Universitat Autònoma de Barcelona

Stochastic modelling of epigenetic regulation: analysis of its heterogeneity and its implications in cell plasticity

Núria Folguera Blasco

Stochastic modelling of epigenetic regulation: analysis of its heterogeneity and its implications in cell plasticity

Tesi Doctoral

Autora: Núria Folguera Blasco

Director: Tomás Alarcón Cor

Codirector: Javier Menéndez Menéndez

Universitat Autònoma de Barcelona (UAB)

Centre de Recerca Matemàtica (CRM)

Programa de Doctorat en Matemàtiques

Setembre de 2018

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Mathematics

at the

Universitat Autònoma de Barcelona

September 2018

We are more than the sum of our genes.

Klar, 1998

Abstract

Epigenetic regulation (ER) is key for understanding cell fate decisions and transitions. The inherent heterogeneity existing in ER systems exhibiting the differentiated cell epi-phenotype allows to identify ER systems which can give rise to either a plastic behaviour, where cell reprogramming is possible, or to a resilient one, where reprogramming is not feasible. The appearance of the plastic scenario has been linked with ageing. By extending the ER model, ageing associated effects can be taken into account. The extended ER model allows us to define a healthy plastic scenario. When further dysregulations of the ER system are considered, a pathological form of reprogramming is identified so that the ER system gets locked in a non-differentiated state. When coupling ER to a gene regulatory network (GRN) model, the ability of the ER systems to switch state allows us to identify differentiation-primed and pluripotency-locked systems, as well as to design strategies able to unlock those cells that remain in a stem cell like state. These strategies may be of key importance in order to avoid cancer progression.

Acknowledgements

Firstly, I would like to acknowledge my supervisor, Tomás Alarcón. I would need lots of pages to thank him for many things, but I will summarise how he has made me feel from the very beginning to the end of these 4 years. I still remember the phone call he gave me when I was studying my MSc in Bath. At that moment, he told me that he had a very nice project in mind and that he thought that I could be a good candidate for it. He could not be more right! He was absolutely right when stating that the project was nice. In fact, it was, and still is, awesome! I would be always grateful to you, Tomás, for having given me this nice project and having relied on me for carrying out this research idea. As you know, Tomás, I have loved this project a lot, I have really got passionated about it and I think this is also because of you. You have worked hard with me in order to make the most of it and I think we have done it quite well. That day, when you called me, I felt that you were valuing me a lot. And I have had this feeling during all the PhD. You have given me valuable support whenever I have needed, not just scientifically. You have provide me advice in research and future career and you have been implied in this project as much as I have been (even reading this thesis during August to make suggestions and corrections to improve it, for which I am extremely grateful). Furthermore, you have created a really nice working atmosphere where I have never felt like being somebody at a lower level. You have treated me as if I were a researcher to whom you could discuss anything. In some cases, I was not clearly at the same level as you were, but in these cases, you had the patience to explain me whatever was needed. With you, I have laughed a lot because you have this nice way of approaching research, and life in general. All this has made me feel really comfortable working with you.

I could write lots of positive things about you, but I know that I need to stop at some point. I would stop in here by saying that when I came back from the UK for the PhD you offered me, I was hesitating a lot if I had taken the right decision or not. Right now, I have no single doubt that that decision is probably one of the best I have ever taken. So, from the deep of my heart, thank you so much Tomás. As you know, and as I have told you so many times, if I could, I would stay at CRM working with you. It has been an incredible pleasure.

As I have just mentioned, if I could, I would stay at CRM. And this is partly thanks to Tomás, but not only because of him. CRM has been to me like a second home and I have felt really valued and appreciated. It is really nice going to work feeling like this. I start by thanking Joaquim Bruna, the former director, for awarding me one of the scholarships from 'La Caixa Foundation'. Consequently, I would like

to extend my gratitude to 'La Caixa Foundation' for its funding during these PhD years. Thank you for funding research in Collaborative Mathematics, something I really believe in. Regarding funding, I would also like to thank Fundació Ferran Sunyer i Balaguer for giving me a travel scholarship which allowed me to work at the Mathematical Institute in Oxford.

During my PhD period, I have met at CRM incredible people who started being my office mates and who now they have become my friends. I would like to say some words about them since I have to acknowledge them for many reasons. To Gemma, I am really grateful for always being there ready to listen, to talk or to let me use her computer to run my simulations. To Roberto, for being a really nice group mate willing to help and for staying late with me in the office, so I was not alone. To Víctor, for always making me laugh, something really needed, specially in those days where research does not work the way you want. To Daria, for her extremely nice ability for drawing awesome plots (like the 'supercell') and for her passion for helping, both in research and in life. To Claudia, for her kindness, her 'extraempathy' sense and for her ability to see when you need somebody to cheer you up. To Marc, for sharing the musical moments and the craziness. To Alberto, Carmelo, Genís and Bernat for making lunch time a really funny part of the day and the perfect disconnection moment.

