imaged in 4 hour intervals for 280 hours (10days). Raw values are shown. The cables used in this case were like the thinner ones in **Figure: 6.27a**.Distances are given in "image pixels", with each pixel representing 1.5μm.

A very notable feature in these experiments, is that even in the absence of the positive feedback a distinct signal propagated spatiotemporally throughout the cable structures. On closer inspection, however, It became apparent that this signal was present even in the absence of any HGF_{ext} (data not shown). As can be seen in **PF-off Rep 2** and **PF-on Rep 3** of **Figure 6.29**, this moving signal of GFP expression emanated from the ends of the cables after about 180 hours of observation. The signal seemed to precede cell death, as assessed by morphological changes in the overall cable structure (i.e. lack of cell movement, and apparent disruption of tight junctions within individual cells). The death could be attributed to the fact that cells remain under observation for prolonged periods of times, without the replenishing of growth medium. Upon closer inspection of the cables after the experiments, it would appear as though an unidentified microbial contamination was present at great number in the cables. It was not immediately clear at what point this contamination was acquired, or whether it contributed to the death of the cables.

In conclusion it would seem that the positive feedback can transmit the initial impulse of HGF temporally in the section of the cables exposed to the stimulus.

This observation supports the hypothesis, that the PF cell line indeed undergoes an autocatalytic reaction. From these experiments it is also clear that the internal production of HGF is not not propagating in space along the cables. As of this moment it is not clear why this is the case, however several hypotheses have been briefly touched upon above. A likely explanation, as I have shown, is that the secretion of HGF and the localization of the receptor is in two distinct parts of the cables. In the next section I will show preliminary results of how one could overcome this obstacle.

6.3.4 Disruption of epithelial integrity enhances spatial signal propagation

As I have shown, MDCK cells respond to a HGF stimulus when grown both as cysts in collagen, or in matrigel as assessed by a GFP reporter. Furthermore, when cells were coerced to extend and form cylindrical "cables", these structures also responded to a stimulus of HGF. The cells all contain a putative autocatalytic module of HGF, that, when turned on propagates a signal temporally. In contrast to my expectations, regardless of the morphological form that MDCK cells were grown in, there was no apparent spatial propagation of HGF from a localized source. One potential culprit, is that the intrinsic polarization of MDCK cells causes a spatial segregation of morphogen and receptor to different sections of the cells. To solve this, I reasoned that polarity disruption might be a strategy to enable communication between cells.

A wide range of strategies have been proposed to disrupt cell polarity reviewed extensively in [Deli, 2009]. Genetically, for example, it is possible to redirect the sorting of proteins by adding signal peptides to the amino acid sequences of a protein of interest. Alternatively, it is possible to chemically degrade the polarity, by attacking the tight junctions which serve as a quasi barrier between apical and basal side. A key component of tight junctions is the calcium dependent binding of the extracellular component of E-Cadherin. In [Deli, 2009], two studies are mentioned by which addition of Ethylenediaminetetraacetic acid (EDTA) to cells, increases the permeability of epithelial monolayers to small molecules. The mechanism proposed being that the binding affinity of EDTA to calcium cations is higher than that of E-cadherin. Addition of EDTA to the medium, sequesters calcium reducing the availability at the tight junctions which in consequence are disrupted. Among the multiple consequences of this disruption are the permeabilization of the cell layer, and further, the disruption of polarity in cells.

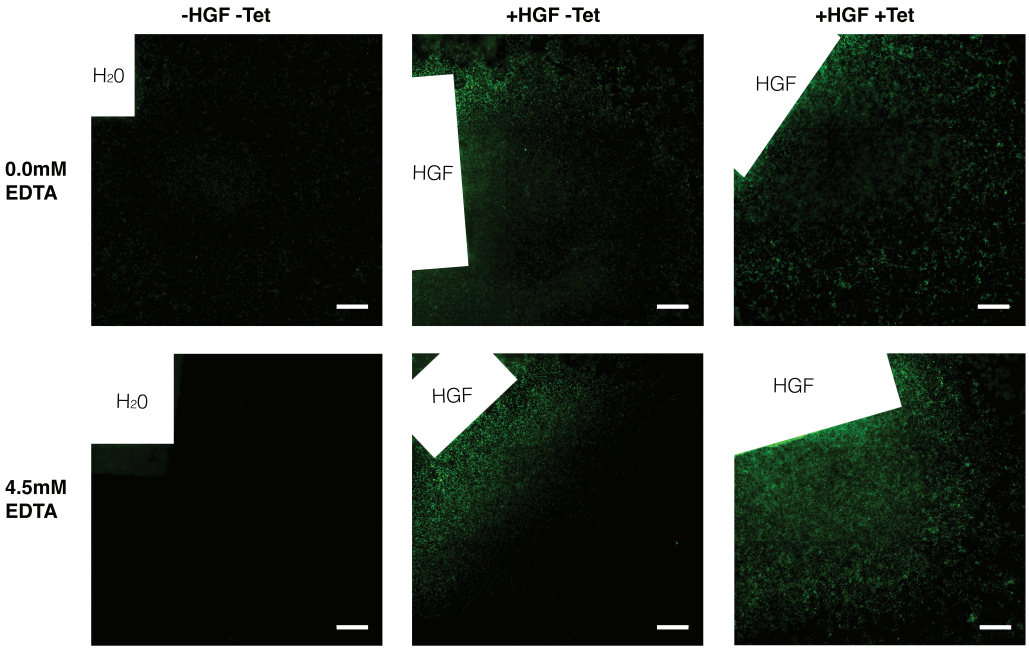


Figure 6.30 – Disruption of E-cadherin tight junctions via EDTA in 2D monolayers. PF-cells were plated on a solidified collagen layer and grown until completely confluent. HGF_{ext} was applied on a sterile Whatman paper where indicated, after which cells were covered with a collagen layer. Tetracycline and EDTA were added as indicated. Cells were imaged 13 hours post stimulus. Experiments were performed by Natalie Scholes Sian. Data analysis and interpretation was performed by me and N.S.S.

To test whether inter-cellular HGF signaling could be enhanced in MDCK cells by disrupting tight junctions, I grew a confluent layer of the positive feedback (PF)-cell line in 2D cultures on top of a collagen layer, and treated the cells with the strong bivalent cation chelator EDTA. Indeed after treatment with 4.5mM EDTA for 1 hour, MDCK cells displayed a rounded morphology, close to an idealized sphere with no extrinsic force interactions. In **Figure 6.30 I** show that cells treated with EDTA can indeed respond to HGF by expressing GFP. Qualitatively, the GFP signal seems to extend further than in the cells not treated with EDTA. In the presence of feedback there was a strong activation of all cells, even in the absence of EDTA. As I had previously observed high variability in these experiments, and given that the experimental setup was less laborious than the cysts or cables, I increased the signal to noise ratio, by performing more observations.

I repeated the above experiment with EDTA biologically, and increased the number of technical repeats for each biological replicate. In **Figure 6.31**, I show the results of two biological replicates each consisting of the means of five technical repeats (n=5). To calculate the means and standard deviations, I first generated the kymographs for each individual technical repeat, and then averaged the images

on a pixel by pixel basis.

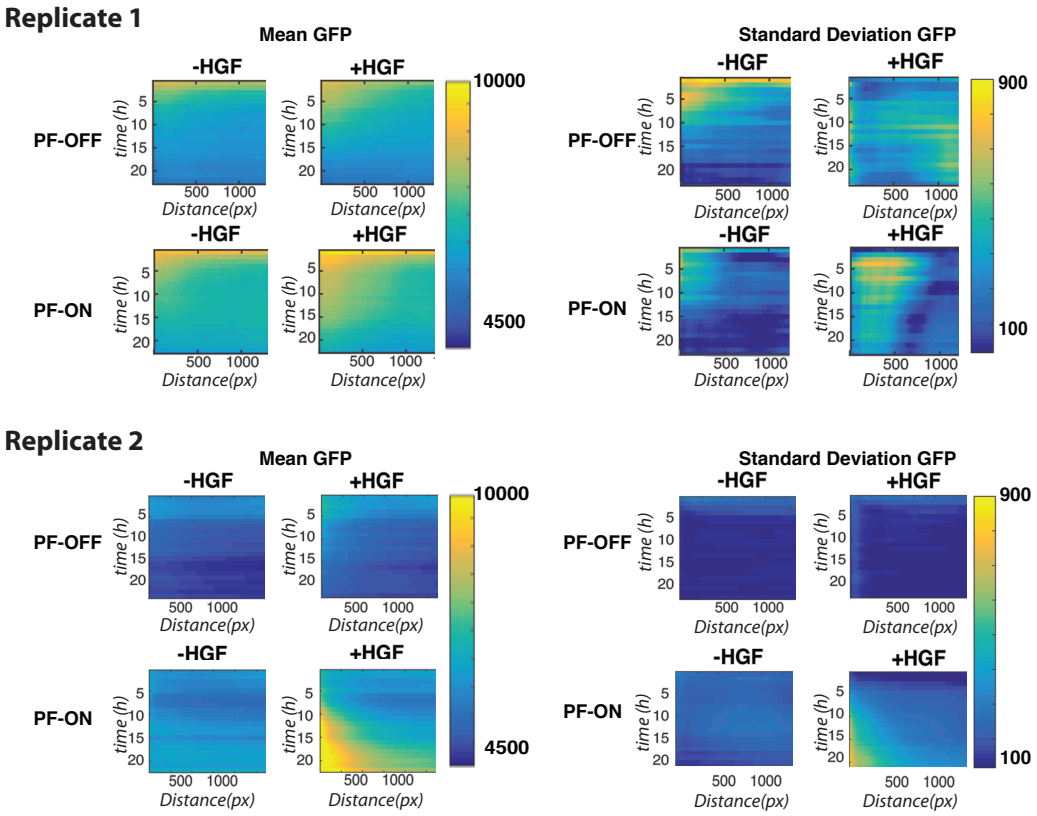


Figure 6.31 – Disruption of E-cadherin tight junctions via EDTA enhances positive feedback signalling in 2D monolayers. HGF_{ext} was applied on a sterile Whatman paper and, cells were imaged continuously for 24 hours post stimulus. Each condition for the replicates is the averages fluorescence intensity of kymographs of 5 technical repeats. Standard Deviation is calculated from the same 5 technical repeats. Distances are given in "image pixels", with each pixel representing 1.5μm.