Special mention goes to Marina and Carles, who have been an extension of my family. You have always been there. Marina and Carles, Carles and Marina, I have shared really special moments with you. You have listened to me, given advice and you have treated me as if I was part of your family. I have no words to describe how you have made me feel. I find it hard to find people to whom I can really trust and connect, but with you, from the very beginning, it has been very different. I don't know exactly why, but, I guess the answer is on you. I am sure that one of the best things the PhD has given me is meeting you, and I will not let go. No matter how far we are, because I know that a part of you goes with me. From you, I have learned both to be a better person and to be more critical about everything surrounding us. I admire your fairness and kindness, the way you approach life in general. You fight for your ideals regardless of what others think or where the world is leading to. You are not afraid of saying your thoughts. I strongly believe that the world needs more people like you, as this is the only way progress will be made. So many thanks for all what you have showed me. I have become a better human, which is priceless, and all because of you. Thanks for letting me enter in your lives.

At CRM, I have also met other research staff who have made the PhD period better. I thanks Juan for listening and givimg advice. I thanks Josep for trying to help with all his contacts and trying to look for collaborations. To Isa, my gratitude for always been thinking about me and trying to involve me in any project she had in mind. I really appreciate it! And, of course, to Álvaro, for his incredible sense of humour, and for his lovely e-mails, which I am already missing. Álvaro, you have been really kind with me. Many thanks! In general, I would have nice words for 'almost' all the people I have met at CRM. Hence, I would also like to thank people from administration. I thank Núria, Ana, Alba, Consol, Pau, Mari Paz, Àngels and Arantxa for treating me so well. I also thanks Jordi for his patience with all my computer problems. And last, but not least, I would like to express my gratitude to the current director, Lluís Alsedà. I met Lluís at my first year at University and, at that time, I already thought that he was a really nice person, extremely kind, and he treated me really well as a student. Now, as a director of the CRM, he has made me feel like an important piece in the centre, he has always worried about me and he has taken care about me almost as a father. I have no words to describe how important is to feel like this. So Lluís, thank you for always having the door open for me, for being in your office with a smile (independently of how difficult the situation was) and for trying to do the best you could so as to help me, no matter with whom you had to fight. As you always say, you care about people, not about money. I truly believe you do, and I can tell you this is one of the reasons why I love CRM so much. To me, it is like a big family and working within this environment is something which I really love. So I thanks to all the people who has contributed to it, and I hope that it will continue like this for many years. I am really proud of having been a member of CRM, it has been a huge pleasure.

During my PhD period, I have collaborated with several people, of whom I would like to say a few words. I thank Javier Menéndez, my cosupervisor, for his valuable biological comments and for his efforts in trying to interpret our results in biological terms. I really appreciate his dedication in reading our research and for always being enthusiastic about it. I also thanks Professor Byrne, for opening me the doors at the Mathematical Institute in Oxford. While being there, Helen, you were really nice and your input was extremely useful, always pointing to the right approach or to the key question to study. I really admire you for this, and it has been a really pleasure collaborating with you. Many thanks, Helen, for giving me this nice opportunity and for all the time you have devoted to me. Hopefully, we will keep working together in the near future!

I send also my gratitudes to Rubén Pérez, who has been an extremely hard-worker, improving the MAP method to fix any issue I encountered, always replying to my e-mails very quickly and with very detailed answers. I also thank you, Rubén, for your enormous effort to improve the paper and for your careful reading. Working with people like you is a pleasure because since you get that much involved, you make the others feel that, at least, they need to be involved as much as you did. So thanks to you, I am pretty sure that I have even worked harder, for instance, on improving some plots; you made me feel that I could not let you down, because you had already done your best, so I had to do mine too.

Another person with whom I have worked with is Toni Guillamon. Toni, I really admire your calm and patience. I am really grateful for the time you devoted for doing the xppauto tutorial so I could learn how to use it really quickly. And you succeed! I am thankful too for your willingness to work together, for reading whatever I was sending you and in general, for finding time and putting your efforts in one of my PhD questions. Thanks to you, I have learned to approach the problem in a better way than the one I had been using. I wish I had learned xppauto before! Probably, I would have been able to do more things, or at least, I would have spent less time waiting for Matlab to finish. In any case, thanks to you, Toni, I have another tool in my bag for its future use. I have really enjoyed the period of time we have worked together.

I would also like to thank people outside the CRM world. In general, I would like to thank my friends for their patience with me, for understanding every time I was saying 'no' to meet them or going somewhere, because I was working and I could not make it. For keeping me updated with everything, even when I was completely out for a longer period of time. And in general, for showing me that they care about me a lot. Also, for showing their interest in whatever I was researching and for giving me the right moments to disconnect from the scientific world. Specially, I would like to thank you, Alba, for always sending me a text for catching up on how things are going on, for asking how I am feeling and, in general, for making me feel that although living in another country you are really close to me. My thanks also go for you, Joan. Thanks for always being there to help in any shape of form (and at any hour!), to cheer me up or to trust in me even when I was not doing it. Having people like you is something anyone should have in their lives because it really helps to overcome any issue you may encounter. So many thanks, Joan, for having helped to feel that I was not alone. I will be always in debt with you for this.

To conclude, I would like to give a general thank you to anyone who has believed in me. For me, it is really important feeling valued and during these years, I have really felt like this. Before finishing, I would like to thank my family, for their support, for providing me anything I have needed and for having given me the opportunity to live with no further preocupations than my research questions. I am really grateful for all the efforts you have put in trying to make my life as easy as possible and, of course, my life would not be the same without you. " Papa i mama, mai us podré agrair el que heu lluitat per mi. Sé que segurament sense vosaltres no estaria on sóc ara. Moltes gràcies per tot. Us estimo molt".

Finally, I would like to say that I see all the people I have mentioned (and those who I have probably forgotten, to whom I apologise), as my 'epigenetics'. All of you have given me lots of inputs, mainly positive, during these years, which have had an impact on my research, on my behaviour and on my life, in general. So the way I have been these years is a consequence of all the ones who have shared them with me. As mentioned by Klar, *we are more than the sum of our genes*. And all of you are, from now on, part of these extra layer beyond my genetic code. So my most sincere gratitude to all the ones who have shared this period of my life with me. This thesis would not have been the same without your support.

Contents

Chapter 1

Introduction

1.1 Biological motivation

During embryonic development, a single cell, the fertilised egg, also known as zygote, divides to give rise to many millions of cells forming a multicellular organism. These cells go on to organise themselves into complex structures as tissues and organs. One important question is how such variety of cells can emerge from a single cell.

In particular, it is interesting to understand how the organising principles of development are encoded within the genetic material of the fertilised egg, which is key in order to understand the final form of a multicellular organism. In this sense, the distinction between genotype, the genetic information inherited by an organism from its parents and encoded into its DNA, and phenotype, the way this genetic information is expressed regarding physical appearance or internal structure, first set in 1909 by the Danish botanist Wilhelm Johannsen, shed some light on understanding that, in spite of having the same genetic information (genotype), individuals could be really different [117, 173].

In turn, this distinction helped towards the comprehension of how from a single cell, the whole variety of cell types contributing to the bodyplan of complex organisms could arise. The genotype/phenotype distinction puts the emphasis on the way genetic endowment is expressed/translated during development, since although individual cells carry the same genetic information, their final aspect is very different.

1.1.1 Cell differentiation

Cellular differentiation refers to the process by which embryonic stem cells (ESCs) become structurally and functionally different, thus acquiring distinct identities and specialised functions. In other words, cell differentiation involves the appearance of cell types that have a clear identity in an adult organism, such as muscle, nerve, blood or skin cells. In humans, the fertilised egg gives rise to at least 250 clearly distinguishable types of cells [173, 167].

Although the differentiation process has been described in the previous paragraph starting from the fertilised egg (ESCs), in fact, it can be generalised as starting with a pluripotent founder cell, such as a stem cell. Stem cells are a type of cell which are able to repeat division and whose daughter cells can either remain stem cells (i.e. they are able to self-renew) or differentiate into a variety of cell types, in analogy of what occurs with ESCs. Two types of self-renewal are possible: symmetric division, where both daughters remain stem cells, and asymmetric division, in which case, one of the daughter cells undergoes differentiation, whereas the other stays as stem cell. In addition to ESCs, tissue-specific progenitor cells exist. These cells have a limited capacity for regenerating those tissues for which they are specific. An example of tissue-specific progenitor cells are hematopoietic stem cells (HSCs).

In general, cell differentiation is a gradual process in the sense that cells go through several divisions before being terminally differentiated. From the initial progenitor cell to the final differentiated cells, which are considered stable, there are some intermediate steps, which are viewed as transient amplifying stages, corresponding to early stages of the differentiation process.