There were similarities in both biological replicates. For example, the condition with a turned on positive feedback and an initial HGF_{ext} stimulus could be distinguished from the other samples in that the GFP response persisted both spatially and temporally compared to the other conditions within each replicate. These results support the hypothesis that the disruption of the tight junctions may indeed enhance the diffusion of HGF produced by MDCK cells. In both biological replicates, the condition with the positive feedback displayed the highest variability (as assessed by the standard deviation plot in **Figure 6.31**). This again serves to highlight the properties of the system.

The background signal detected for each individual experiment differs signif-

icantly. On the technical side, the background in replicate 1, hovers around 5000 fluorescent units, whereas in replicate 2 it hovers at 4500. This difference is also markedly high between technical repeats, evidence by the standard deviation in replicate 1, and can be attributed to the intrinsic background fluorescence of collagen in the blue/green spectrum. In the experiments with cysts, it was feasible to estimate the background fluorescence, by using the area where no cysts were present. In a epithelial cell layer, there is no such cell-free area, thus the quantitative measures will be non-trivial.

On the biological side, there are differences in respect with the timing and duration of the GFP signal after addition of HGF_{ext}. In the case of cells without feedback, both replicates show an initial increase in GFP signal followed by a fade out at roughly t=10h. The maximum distance reached by the signal is roughly at 800μm from the source. In the case of cells with feedback, the variability is higher. In replicate 1, the signal increases monotonically and persists temporally up to T=20h. No difference was observed in terms of the spatial propagation. In replicate 2, there is a temporal lag in the response, however at T=10h, the signal initiates and propagates at a rate of 1.1μm/min, which is somewhat below the estimated diffusion coefficient of HGF we previously measured in [Carvalho et al., 2014].

The reason for the biological variability within experiments remains unclear. There are several factors that might contribute to this variability, some of which can be addressed technically. First and foremost, a technical solution to the method of delivery of HGF_{ext} throughout the spatial experiments must be found, as the cellulose Whatman paper used, might strongly influence the diffusion properties of HGF_{ext}. Furthermore, it is unclear how crucial the timing of application of the reagents including, tetracycline to derepress the PF or EDTA to disrupt the tight junctions might influence the system. The nature of gellification of Collagen/Matrigel, makes it difficult to exactly reproduce this timing, as it can sometimes take up to 2 hours for the matrix to solidify. In this sense, it could be interesting to attempt to grow the cells on different even inorganic substrates, which might be more homogenous in their composition.

Chapter 7

CONCLUSIONS

A main goal for biologists is to understand the incredible spatiotemporal order that emerges from the sack of organic and inorganic chemicals we refer to as cells. Autocatalytic reactions are hypothesized to play a fundamental role in the generation of order in these and other systems [Kauffman, 2011]. Synthesis can be proof and aide of understanding, thus to help understand autocatalytic reactions, in this work I presented the engineering and characterization of an inducible autocatalytic positive feedback (PF) into the MDCK cell line reported in **Part II** by genetic introduction of a Hepatocyte Growth Factor (HGF) module. The effects of the PF on both the temporal and spatial response of an externally supplied source of HGF_{ext}, were weak and I proposed possible solutions to strengthen this interaction.

I also proposed the adoption of a universal measure of strength of feedbacks in biological systems through the feedback gain. I demonstrate how to measure this gain in the MDCK cysts. Using this method, I can detect the presence of feedback only at certain concentrations of initial HGF_{ext} stimulus, and found a maximum gain of 0.4679 to the system at a concentration of 4.2ng/ml HGF_{ext}. The fact that feedback gain was not observed at other concentrations of externally supplied HGF, seems to indicate inefficient communication or some sort of negative feedback present in the system. I presented a hypothesis as to why this might be, namely the mis-localization of HGF_{int} secretion in respect to the c-Met receptor.

Roughly 24 hours after an initial stimulus of HGF_{ext}, the mean fold activation of GFP reaches a prolonged steady state. Here, the levels of GFP in cysts with the PF turned on remain invariably higher than their non-PF counterparts. Although this observation is supportive of the presence of a feedback, a higher mean does not offer an irrefutable proof thereof. This was clearly evident in a stochastic model presented in 6.2.3. I used this model to present three different mathematical methods that can accurately distinguish the temporal effects of a process with feedback from one without. On the one hand two measures of time dependent self

similarity, the autocorrelation function and the Hurst Exponent of the rescaled range, both increased in the presence of feedback as opposed to the non-feedback process. On the other hand, I found that the temporal "noise" of the stochastic process, as estimated by the Fano factor of the time series, also increases in the presence of feedback.

All these measures of feedback have two important caveats. Firstly, a non feedback process is needed to provide a reference although the inducible system presented here addresses this issue. Secondly, the increases in both autocorrelation and Fano factor, become evident only after a large number of observations. This caveat is somewhat more challenging to overcome in biological experiments, and will depend on the available resources. For example, when I measured the autocorrelation function of the data generated in section 6.2.1, in contrast to the model, the autocorrelation function decreased in the presence of PF. Thus, the autocatalytic feedback loop seems to reduce the temporal transmission of information in cysts that have a self activating HGF module. In terms of the Fano factor, the results seemed to agree with the model, the feedback increases the variability of the system. However, as with the gain, this effect is only detectable at certain HGF_{ext} input concentrations.

In parallel to the temporal characterization of the PF cell line, I presented a method to study the effects of feedback on the spatial transmission of information. This was important, as a feedback mechanism can help amplify multicellular signaling. MDCK cells have been grown in different forms, and are known to be quite maleable [Harris et al., 2012]. In section 6.3 I demonstrate my efforts to coerce cells to grow in different forms. It was clear to me from the temporal exploration that the feedback was weak, and given the non-continuous methodology to explore spatial diffusion used in [Carvalho et al., 2014], it would prove challenging to observe differences in the presence of feedback. I thus attempted several different protocols to create a continuous MDCK cell layer.

I confirmed that the substrate of growth for MDCK cells strongly influences their phenotypic response to an HGF stimulus. For example, cysts which were grown in a substrate containing matrigel, did not tubulate in the presence of HGF, but responded by production of GFP. MCDCK cells tend to form polarized layers, both when grown in 2D cultures or as 3D cysts. One key aspect of the polarization, is the localization of the receptor for HGF (c-met) to the basolateral side [Mojallal et al., 2014]. In 2D monolayers, it was non-trivial to deliver HGF to this side, as it is usually the "bottom" side. To generate a continuous layer of MDCK cysts, I developed a protocol for generating a scaffold for cells made of Collagen and Matrigel. When the cells were grown under these conditions, they formed long cylindrical structures resembling "cables". I demonstrated that these cables were still receptive to an HGF_{ext} stimulus. Moreover, these structures seemed

to confirm the weak temporal signal propagation via the PF. Interestingly, I could not detect any spatial propagation, even after extended periods of observation (up to 10 days).

Finally, I reasoned that disruption of the epithelial polarization in the MDCK cells would facilitate the intercellular diffusion of HGF produced internally (HGF_{int}). To test this hypothesis, I cultured cells in the presence of a bivalent ion chelator EDTA, which in principle should disrupt tight junctions by removing the necessary Calcium ions from the medium. Under these conditions, a propagation of GFP signal could be observed in response to an localized stimulus of HGF_{ext} only in the presence of the positive feedback.

Although culturing of the cables or cysts under the same conditions proved to be toxic to the cells, it is possible that further adjustments to the growth protocols could find a window in which signal propagation will be observed. It would be useful to further genetically engineer the MDCK cells by adding either basolateral localization signals to the HGF_{int} module, or by introducing a modified c-Met receptor with apical localization signals. From the current standpoint and results, it is conceivable that such modifications of the system will generate stable long-term spatiotemporal patterns.

Chapter 8

MATERIALS AND METHODS

8.1 Supplementary Materials for Part II

Carvalho A, Bárcena D, Senthivel VR, Zimmermann T, Diambra L, Isalan M. Genetically encoded sender-receiver system in 3D mammalian cell culture. Suplementary material. ACS Synth Biol. 2014 May 16;3(5):264–72. DOI: 10.1021/sb400053b

8.2 Experimental

8.2.1 Plasmid Cloning

pMMP-1-d2tdTomato-p2A-HGF I used the Gibson method to clone the Positive Feedback plasmid. I used the receiver plasmid used in ACS1 as a template for the MMP1 and the PEST degradation domain. HGF cDNA was provided by Andreia Carvalho in the lab.

Key	Position (bp)
Human MMP-1 promoter (bp -512 to +63)	4 - 627
tdTomato with c-Terminal PEST	634 - 2187
P2A linker	2188 - 2253
HGF	2254 - 4440
Gentamicin resistance	5683 - 6478

```
   1 CTACTAGCGC TTAACAAAGG CAGAAGGGAA CCTCAGAGAA
  41 CCCCGAAGAG CCACCGTAAA GTGAGTGCTG GGGGAGCTGA
  81 ACTTCAGTCA GTACAGGAGC CGAACAGCCA TCAGGTGCGC
 121 AGTGTTAGTA ATTCCACCCT CTGCCCTGGG AGCAAGGTGT
 161 GTGGAGAAAC CTGTAGCACT TTATGACCAT CAGAACCAGT
 201 CTTTTTCAAA AAGACCATGG AGTACTCTTT GACCTGTGTA
 241 TATAACAAGA ACCTTTCTCA AATAGGAAAG AAATGAATTG
 281 GAGAAAACCA CTGTTTACAT GGCAGAGTGT GTCTCCTTCG
 321 CACACATCTT GTTTGAAGTT AATCATGACA TTGCAACACC
 361 AAGTGATTCC AAATAATCTG CTAGGAGTCA CCATTTCTAA
 401 TGATTGCCTA GTCTATTCAT AGCTAATCAA GAGGATGTTA
 441 TAAAGCATGA GTCAGACACC TCTGGCTTTC TGGAAGGGCA
 481 AGGACTCTAT ATATACAGAG GGAGCTTCTC CCTATCAGTG
 521 ATAGAGACTT CCCTATCAGT GATAGAGATC CCTAGCTGGG
 561 ATATTGGAGC AGCAAGAGGC TGGGAAGCCA TCACTTACCT
 601 TGCACTGAGA AAGAAGACAA AGGCCATACc ACCATGGTGA
 641 GCAAGGGCGA GGAGGTCATC AAAGAGTTCA TGCGCTTCAA
 681 GGTGCGCATG GAGGGCTCCA TGAACGGCCA CGAGTTCGAG
 721 ATCGAGGGCG AGGGCGAGGG CCGCCCCTAC GAGGGCACCC
 761 AGACCGCCAA GCTGAAGGTG ACCAAGGGCG GCCCCCTGCC
 801 CTTCGCCTGG GACATCCTGT CCCCCCAGTT CATGTACGGC
 841 TCCAAGGCGT ACGTGAAGCA CCCCGCCGAC ATCCCCGATT
 881 ACAAGAAGCT GTCCTTCCCC GAGGGCTTCA AGTGGGAGCG
 921 CGTGATGAAC TTCGAGGACG GCGGTCTGGT GACCGTGACC
 961 CAGGACTCCT CCCTGCAGGA CGGCACGCTG ATCTACAAGG
1001 TGAAGATGCG CGGCACCAAC TTCCCCCCCG ACGGCCCCGT
```

1041 AATGCAGAAG AAGACCATGG GCTGGGAGGC CTCCACCGAG
1081 CGCCTGTACC CCCGCGACGG CGTGCTGAAG GGCGAGATCC
1121 ACCAGGCCCT GAAGCTGAAG GACGGCGGCC ACTACCTGGT
1161 GGAGTTCAAG ACCATCTACA TGGCCAAGAA GCCCGTGCAA
1201 CTGCCCGGCT ACTACTACGT GGACACCAAG CTGGACATCA
1241 CCTCCCACAA CGAGGACTAC ACCATCGTGG AACAGTACGA
1281 GCGCTCCGAG GGCCGCCACC ACCTGTTCCT GGGGCATGGC
1321 ACCGGCAGCA CCGGCAGCGG CAGCTCCGGC ACCGCCTCCT
1361 CCGAGGACAA CAACATGGCC GTCATCAAAG AGTTCATGCG
1401 CTTCAAGGTG CGCATGGAGG GCTCCATGAA CGGCCACGAG
1441 TTCGAGATCG AGGGCGAGGG CGAGGGCCGC CCCTACGAGG
1481 GCACCCAGAC CGCCAAGCTG AAGGTGACCA AGGGCGGCCC
1521 CCTGCCCTTC GCCTGGGACA TCCTGTCCCC CCAGTTCATG
1561 TACGGCTCCA AGGCGTACGT GAAGCACCCC GCCGACATCC
1601 CCGATTACAA GAAGCTGTCC TTCCCCGAGG GCTTCAAGTG
1641 GGAGCGCGTG ATGAACTTCG AGGACGGCGG TCTGGTGACC
1681 GTGACCCAGG ACTCCTCCCT GCAGGACGGC ACGCTGATCT
1721 ACAAGGTGAA GATGCGCGGC ACCAACTTCC CCCCCGACGG
1761 CCCCGTAATG CAGAAGAAGA CCATGGGCTG GGAGGCCTCC
1801 ACCGAGCGCC TGTACCCCCG CGACGGCGTG CTGAAGGGCG
1841 AGATCCACCA GGCCCTGAAG CTGAAGGACG GCGGCCACTA
1881 CCTGGTGGAG TTCAAGACCA TCTACATGGC CAAGAAGCCC
1921 GTGCAACTGC CCGGCTACTA CTACGTGGAC ACCAAGCTGG
1961 ACATCACCTC CCACAACGAG GACTACACCA TCGTGGAACA
2001 GTACGAGCGC TCCGAGGGCC GCCACCACCT GTTCCTGTAC
2041 GGCATGGACG AGCTGTACAA GAAGCTTAGC CATGGCTTCC
2081 CGCCGGAGGT GGAGGAGCAG GATGATGGCA CGCTGCCCAT
2121 GTCTTGTGCC CAGGAGAGCG GGATGGACCG TCACCCTGCA
2161 GCCTGTGCTT CTGCTAGGAT CAATGTGGGA TCCGGAGCCA
2201 CGAACTTCTC TCTGTTAAAG CAAGCAGGAG ACGTGGAAGA
2241 AAACCCCGGT CCTgactggg tgaccaaact cctgccagcc
2281 ctgctgctgc agcatgtcct cctgcatctc ctcctgctcc
2321 ccatcgccat cccctatgca gagggacaaa ggaaaagaag
2361 aaatacaatt catgaattca aaaaatcagc aaagactacc
2401 ctaatcaaaa tagatccagc actgaagata aaaaccaaaa
2441 aagtgaatac tgcagaccaa tgtgctaata gatgtactag
2481 gaataaagga cttccattca cttgcaaggc ttttgttttt
2521 gataaagcaa gaaaacaatg cctctggttc cccttcaata
2561 gcatgtcaag tggagtgaaa aaagaatttg gccatgaatt
2601 tgacctctat gaaaacaaag actacattag aaactgcatc
2641 attggtaaag gacgcagcta caagggaaca gtatctatca

2681 ctaagagtgg catcaaatgt cagccctgga gttccatgat
2721 accacacgaa cacagctttt tgccttcgag ctatcggggt
2761 aaagacctac aggaaaacta ctgtcgaaat cctcgagggg
2801 aagaaggggg accctggtgt ttcacaagca atccagaggt
2841 acgctacgaa gtctgtgaca ttcctcagtg ttcagaagtt
2881 gaatgcatga cctgcaatgg ggagagttat cgaggtctca
2921 tggatcatac agaatcaggc aagatttgtc agcgctggga
2961 tcatcagaca ccacaccggc acaaattctt gcctgaaaga
3001 tatcccgaca agggctttga tgataattat tgccgcaatc
3041 ccgatggcca gccgaggcca tggtgctata ctcttgaccc
3081 tcacacccgc tgggagtact gtgcaattaa aacatgcgct
3121 gacaatacta tgaatgacac tgatgttcct ttggaaacaa
3161 ctgaatgcat ccaaggtcaa ggagaaggct acaggggcac
3201 tgtcaatacc atttggaatg gaattccatg tcagcgttgg
3241 gattctcagt atcctcacga gcatgacatg actcctgaaa
3281 atttcaagtg caaggaccta cgagaaaatt actgccgaaa
3321 tccagatggg tctgaatcac cctggtgttt taccactgat
3361 ccaaacatcc gagttggcta ctgctcccaa attccaaact
3401 gtgatatgtc acatggacaa gattgttatc gtgggaatgg
3441 caaaaattat atgggcaact tatcccaaac aagatctgga
3481 ctaacatgtt caatgtggga caagaacatg gaagacttac
3521 atcgtcatat cttctgggaa ccagatgcaa gtaagctgaa
3561 tgagaattac tgccgaaatc cagatgatga tgctcatgga
3601 ccctggtgct acacgggaaa tccactcatt ccttgggatt
3641 attgccctat ttctcgttgt gaaggtgata ccacacctac
3681 aatagtcaat ttagaccatc ccgtaatatc ttgtgccaaa
3721 acgaaacaat tgcgagttgt aaatgggatt ccaacacgaa
3761 caaacatagg atggatggtt agtttgagat acagaaataa
3801 acatatctgc ggaggatcat tgataaagga gagttgggtt
3841 cttactgcac gacagtgttt cccttctcga gacttgaaag
3881 attatgaagc ttggcttgga attcatgatg tccacggaag
3921 aggagatgag aaatgcaaac aggttctcaa tgtttcccag
3961 ctggtatatg gccctgaagg atcagatctg gttttaatga
4001 agcttgccag gcctgctgtc ctggatgatt ttgttagtac
4041 gattgattta cctaattatg gatgcacaat tcctgaaaag
4081 accagttgca gtgtttatgg ctggggctac actggattga
4121 tcaactatga tggcctatta cgagtggcac atctctatat
4161 aatgggaaat gagaaatgca gccagcatca tcgagggaag
4201 gtgactctga atgagtctga aatatgtgct ggggctgaaa
4241 agattggatc aggaccatgt gaggggatt atggtggccc
4281 acttgtttgt gagcaacata aaatgagaat ggttcttggt