The main characteristic of cell differentiation is changes in gene expression, i.e. changes in the phenotype under an unchanging genotype. In general, a differentiated cell has a target protein (usually labelled as the 'luxury' or cell-specific protein [173]) to be produced, for instance, hemoglobin in red blood cells. Besides upregulation of the target protein, other genes that help in the production of 'housekeeping' proteins, for instance, those needed for metabolism, are also activated [173].

A metaphor of cell differentiation in development was proposed by Conrad Waddington [162] (see Fig. 1.1 for a representation), which later was known as Waddington landscape. The picture Waddington proposed depicts cell differentiation as a ball sliding down a sloped 'landscape' with some valleys separated by ridges. The valleys

correspond to different cell types and how far the ball travels in this landscape mimics the cell differentiation process; as development proceeds, the ball travels down the valley and it encounters bifurcation points, leading to new valleys, i.e. new cell fates. The branching points, which are the points of lowest local elevation, correspond to the intermediate steps described above, i.e. the transitory states before reaching the fully differentiated states, which will be at the bottom of the landscape. Since once the ball has decided to take one path, all the others are 'rejected', this implies that the cell (represented by the ball) has gone one step away from being pluripotent, and closer to its final differentiated state.

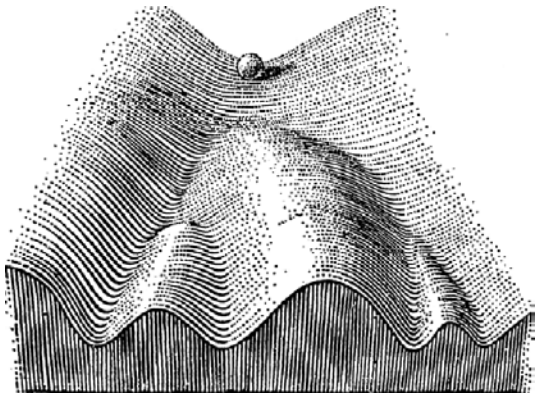


Figure 1.1: Waddington landscape represents cell differentiation as a ball going down a hill.

This landscape representation can be further used in the context of the genotype/phenotype distinction mentioned earlier. The structure of the landscape as defined by its valleys and branching points, is identical in all the cells (all of them have the same genotype). By contrast, the way the ball transits over this landscape is something not predetermined and hence, when deciding to take one path or another, the final fate is noticeably different, implying a completely different phenotype.

For decades, it was believed that cell differentiation was an irreversible process, that is, when a cell was differentiated into a terminal phenotype, the cell fate was determined and it could not be changed. For this reason, Waddington imagined his landscape with a downwards slope, as, by viewing it in this manner, the cell could only go forward in the differentiation process. This was proved wrong in 2006, when a Japanese scientist, Shinya Yamanaka, showed that cell differentiation

could be reverted, i.e. somatic, differentiated cells could be reprogrammed into the pluripotent (stem cell) state [1] [150].

1.1.2 Cell reprogramming

In 2006, Yamanaka and coworkers proved that differentiated cells could be reprogrammed into behaving like pluripotent stem cells [150, 23]. This process was called cell reprogramming. Although reprogramming may follow a completely different path to reach the pluripotent stem cell state compared to the one differentiation process followed to reach the differentiated state, the initial and final state of the differentiation process can be seen, respectively, as the final and initial points for reprogramming.

The first successful reprogramming experiments were carried out with adult mouse fibroblasts [150]. Later, reprogramming was also achieved with human cells [149]. The pluripotent cells obtained by reprogramming were coined as induced pluripotent stem cells (iPSCs). Such cells have the same morphology and growth properties as ESCs [150, 167, 115, 124]. ESCs are able to divide indefinitely while maintaining pluripotency, as well as having the ability to differentiate into cells of the three germ layers (mesoderm, endoderm and ectoderm), that is, into any cell type. This is the main reason why iPSCs are instrumental in treating many degenerative diseases, such as Parkinson's disease or a spinal chord injury.