4321 gtcattgttc ctggtcgtgg atgtgccatt ccaaatcgtc
4361 ctggtatttt tgtccgagta gcatattatg caaaatggat
4401 acacaaaatt attttaacat ataaggtacc acagtcatag
4441 CCCGGGATCC ACCGGATCTA GATAACTGAT CATAATCAGC
4481 CATACCACAT TTGTAGAGGT TTTACTTGCT TTAAAAAACC
4521 TCCCACACCT CCCCCTGAAC CTGAAACATA AAATGAATGC
4561 AATTGTTGTT GTTAACTTGT TTATTGCAGC TTATAATGGT
4601 TACAAATAAA GCAATAGCAT CACAAATTTC ACAAATAAAG
4641 CATTTTTTTC ACTGCATTCT AGTTGTGGTT TGTCCAAACT
4681 CATCAATGTA TCTTAAGGCG TAAATTGTAA GCGTTAATAT
4721 TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA
4761 TTTTTTAACC AATAGGCCGA AATCGGCAAA ATCCCTTATA
4801 AATCAAAAGA ATAGACCGAG ATAGGGTTGA GTGTTGTTCC
4841 AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC
4881 AACGTCAAAG GGCGAAAAAC CGTCTATCAG GGCGATGGCC
4921 CACTACGTGA ACCATCACCC TAATCAAGTT TTTTGGGGTC
4961 GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC
5001 CCCCGATTTA GAGCTTGACG GGGAAAGCCG GCGAACGTGG
5041 CGAGAAAGGA AGGGAAGAAA GCGAAAGGAG CGGGCGCTAG
5081 GGCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC
5121 ACACCCGCCG CGCTTAATGC GCCGCTACAG GGCGCGTCAG
5161 GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG
5201 TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG
5241 AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
5281 GGAAGAGTCC TGAGGCGGAA AGAACCAGCT GTGGAATGTG
5321 TGTCAGTTAG GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG
5361 GCAGAAGTAT GCAAAGCATG CATCTCAATT AGTCAGCAAC
5401 CAGGTGTGGA AAGTCCCCAG GCTCCCCAGC AGGCAGAAGT
5441 ATGCAAAGCA TGCATCTCAA TTAGTCAGCA ACCATAGTCC
5481 CGCCCCTAAC TCCGCCCATC CCGCCCCTAA CTCCGCCCAG
5521 TTCCGCCCAT TCTCCGCCCC ATGGCTGACT AATTTTTTTT
5561 ATTTATGCAG AGGCCGAGGC CGCCTCGGCC TCTGAGCTAT
5601 TCCAGAAGTA GTGAGGAGGC TTTTTTGGAG GCCTAGGCTT
5641 TTGCAAAGAT CGATCAAGAG ACAGGATGAG GATCGTTTCG
5681 CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC
5721 GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC
5761 AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC
5801 AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG
5841 TCCGGTGCCC TGAATGAACT GCAAGACGAG GCAGCGCGGC
5881 TATCGTGGCT GGCCACGACG GGCGTTCCTT GCGCAGCTGT
5921 GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA

5961 TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCACC
6001 TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT
6041 GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC
6081 GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC
6121 GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA
6161 AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG
6201 CTCAAGGCGA GCATGCCCGA CGGCGAGGAT CTCGTCGTGA
6241 CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA
6281 TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT
6321 GTGGCGGACC GCTATCAGGA CATAGCGTTG GCTACCCGTG
6361 ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT
6401 CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC
6441 ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG
6481 GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC
6521 CTGCCATCAC GAGATTTCGA TTCCACCGCC GCCTTCTATG
6561 AAAGGTTGGG CTTCGGAATC GTTTTCCGGG ACGCCGGCTG
6601 GATGATCCTC CAGCGCGGGG ATCTCATGCT GGAGTTCTTC
6641 GCCCACCCTA GGGGGAGGCT AACTGAAACA CGGAAGGAGA
6681 CAATACCGGA AGGAACCCGC GCTATGACGG CAATAAAAAG
6721 ACAGAATAAA ACGCACGGTG TTGGGTCGTT TGTTCATAAA
6761 CGCGGGGTTC GGTCCCAGGG CTGGCACTCT GTCGATACCC
6801 CACCGAGACC CCATTGGGGC CAATACGCCC GCGTTTCTTC
6841 CTTTTCCCCA CCCCACCCCC CAAGTTCGGG TGAAGGCCCA
6881 GGGCTCGCAG CCAACGTCGG GGCGGCAGGC CCTGCCATAG
6921 CCTCAGGTTA CTCATATATA CTTTAGATTG ATTTAAAACT
6961 TCATTTTTAA TTTAAAAGGA TCTAGGTGAA GATCCTTTTT
7001 GATAATCTCA TGACCAAAAT CCCTTAACGT GAGTTTTCGT
7041 TCCACTGAGC GTCAGACCCC GTAGAAAAGA TCAAAGGATC
7081 TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG
7121 CAAACAAAAA AACCACCGCT ACCAGCGGTG GTTTGTTTGC
7161 CGGATCAAGA GCTACCAACT CTTTTTCCGA AGGTAACTGG
7201 CTTCAGCAGA GCGCAGATAC CAAATACTGT TCTTCTAGTG
7241 TAGCCGTAGT TAGGCCACCA CTTCAAGAAC TCTGTAGCAC
7281 CGCCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC
7321 TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC
7361 TCAAGACGAT AGTTACCGGA TAAGGCGCAG CGGTCGGGCT
7401 GAACGGGGGG TTCGTGCACA CAGCCCAGCT TGGAGCGAAC
7441 GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA
7481 GAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT
7521 ATCCGGTAAG CGGCAGGGTC GGAACAGGAG AGCGCACGAG
7561 GGAGCTTCCA GGGGGAAACG CCTGGTATCT TTATAGTCCT

```
7601 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTTGT
7641 GATGCTCGTC AGGGGGGCGG AGCCTATGGA AAAACGCCAG
7681 CAACGCGGCC TTTTTACGGT TCCTGGCCTT TTGCTGGCCT
7721 TTTGCTCACA TGTTCTTTCC TGCGTTATCC CCTGATTCTG
7761 TGGATAACCG TATTACCGCC ATGCATg
```

8.2.2 Cell Line Generation

I co-transfected the $2e5$ACS1 cells with the PF plasmid and a Tet-Repressor plasmid (**Tet-OFF purchased from Invitrogen**) in a 1:3 ratio to a total of 5μg DNA and 20μl Lipofectamine 2000® according to the manufacturers instructions. Antibiotic resistance of the PF plasmid where the same as the ACS1 cell line, nonetheless G418 was used in the medium (3mg/ml). For selection of TetR integration, cells were treated with Blasticidine (1mg/ml).

Previous experience gained whilst generating the ACS1 line, led to the three-step protocol detailed in **figure 8.1a**. Briefly, initially, ACS1 cells are co-transfected with the TetR and PF plasmids and treated with and without tetracycline. Although cells expressing GFP and tdTomato were observed under both conditions, I could detect a consistent sub-population of cells which were brighter in the presence of tetracycline (ranging between 1-5% of the total cell population). These cells are FACS sorted into 5ml tubes (**Sort 1**) and kept in culture under Blasticidine and Gentamycin selection for 8-10 days. I next performed a second round of tetracycline stimulation and FACS sorted cells (**Sort 2**) into individual 96-wells with the highest stringency available. These cells were left to divide for up to two weeks, after which a final tetracycline stimulation was applied to select for clones which expressed tdTomato and GFP.

8.2.3 3D Cell culture

MDCK cells were purchased from the American Type Culture Collection (ATCC) kept in minimum essential medium (MEM) supplemented with 10mg/ml L-Glutamine, and 10% Fetal Bovine Serum.

To prepare cysts in Collagen I for imaging, I used the same protocol as in II. To grow "cables", the substrate was a composition of 160μl of High Concentrated Matrigel from Cornig® with 40μl of Collagen I (3mg/ml) from Advanced BioMatrix® . This substrate was first deposited on a glass bottom MatTek 6-well dish with a 14mm diameter, and spread evenly. Then up to two purpose fire bent borosilicate 1μl HPLC glass capillaries per well were placed on top. The plates were then incubated at 37C without CO_2 for 30-60 minutes, or until the substrate had solidified. Glass capillaries were then carefully removed, and 5000 cells were pipetted under the microscope and aseptic conditions onto the through left by the

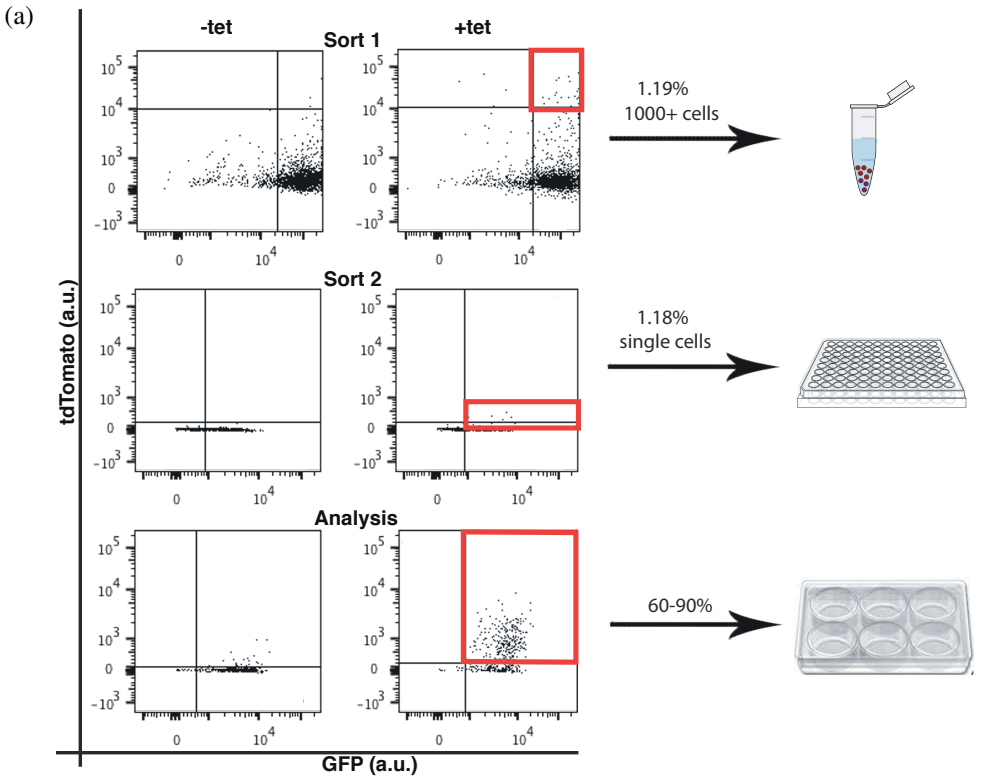


Figure 8.1 – The General Protocol for generation of Positive feedback cell lines is shown.

capillary. The cells were left in the incubator overnight to form a thin eptithelium, after which a fresh layer of 100μl matrigel/collagen was delivered and solidified as before. “Cable” like structures formed within 24 hours of this step.

8.2.4 Image Acquisition

Time-Lapse Microscopy

Due to logistical move of the laboratories (CRG in Barcelona, and Imperial College London)during my PhD, two different microscopes were used for the acquisition of the microscopy timelapses. These were:

	CRG	Imperial
Lightsource	Xenon Lamp	Mercury Lamp
Camera Module	Hammamatsu EM-CCD	Hammamatsu Orca Flash 4 CMOS
Objectives	Zeiss 20x PH 0.4NA	Zeiss 10x PH 0.3NA
Acquisition Time	3s	200ms
Filter Cube GFP	38HE	38HE
Filter Cube tdTomato	NA	46HE
Software	Axiosvision	ZEN and Volocity

For time-lapse acquisitions the cells were kept under a 37C and 5% CO_2 atmosphere. Experiments were setup using the proprietary Zeiss Zen software. To account for the collagen variability, 5 Z-stacks were taken in each of the fluorescent channels, and then the Maximum z-Projections were reconstructed in ImageJ. For the spatial images, an overlap of 5% was chosen.

Confocal Microscopy

For the 3D confocal images in section 6.3.2 the following protocol was followed.

Other Materials:
Permeabiliztion solution
0.5% Triton-X in PBS
test 0.1% vs 0.5%
Glycine rinse solution
PBS 1X
100 mM glycine
Primary Block (IF) Buffer
PBS 1X
7.7 mM NAN3
10% goat serum or horse
0.1% bovine serum albumin
0.2% Triton X-100
0.05% Tween-20
Secondary Block Buffer
IF Buffer
Materials for staining
Phalloidin: FITC, TRITC, 594
E-Cadherin: DECMA-Alexa 594 conjugated antibody Cell Signaling Technologies (Catalog number: 7687).
C-Met: N17 (Santa Cruz)

Nuclei: To-PRO3,DAPI
Secondary Ab anti-goat: Alexa-546
Method:

1. Wash with PBS and fix with 4% PFA for 20min at RT. Wash 3xPBS and proceed to staining. 10% goat or horse serum.

2. Permeabilization: If formalin or paraformaldehyde is used for fixation, permeabilize with PBS containing 0.5% Triton X-100 for 10 min at 4C. Depending on the antibody used for immunostaining, the detergent concentration or duration of permeabilization may require modification.

3. Glycine rinse: Rinse three times with PBS/glycine 10–15 min per wash at room temperature.

4. Primary block: Incubate with 200 μl/well IF Buffer for 1–1.5 h at room temperature.

5. Secondary block: Aspirate the primary block and incubate with 100 μl/well secondary block for 30 min.

6. Primary antibody: Incubate with primary antibody in the primary block solution (see step 5) overnight (15–18 h) at 4C. 1:100 to 1:200 dilutions of the primary antibody.

7. Rinse three times (20 min each) with IF Buffer at room temperature with gentle rocking.

8. Secondary antibody: Incubate with fluorescent conjugated secondary antibody in IF Buffer +10% goat serum for 40–50 min at room temperature.

9. Rinse three times (20 min each) with IF Buffer at room temperature with gentle rocking.

10. To counterstain nuclei, incubate with PBS containing 5 μM TOPRO-3 (Molecular Probes ®) and/or 0.5 ng/ml (DAPI, Sigma®) for 15 min at room temperature.

11. Rinse once with PBS for 5 min at room temperature.

12. Mount with freshly prepared Prolong Antifade Reagent (Molecular Probes) and allow to dry overnight at room temperature. Once dry, slides can be stored at 4C for up to 1 week or at -20C for up to 2 months.

Imaging was done on a Leica SP5 upright confocal microscope using a 63X low NA, long working distance water objective (HCX APO L 63x 0.90 W U-V-I). The illumination sources were an Argon laser for 488nm and HeNe2 for 534nm and 633nm, all were used at 30% intensity. The acquisition was done with a HyD detector, The z-Stacks were 40μm deep, and acquired by line scanning. The images were reconstructed in ImageJ after applying a gaussian blur filter of 2.0 pixels.

8.3 Computational

8.3.1 Model

The stochastic model was coded in R. The inputs for the model are the parameters k_1, k_2, k_3; the maximum time for the simulation, the timestep at which the stochastic model will add a molecule to the system, the number of realisations and the initial concentration of activator.

```
simple_positive_feedback <- function(k_1,k_2,k_3, maxtime,
   timestep, numberofrealisations, initial_condition=0){
output_GFP =
   matrix(0,nrow=(maxtime/timestep)+1,ncol=numberofrealisations);
for (realisation in seq(1,numberofrealisations,by=1)){
#print(realisation)
GFP= initial_condition;
a = rep(0,3);
a[1] = k_1;
a[2] = k_2*GFP;
a[3] = k_3*GFP;
asum = sum(a);
time=0.0;
count = 1;
for (ts in seq(0,maxtime,by=timestep)) {
while (time < ts){
tau = log(1/runif(1))/asum;
time = time + tau;
#DETERMINE WHICH REACTION WILL OCCUR
psm = 0;
drxn = asum*runif(1);
j = 0;
while (psm < drxn){
j = j + 1;
psm = psm+ a[j];
}
#UPDATE SYSTEM CONFIGURATION AND FURTHER PARAMETERS
if (j == 1){
GFP = GFP + 1;
a[2] = a[2] + k_2;
a[3] = a[3] + k_3;
asum = asum + k_2 + k_3;}
else if (j == 2){
GFP = GFP - 1;
```

```
#HGF diffusion
a[2] = a[2] - k_2;
a[3] = a[3] - k_3;
asum = asum - k_2 - k_3;
}
#positive feedback
else if (j == 3){
GFP = GFP + 1;
a[2] = a[2] + k_2;
a[3] = a[3] + k_3;
asum = asum + k_2 + k_3;
}
} # End of integration loop
output_GFP[count,realisation] = GFP;
count = count + 1;
}
}
return(output_GFP)
}
```

8.3.2 Image Analysis

Cysts

The code for the analysis of cysts in time can detect and track single cysts from TIF files.

```
clear all
direc='E151215B_GFP'; %specify directory where experiment
   can be found
ini=2; % specify time in hours, post initial stimulus
   stimulus
interv=1; %interval in hour*e-100
tamx=1024;
tamy=1024;
binv=4;
bin=0:tamx/binv:tamx;
se90=strel('line',5,90);
se0=strel('line',5,0);
se90b=strel('line',10,90);
se0b=strel('line',10,0);
xbin=0:800;
xbin8=xbin/800;
```

```
%% read files
cd(direc);
filist=dir('*GFP*.tif');
filist2=dir('*TOM*.tif');
lentot=length(filist);
scenelist=(dir('*GFP*t01m01_ORG.tif'));
lenscene= length(scenelist);
mosaiclist=(dir('*GFP*s01t01m*ORG.tif'));
lenmos=length(mosaiclist);
timelist=(dir('*GFP*s01t*m01_ORG.tif'));
lentime=length(timelist);
times=lentime;
%% first part of the code segments images
for iiii = 0:lenscene-1

   for iii=0:lenmos-1
      clock

      BW=zeros(tamx,tamy);%%GFP
      M=zeros(times,5);
      background_GFP=zeros(times,1);
      background_TOM=zeros(times,1);
      cyst=zeros(times,1);
      M(:,1)=ini:interv:ini+((times-1)*interv);
      dali=zeros(times,tamx,tamy);
      dalitom=zeros(times,tamx,tamy);
      ali=zeros(times,tamx,tamy);
      clock


      j=1;
      for jn=
         (lentime*lenmos*iiii+1+iii):lenmos:(lentime*lenmos*(iiii+1))
         strfil=filist(jn).name %looking for file with
            name depedent on jn
         j
         jn
         strfilsplt=strsplit(strfil,'_');
         filsplt1=char(strfilsplt(1));
         filsplt2=char(strfilsplt(2));
         filsplt3=char(strfilsplt(3));
         filsplt4=char(strfilsplt(4));
```

```
%% % % %GFP segmentation and data extraction

image=imread(filist(jn).name);
yy=reshape(image,tamx*tamy,1);
peaks=histc(yy,xbin);
pks=max(peaks);
locs=find(peaks==pks);
shiftb=xbin(locs(1));
background_GFP(j)=shiftb;
dali(j,:,:)=image;
image=wiener2(image);
image =
   adapthisteq(mat2gray(image),'Range','Original');
yy=reshape(image,tamx*tamy,1);
peaks=histc(yy,xbin8);
pks=max(peaks);
locs=find(peaks==pks);
if locs(1)==800
   shift=0.36;
else
   shift=xbin8(locs(1))+0.09+0.03*j/times;
end
BW=im2bw(image,shift);

BW=im2bw(BW,0.9);

BW=bwareaopen(BW,200);
BW=imdilate(BW,[se90,se0]);

BW=bwareaopen(1-BW,2000);

BW=imerode(1-BW,[se90,se0]);
BW=bwareaopen(BW,2000);

BW=bwareaopen(BW,20);
BW=imerode(BW,[se90,se0]);
BW=imclearborder(BW);
[labi,num]=bwlabel(BW);
res=regionprops(labi,'Area','PixelList');
%

%% TOMATO segmentation and data extraction only
   if Tom channel present
```

```
  while
     exist(strcat(filsplt1,'_','TOM','_',filsplt3,'_',filsplt4),'file
   strfilTOM=strcat(filsplt1,'_','TOM','_',filsplt3,'_',filsplt4);
   image=imread(strfilTOM);
   yy=reshape(image,tamx*tamy,1);
   peaks=histc(yy,xbin);
   pks=max(peaks);
   locs=find(peaks==pks);
   shiftb=xbin(locs(1));
   background_TOM(j)=shiftb;
   dalitom(j,:,:)=image;
   break
  end
shiftb_GFP=3000;
shiftb_TOM=3000;
for jk=1:num
   marea=res(jk).Area;
   if marea>1000
      vec=res(jk).PixelList;
      long=size(vec,1);
      veco_GFP=zeros(long,1);
      veco_TOM=zeros(long,1);
      for it=1:long
         veco_GFP(it)=dali(j,vec(it,2),vec(it,1));
         veco_TOM(it)=dalitom(j,vec(it,2),vec(it,1));
      end
      q_GFP=quantile(veco_GFP,.05)
      q_TOM=quantile(veco_TOM,.05)

      if q_GFP < shiftb_GFP
         shiftb_GFP=q_GFP;
      end
      if q_TOM < shiftb_TOM
         shiftb_TOM=q_TOM;
      end
   end
end
background_GFP(j)=shiftb_GFP;
background_TOM(j)=shiftb_TOM;
ali(j,:,:)=BW; %save cyst binary mask cysts=1
   (cyst location)
j=j+1
```

```
end
%second part of the script data extraction
[lab,num]=bwlabeln(ali);
res=regionprops(lab,'Area');
h=1;
ll=zeros(num,1); %cyst label
for j=1:num
   if res(j).Area > 3000 %label cysts only if volume
      in time is higher than
      ll(h)=j;
      h=h+1;
   end
end
h=h-1;
media_GFP=zeros(h,times);
media_TOM=zeros(h,times);
sumtot_GFP=zeros(h,times);
sumtot_TOM=zeros(h,times);
sup=zeros(h,times);
correc=zeros(h,times);
correc_GFP=zeros(h,times);
correc_TOM=zeros(h,times);
maxarea=zeros(h,1);
partix=zeros(h,times);
partiy=zeros(h,times);
centx=zeros(h,times);
centy=zeros(h,times);
for j=1:times
   [labxy,xy]=bwlabeln(ali(j,:,:));
   resc=regionprops(labxy,'Centroid');

   AA(:,:)=lab(j,:,:);
   for ii=1:h
      %% look for real cysts in image
      [r, c]=find(AA==ll(ii));
      sla=size(r);
      average_GFP=0;
      average_TOM=0;
      area=0;
      if sla(1)~=0
         for k=1:sla(1)
            average_GFP=average_GFP+dali(j,r(k),c(k));
            average_TOM=average_TOM+dalitom(j,r(k),c(k));
```

```
            area=area+1;
        end
        % substract background
        average_GFP=average_GFP-area*background_GFP(j);
        average_TOM=average_TOM-area*background_TOM(j);
        if area>maxarea(ii)
            maxarea(ii)=area;
        end

        media_GFP(ii,j)=average_GFP/area;
        media_TOM(ii,j)=average_TOM/area;
        sumtot_GFP(ii,j)=average_GFP;
        sumtot_TOM(ii,j)=average_TOM;
        correc_GFP(ii,j)=background_GFP(j);
        correc_TOM(ii,j)=background_TOM(j);
        sup(ii,j)=area;
    else
        media_GFP(ii,j)=0;
        media_TOM(ii,j)=0;
        sumtot_GFP(ii,j)=0;
        sumtot_TOM(ii,j)=0;
        correc_GFP(ii,j)=background_GFP(j);
        correc_TOM(ii,j)=background_TOM(j);
        sup(ii,j)=0;
    end

  end

end
%% save the extracted parameters per good cyst
for ii=1:h
  M(:,2)=media_GFP(ii,:); % background substracted
     and area corrected GFP
  M(:,3)=sumtot_GFP(ii,:); % sum of GFP
  M(:,4)=correc_GFP(ii,:); % measured GFP
     background value
  M(:,5)=media_TOM(ii,:); % background substracted
     and area corrected TOM
  M(:,6)=sumtot_TOM(ii,:); % sum of TOM
  M(:,7)=correc_TOM(ii,:); % measured TOM
     background value
  M(:,8)=sup(ii,:,1); % area of cyst
  cleo=find(sup(ii,:)==0);
```

```
        ss=size(cleo);
        % save cysts which are not lost in time
        if ss(2)<4
filename_of_responsive_cyst=strcat(strcat('11.final.time._scene_',int2s
           dlmwrite(filename_of_responsive_cyst, M);
        end
     end
  end
end
cd ..

cd ..
```

Cables

The original matlab code to analyze the cables described in **Section 6.3.2**, has a GUI interface. For simplicity I have included only the key sections of the code here.

```
% Image is previously saved in obj
% obj.sources_mask contains localization of source Whatman
   paper
% obj.object_mask contains localization of cable
% obj.imgdata contains multidimensional time-lapse image
   data

function [X,Y,sd,bin_thresh,t,a]= calculate_time_stuff(obj)
 c = 1;
 [sizeX,sizeY,~,sizeC,nT] = size(obj.imgdata);
 if c > sizeC, return, end;
 bin_num=1; end;
 time_num=1; end;

 %%Background substraction
 ylen=length(obj.imgdata(:,1,1,1,1));
 xlen=length(obj.imgdata(1,:,1,1,1));
 tlen=length(obj.imgdata(1,1,1,1,:));
 backmat=nan(ylen,xlen,tlen);
 sturmat=nan(ylen,xlen,tlen);
 sturback=nan(ylen,xlen,tlen);
 sturbacki=nan(ylen,xlen,tlen);
 sturcont=[];
```

```
for jj = 1:tlen
 backmat(:,:,jj)=
    obj.imgdata(:,:,1,c,jj).*~obj.object_mask;
 sturmat(:,:,jj)=
    obj.imgdata(:,:,1,c,jj).*obj.object_mask;

 parfor ll = 1: xlen
   sturback(:,ll,jj)=sturmat(:,ll,jj)-quantile(backmat(:,ll,jj),.05);
 end

 sturbacki(:,:,jj)=sturback(:,:,jj);%./sturcont(jj);
end
sturbacki=bsxfun(@times,sturbacki,obj.object_mask);
ucen=zeros(sizeX,sizeY);
timesb = (1:time_num:nT);
timesb = length(timesb);
uc=regionprops(obj.sources_mask,'Centroid');
ucx=round(uc.Centroid(2));
ucy=round(uc.Centroid(1));

ucen(ucx,ucy)=1;
u = bwdist( ucen,'cityblock' ); %obj.sources_mask;
u(~obj.object_mask)=0;
uthresh=(min(min(u)):bin_num:max(max(u)));
bins=(1:length(uthresh));
ubin=imquantize(u,uthresh);%alternatively maybe user can
   input what to quantize, as multithreshold is limited
   to 20 levels
mean_activity= zeros(timesb,length(uthresh));%bin_num
sd_activity=zeros(timesb,length(uthresh));%bin_num
area=zeros(timesb,length(uthresh));
time_count=zeros(timesb,1);
max_act=zeros(timesb,length(uthresh));
median_activity=zeros(timesb,length(uthresh));

for k = 1 :time_num :nT
 plane=squeeze(sturbacki(:,:,k));
     for j = 2:length(uthresh);%bin_num
   [x, y] = find(ubin == j);
   mean_act=zeros(length(x),1);
   for n = 1:length(x);
     mean_act(n)=(plane(x(n),y(n)));
   end
```

```
    mean_activity(k,j)=(mean(mean_act));
    max_act(k,j)=max(max(mean_act));
    sd_activity(k,j)=std(mean_act);
    area(k,j)=length(x);
  end
  time_count(k)=k;
 end
 X = bins;
 Y = mean_activity(1:time_num:nT,:);
 sd = sd_activity(1:time_num:nT,:);
 bin_thresh=transpose(uthresh);
 t=time_count(1:time_num:nT);
 a=area(1,:);
end
```

Part IV

Global Conclusions

The projects presented in this thesis are interlinked and construct upon each other. The common denominator is to synthesize in order to understand sender-receiver systems (S-R) in cell biology. It is important to make the distinction between studies of cell-cell communication and an S-R system. Whereas the former deals with the mechanistic view of the transmission of information, in an S-R system the observable is the macroscopic dynamic behavior emerging from the underlying mechanism. There is not (yet) a scientific field dealing solely with the studies of S-R systems. I would argue that, such a field is necessary and has the potential as a tool or approach to understand the "big-picture" of biological phenomena. Given enough data, the underlying rules and principles of S-R systems, could be derived from the mathematical theory of communication.

In the first project presented in Part II, I have documented how the promoter sequence of the Matrix Metalloproteinase I in the Madine Darby Canine Kidney (MDCK) cell line, can serve as a transcriptional reporter to a stimulus of Hepatocyte Growth Factor (HGF). The transcriptional reporter utilizes the intrinsic sensing machinery of the MDCK cells.

It is apparent to me, and hopefully to the reader that the generation of a true positive feedback mammalian cell line in which an initial input is both the input and the output is both a daunting and interesting proposition. Autocatalytic reactions are a natural force to be reckoned with, although their control in biological systems, requires further investigation. In this work, I started with the assumption that introducing an autocatalytic Hepatocyte Growth Factor (HGF) module into an HGF responding cell, would lead to uncontrollable feedback, and to cell death. I designed experiments with the prospect of taming the feedback to make it observable. In reality the contrary was the case, the feedback was weak,and barely detectable. I can only speculate as to the reasons, however I find it plausible that nature has optimized systems to inhibit positive feedbacks as much as possible.

Because feedbacks are immensely powerful. As I write this, the atmospheric CO_2 levels have for the first time in our species' history surpassed the 400ppm mark. Despite governmental efforts to curtail these emissions, last years' estimated emissions of CO_2 levels is about 35.7 Gt (as a comparison in 1997 the emissions were "just" 25 Gt)[Jackson et al., 2015]. The consensus among the scientific community, is that the increase in CO_2 levels is due to the burning of fossil fuels, and is a major contributor to the warming of the planet [Cook et al., 2013]. There is also a consensus, that beyond a certain atmospheric concentration of greenhouse gases, some of the positive feedbacks mentioned in **Section 3.2** will increase to a strength which is beyond our control. What is not clear is the temporal dimension of the upcoming changes. As a scientific community we can and should adapt to the urgency of the matter and address this question together.

Bibliography

[Kau, 1994] (1994). *The atlas of mouse development, by M.H. Kaufman, Academic Press, San Diego, CA, 1992, 512 pp, 80*, volume 37. Wiley Subscription Services, Inc., A Wiley Company.

[Aerni et al., 2015] Aerni, H. R., Shifman, M. A., Rogulina, S., O'Donoghue, P., and Rinehart, J. (2015). Revealing the amino acid composition of proteins within an expanded genetic code. *Nucleic Acids Res*, 43(2):e8.

[Ajo-Franklin et al., 2007] Ajo-Franklin, C. M., Drubin, D. A., Eskin, J. A., Gee, E. P., Landgraf, D., Phillips, I., and Silver, P. A. (2007). Rational design of memory in eukaryotic cells. *Genes & development*, 21(18):2271–2276.

[Arrhenius and Holden, 1897] Arrhenius, S. and Holden, E. S. (1897). On the influence of carbonic acid in the air upon the temperature of the earth. *Publications of the Astronomical Society of the Pacific*, 9(54):14–24.

[Avery et al., 1944] Avery, O. T., Macleod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J Exp Med*, 79(2):137–58.

[Azua-Bustos and Vega-Martínez, 2013] Azua-Bustos, A. and Vega-Martínez, C. (2013). The potential for detecting 'life as we don't know it' by fractal complexity analysis. *International Journal of Astrobiology*, 12:314–320.

[Bagnoli et al., 2005] Bagnoli, F., Buti, L., Tompkins, L., Covacci, A., and Amieva, M. R. (2005). Helicobacter pylori caga induces a transition from polarized to invasive phenotypes in mdck cells. *Proc Natl Acad Sci U S A*, 102(45):16339–44.

[Barcena Menendez et al., 2015] Barcena Menendez, D., Senthivel, V. R., and Isalan, M. (2015). Sender-receiver systems and applying information theory for quantitative synthetic biology. *Curr Opin Biotechnol*, 31:101–7.

[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. *The EMBO journal*, 20(10):2528–2535.

[Bialek et al., 2000] Bialek, W., Nemenman, I., and Tishby, N. (2000). Predictability, complexity and learning. *ArXiv Physics e-prints*.

[Bialek et al., 2001] Bialek, W., Nemenman, I., and Tishby, N. (2001). Complexity through nonextensivity. *Physica A: Statistical Mechanics and its Applications*, 302(1,§?4):89 – 99. Proc. Int. Workshop on Frontiers in the Physics of Complex Systems.

[Carvalho et al., 2014] Carvalho, A., Menendez, D. B., Senthivel, V. R., Zimmermann, T., Diambra, L., and Isalan, M. (2014). Genetically encoded sender-receiver system in 3d mammalian cell culture. *ACS Synth Biol*, 3(5):264–72.

[Chacon-Heszele et al., 2014] Chacon-Heszele, M. F., Zuo, X., Hellman, N. E., McKenna, S., Choi, S. Y., Huang, L., Tobias, J. W., Park, K. M., and Lipschutz, J. H. (2014). Novel mapk-dependent and -independent tubulogenes identified via microarray analysis of 3d-cultured madin-darby canine kidney cells. *Am J Physiol Renal Physiol*, 306(9):F1047–58.

[Chang and Ferrell, 2013] Chang, J. B. and Ferrell, Jr, J. E. (2013). Mitotic trigger waves and the spatial coordination of the xenopus cell cycle. *Nature*, 500(7464):603–7.

[Chin, 2014] Chin, J. W. (2014). Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem*, 83:379–408.

[Ciais et al., 2013] Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quere, C., Myneni, R., Piao, S., and Thornton, P. (2013). *Carbon and Other Biogeochemical Cycles*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[Clark et al., 2016] Clark, P. U., Shakun, J. D., Marcott, S. A., Mix, A. C., Eby, M., Kulp, S., Levermann, A., Milne, G. A., Pfister, P. L., Santer, B. D., Schrag, D. P., Solomon, S., Stocker, T. F., Strauss, B. H., Weaver, A. J., Winkelmann, R., Archer, D., Bard, E., Goldner, A., Lambeck, K., Pierrehumbert, R. T., and Plattner, G.-K. (2016). Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nature Clim. Change*, advance online publication:–.

[Cook et al., 2013] Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., and Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2):024024.

[Crick et al., 1961] Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, 192:1227–32.

[Daniel et al., 2013] Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013). Synthetic analog computation in living cells. *Nature*, 497(7451):619–23.

[de Ronde et al., 2010] de Ronde, W. H., Tostevin, F., and ten Wolde, P. R. (2010). Effect of feedback on the fidelity of information transmission of time-varying signals. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(3 Pt 1):031914.

[Deli, 2009] Deli, M. A. (2009). Potential use of tight junction modulators to reversibly open membranous barriers and improve drug delivery. *Biochim Biophys Acta*, 1788(4):892–910.

[Diambra et al., 2015] Diambra, L., Senthivel, V. R., Menendez, D. B., and Isalan, M. (2015). Cooperativity to increase turing pattern space for synthetic biology. *ACS Synth Biol*, 4(2):177–86.

[Dollo, 2016] Dollo, L. (2016). Dollos law.

[Dublanche et al., 2006] Dublanche, Y., Michalodimitrakis, K., Kümmerer, N., Foglierini, M., and Serrano, L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol*, 2:41.

[Dubuis et al., 2013] Dubuis, J. O., Tkacik, G., Wieschaus, E. F., Gregor, T., and Bialek, W. (2013). Positional information, in bits. *Proc Natl Acad Sci U S A*, 110(41):16301–8.

[Duran and Reyes, 2014] Duran, R. F. and Reyes, L. G. (2014). *En la espiral de la energia. Historia de la humanidad desde el papel de la energía (pero no solo)*, volume 1. Libros en Accion.

[Elowitz et al., 2002] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186.

[Elsässer et al., 2016] Elsässer, S. J., Ernst, R. J., Walker, O. S., and Chin, J. W. (2016). Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat Methods*, 13(2):158–64.

[Esvelt et al., 2011] Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, 472(7344):499–503.

[Flemming, 1880] Flemming, W. (1880). *Beiträge zur kenntniss der zelle und ihrer lebenserscheinungen*. M. Cohen und Sohn, Bonn.

[Friedlingstein et al., 2006] Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al. (2006). Climate-carbon cycle feedback analysis: Results from the c4mip model intercomparison. *Journal of Climate*, 19(14):3337–3353.

[Friedlingstein et al., 2011] Friedlingstein, P., Dufresne, J., Cox, P., and Rayner, P. (2011). How positive is the feedback between climate change and the carbon cycle? *Tellus B*, 55(2).

[Gautier et al., 2011] Gautier, A., Deiters, A., and Chin, J. W. (2011). Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc*, 133(7):2124–7.

[Gillespie, 1977] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361.

[Hammond et al., 2016] Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., and Nolan, T. (2016). A crispr-cas9 gene drive system targeting female reproduction in the malaria mosquito vector anopheles gambiae. *Nat Biotechnol*, 34(1):78–83.

[Hansen et al., 1984] Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J. (1984). Climate sensitivity: Analysis of feedback mechanisms. *Climate processes and climate sensitivity*, pages 130–163.

[Harris et al., 2012] Harris, A. R., Peter, L., Bellis, J., Baum, B., Kabla, A. J., and Charras, G. T. (2012). Characterizing the mechanics of cultured cell monolayers. *Proceedings of the National Academy of Sciences*, 109(41):16449–16454.

[Hellman et al., 2008] Hellman, N. E., Spector, J., Robinson, J., Zuo, X., Saunier, S., Antignac, C., Tobias, J. W., and Lipschutz, J. H. (2008). Matrix metalloproteinase 13 (mmp13) and tissue inhibitor of matrix metalloproteinase 1 (timp1), regulated by the mapk pathway, are both necessary for madin-darby canine kidney tubulogenesis. *J Biol Chem*, 283(7):4272–82.

[Hurst, 1956] Hurst, H. E. (1956). Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers*, 5(5):519–543.

[Isalan, 2012] Isalan, M. (2012). Systems biology: A cell in a computer. *Nature*, 488(7409):40–1.

[Jackson et al., 2015] Jackson, R. B., Canadell, J. G., Le Quéré, C., Andrew, R. M., Korsbakken, J. I., Peters, G. P., and Nakicenovic, N. (2015). Reaching peak emissions. *Nature Climate Change*.

[Jantsch, 1975] Jantsch, E. (1975). *Design for evolution: self-organization and planning in the life of human systems*. G. Braziller, New York.

[Jantsch, 1980] Jantsch, E. (1980). *The self-organizing universe: scientific and human implications of the emerging paradigm of evolution*. Pergamon Press, Oxford, 1st ed edition.

[Johnson et al., 2015] Johnson, H. P., Miller, U. K., Salmi, M. S., and Solomon, E. A. (2015). Analysis of bubble plume distributions to evaluate methane hydrate decomposition on the continental slope. *Geochemistry, Geophysics, Geosystems*, 16(11):3825–3839.

[Kang et al., 2016a] Kang, H.-W., Lee, S. J., Ko, I. K., Kengla, C., Yoo, J. J., and Atala, A. (2016a). A 3d bioprinting system to produce human-scale tissue constructs with structural integrity. *Nat Biotech*, advance online publication:–.

[Kang et al., 2016b] Kang, X., He, W., Huang, Y., Yu, Q., Chen, Y., Gao, X., Sun, X., and Fan, Y. (2016b). Introducing precise genetic modifications into human 3pn embryos by crispr/cas-mediated genome editing. *J Assist Reprod Genet*.

[Kauffman, 2011] Kauffman, S. A. (2011). Approaches to the origin of life on earth. *Life*, 1(1):34.

[Kaufman and Navaratnam, 1981] Kaufman, M. H. and Navaratnam, V. (1981). Early differentiation of the heart in mouse embryos. *J Anat*, 133(Pt 2):235–46.

[Kimoto et al., 2013] Kimoto, M., Yamashige, R., Matsunaga, K.-i., Yokoyama, S., and Hirao, I. (2013). Generation of high-affinity dna aptamers using an expanded genetic alphabet. *Nat Biotechnol*, 31(5):453–7.

[Kwon et al., 2014] Kwon, S.-H., Liu, K. D., and Mostov, K. E. (2014). Intercellular transfer of gprc5b via exosomes drives hgf-mediated outward growth. *Curr Biol*, 24(2):199–204.

[Lahav et al., 2004] Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., and Alon, U. (2004). Dynamics of the p53-mdm2 feedback loop in individual cells. *Nat Genet*, 36(2):147–50.

[Lancaster and Knoblich, 2014] Lancaster, M. A. and Knoblich, J. A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. *Science*, 345(6194):1247125.

[Leduc, 1906] Leduc, S. (1906). *Théorie Physico-chimique de la Vie et Générations Spontanées*. Masson.

[Leduc, 1912] Leduc, S. (1912). *La biologie synthétique*. A. Poinat, Paris.

[Leduc, 1914] Leduc, S. (1914). *The Mechanism of Life*. Rebman Company.

[Levchenko and Nemenman, 2014] Levchenko, A. and Nemenman, I. (2014). Cellular noise and information transmission. *Current Opinion in Biotechnology*, 28:156 – 164. Nanobiotechnology * Systems biology.

[Li and Vitányi, 2009] Li, M. and Vitányi, P. (2009). *An introduction to Kolmogorov complexity and its applications*. Springer Science & Business Media.

[Lillo and Farmer, 2004] Lillo, F. and Farmer, J. D. (2004). The long memory of the efficient market. *Studies in nonlinear dynamics & econometrics*, 8(3).

[Lorenz, 1934] Lorenz, E. (1934). Search for mitogenetic radiation by means of the photoelectric method. *The Journal of General Physiology*, 17(6):843–862.

[Mali et al., 2013] Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). Rna-guided human genome engineering via cas9. *Science*, 339(6121):823–6.

[Maturana and Varela, 1972] Maturana, H. R. and Varela, F. J. (1972). *De máquinas y seres vivos: una teoría sobre la organización biológia*. Cormoran. El Mundo de la cienca, 11. Editorial Universitaria, Santiago de Chile, 1. ed edition.

[Minsky, 1982] Minsky, H. P. (1982). *Can "it" happen again?: essays on instability and finance*. M.E. Sharpe, Armonk, N.Y.

[Mizuno and Nakamura, 2013] Mizuno, S. and Nakamura, T. (2013). Hgf-met cascade, a key target for inhibiting cancer metastasis: the impact of nk4 discovery on cancer biology and therapeutics. *Int J Mol Sci*, 14(1):888–919.

[Mojallal et al., 2014] Mojallal, M., Zheng, Y., Hultin, S., Audebert, S., van Harn, T., Johnsson, P., Lenander, C., Fritz, N., Mieth, C., Corcoran, M., et al. (2014). Amotl2 disrupts apical–basal cell polarity and promotes tumour invasion. *Nature communications*, 5.

[Montesano et al., 1991] Montesano, R., Matsumoto, K., Nakamura, T., and Orci, L. (1991). Identification of a fibroblast-derived epithelial morphogen as hepatocyte growth factor. *Cell*, 67(5):901–8.

[Nawroth et al., 2012] Nawroth, J. C., Lee, H., Feinberg, A. W., Ripplinger, C. M., McCain, M. L., Grosberg, A., Dabiri, J. O., and Parker, K. K. (2012). A tissue-engineered jellyfish with biomimetic propulsion. *Nat Biotechnol*, 30(8):792–7.

[Nielsen et al., 2016] Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016). Genetic circuit design automation. *Science*, 352(6281):aac7341.

[Park et al., 1986] Park, M., Dean, M., Cooper, C. S., Schmidt, M., O'Brien, S. J., Blair, D. G., and Vande Woude, G. F. (1986). Mechanism of met oncogene activation. *Cell*, 45(6):895–904.

[Pascal et al., 2013] Pascal, R., Pross, A., and Sutherland, J. D. (2013). Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol*, 3(11):130156.

[Perez-Pinera et al., 2012] Perez-Pinera, P., Ousterout, D. G., Brown, M. T., and Gersbach, C. A. (2012). Gene targeting to the rosa26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res*, 40(8):3741–52.

[Peters, 1996] Peters, E. E. (1996). *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*. Wiley, New York, 2nd ed edition.

[Pinheiro et al., 2012] Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., Herdewijn, P., and Holliger, P. (2012). Synthetic genetic polymers capable of heredity and evolution. *Science*, 336(6079):341–4.

[Prigogine, 1977] Prigogine, I. (1977). Ilya prigogine - nobel lecture: Time, structure and fluctuations. Nobelprize.org Nobel Media AB 2014.

[Purcell and Lu, 2014] Purcell, O. and Lu, T. K. (2014). Synthetic analog and digital circuits for cellular computation and memory. *Curr Opin Biotechnol*, 29:146–55.

[Rackham et al., 2016] Rackham, O. J. L., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., FANTOM Consortium, Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W., Petretto, E., Forrest, A. R. R., Hayashizaki, Y., Polo, J. M., and Gough, J. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nat Genet*.

[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):566–70.

[Ritchie, 1986] Ritchie, D. (1986). Shannon and weaver unravelling the paradox of information. *Communication research*, 13(2):278–298.

[Rothwell et al., 2010] Rothwell, D. G., Crossley, R., Bridgeman, J. S., Sheard, V., Zhang, Y., Sharp, T. V., Hawkins, R. E., Gilham, D. E., and McKay, T. R. (2010). Functional expression of secreted proteins from a bicistronic retroviral cassette based on foot-and-mouth disease virus 2a can be position dependent. *Human gene therapy*, 21(11):1631–1637.

[Samuels, 2013] Samuels, L. (2013). *In Defense of Chaos: The Chaology of Politics, Economics, and Human Action*. Cobden Press.

[Santos and Nigam, 1993] Santos, O. F. and Nigam, S. K. (1993). Hgf-induced tubulogenesis and branching of epithelial cells is modulated by extracellular matrix and tgf-beta. *Dev Biol*, 160(2):293–302.

[Schikora-Tamarit et al., 2016] Schikora-Tamarit, M. A., Toscano-Ochoa, C., Domingo Espinos, J., Espinar, L., and Carey, L. B. (2016). A synthetic gene circuit for measuring autoregulatory feedback control. *Integr. Biol.*, 8:546–555.

[Shannon and Weaver, 1949] Shannon, C. E. and Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press, Urbana.

[Siciliano et al., 2011] Siciliano, V., Menolascina, F., Marucci, L., Fracassi, C., Garzilli, I., Moretti, M. N., and di Bernardo, D. (2011). Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLoS Comput Biol*, 7(6):e1002074.

[Stegmann,] Stegmann, U. *Animal communication theory: information and influence*.

[Stewart-Ornstein and Lahav, 2016] Stewart-Ornstein, J. and Lahav, G. (2016). Dynamics of cdkn1a in single cells defined by an endogenous fluorescent tagging toolkit. *Cell Rep*, 14(7):1800–11.

[Su et al., 2016] Su, S., Hu, B., Shao, J., Shen, B., Du, J., Du, Y., Zhou, J., Yu, L., Zhang, L., Chen, F., et al. (2016). Crispr-cas9 mediated efficient pd-1 disruption on human primary t cells from cancer patients. *Scientific reports*, 6.

[Summerer et al., 2006] Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W., and Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A*, 103(26):9785–9.

[Sun et al., 2014] Sun, S. B., Schultz, P. G., and Kim, C. H. (2014). Therapeutic applications of an expanded genetic code. *Chembiochem*, 15(12):1721–9.

[Tan et al., 2009] Tan, S., Tompkins, L. S., and Amieva, M. R. (2009). Helicobacter pylori usurps cell polarity to turn the cell surface into a replicative niche. *PLoS Pathog*, 5(5):e1000407.

[Tocchio et al., 2015] Tocchio, A., Martello, F., Tamplenizza, M., Rossi, E., Gerges, I., Milani, P., and Lenardi, C. (2015). Rgd-mimetic poly(amidoamine) hydrogel for the fabrication of complex cell-laden micro constructs. *Acta Biomater*, 18:144–54.

[Tom and Andrew, 1991] Tom, D. and Andrew, L. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313.

[Toporowski, 2005] Toporowski, J. (2005). *Theories of financial disturbance: an examination of critical theories of finance from Adam Smith to the present day*. Edward Elgar, Cheltenham, UK.

[Tostevin and ten Wolde, 2009] Tostevin, F. and ten Wolde, P. R. (2009). Mutual information between input and output trajectories of biochemical networks. *Phys Rev Lett*, 102(21):218101.

[Turing, 1952] Turing, A. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B*, 237:37–72.

[Uehara et al., 2000] Uehara, Y., Mori, C., Noda, T., Shiota, K., and Kitamura, N. (2000). Rescue of embryonic lethality in hepatocyte growth factor/scatter factor knockout mice. *Genesis*, 27(3):99–103.

[Varela et al., 1974] Varela, F. G., Maturana, H. R., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Curr Mod Biol*, 5(4):187–96.

[Wang et al., 2012] Wang, K., Schmied, W. H., and Chin, J. W. (2012). Reprogramming the genetic code: from triplet to quadruplet codes. *Angew Chem Int Ed Engl*, 51(10):2288–97.

[Wang et al., 2001] Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001). Expanding the genetic code of escherichia coli. *Science*, 292(5516):498–500.

[Weber et al., 1988] Weber, B. H., Depew, D. J., and Smith, J. D. (1988). *Entropy, information, and evolution: new perspectives on physical and biological evolution*. MIT Press, Cambridge, Mass.

[Weron, 2002] Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, 312(1):285–299.

[Wikipedia, 2016] Wikipedia (2016). Kolmogorov complexity — wikipedia, the free encyclopedia. [Online; accessed 19-April-2016].