Yamanaka proposed a mechanism for in vitro reprogramming constisting on inducing the expression of certain genes. In particular, Yamanaka's experiment was performed by inducing the expression of 4 genes, the $OSKM$ combination, which stands for *Oct4*, *Sox2*, *Klf4* and *c-Myc* genes. Since then, these 4 genes are known as Yamanaka factors. $OSKM$ were identified by systematic elimination among 24 candidate genes thought to induce pluripotency in somatic cells (any type of cell that is not germinal) [150]. They tried several combinations by removing one gene at a time so as to identify which genes were necessary to produce ESCs. They identified a set of 10 genes which was reduced to a set of 4 following the same procedure, so they could identify the $OSKM$ genes as the key ones in order to obtain ESCs. They tried to further reduce the gene set, but the results were not satisfactory and thus, these 4 genes remained as the key combination. As Takahashi and Yamanaka mention in

[1] For this remarkable discovery, which has helped to redesign the way developmental biology had been thought for centuries, Yamanaka was awarded (jointly with John B. Gurdon) the Nobel Prize for Physiology or Medicine in 2012 [124].

their paper [150], they were quite surprised that *Nanog* was dispensable in this key combination, because it is a gene highly related to pluripotent behaviour. Before Yamanaka's discovery, *Nanog* was thought to be necessary to obtain pluripotency and this could explain why other experimentalists did not succeed when trying to obtain iPSCs.

It is noteworthy that more recent experiments have shown that not all of these 4 genes are essential. There exist other possible combinations which allow to obatin iPSCs without the 4 Yamanaka factors at once [23, 64]. In fact, there is evidence that reprogramming can occur spontaneously in vivo in response to injury or damage conditions [151, 179], although such spontaneous reprogramming is extremely rare.

Something Yamanaka already observed, which has also been noted by many other experimentalists [129, 23], is that although the reprogramming process is robust, well-characterised and reproducible, it has extended latency and it is extremely inefficient, i.e. from the differentiated cells that start the process of reprogramming, all of them genetically identical, only few (about 1%) complete the process of reprogramming and become pluripotent stem cells [129, 124]. These results may suggest an inherent heterogeneity in the set of differentiated cells, whereby some cells may be more prone to reprogramming than others.

The first to investigate whether cell heterogeneity is involved in reprogramming was Yamanaka [176]. He reviewed the evidence available at the time regarding support for two models: the 'elite' model and the 'stochastic' model. The former assumes that only a small subset of predetermined cells has reprogramming potential. By contrast, the stochastic model states that reprogramming is a random event occurring on a homogeneous population. Although, at the time, Yamanaka found the evidence supporting the stochastic model more compelling, later work has provided evidence sustaining that cell populations are heterogeneous regarding their reprogramming potential [66].

Reprogramming potential has also been linked to cell cycle duration [70, 66]. It has been shown that within a genetically identical population of cells, those cells with a shorter cell cycle are more likely to reprogram [66]. Several methods, such as inhibiting the p53 pathway which controls the cell cycle progression, increase the cell division rate and, ultimately, the rate of iPSCs formation [70]. Therefore, the duration of the cell cycle seems to be one identified property that may help to explain the heterogeneity in reprogramming potential. Since this heterogeneous cell response

to reprogramming cannot have a genetic origin and the stochastic hypothesis is not convincing enough, there must be some other form of regulation which can explain this heterogeneity regarding cell ability to reprogram. This regulation is performed through epigenetics. As we will see in the following section, epigenetics may help or impede reprogramming by changing certain properties that control gene expression.

1.2 Basic facts about epigenetics

The term epigenetics, meaning above genetics, was coined by the developmental biologist Conrad Waddington in the 1940s [160, 161, 163], to describe the feedback between genes and environment which, eventually, produces the phenotype. A general, well accepted definition of epigenetics is the study of heritable changes in the cellular state, such as the gene expression profile of a cell (which determines which genes should be expressed and which ones should not), that are not caused by changes in the nucleotide sequence of the DNA [157].

Epigenetic regulation (ER) is essential to determine and maintain cell fate in multicellular organisms, where all somatic cells are genetically identical. Epigenetics provides the additional layers of information on top of the bare genomic sequence, needed to deploy the full potential of the genetic code to create a variety of cells in an otherwise genetically homogeneous organism. This is why it is said that epigenetics gives to the cell a phenotypic plasticity, since depending on the epigenetic signature, the phenotype may be one or another [3].

ER allows to modify the phenotype without modification of the genome (which occurs through mutations on evolutionary time scales, much longer than those characteristic of development). Epigenetic marks that silence or activate portions of the genome are inherited upon cell division. This so-called epigenetic memory allows for adaptations on a much shorter time scale than those associated with evolutionary processes [42, 3]. Thus, information is encoded in transient signals which set the cell in one of at least two regulatory states, which result in bistable gene expression, with the gene being expressed or not (on or off). Both of these states are stable and heritable through cell division. In particular, epigenetic cell memory is responsible for identical genomes to have different functional identities, for example, different phenotypes in cell differentiation [42]. Epigenetic modifications are thus mandatory for correct cell differentiation. Nevertheless, this is not always the case as certain epigenetic modifications are caused by other factors, such as ageing or cancer, and

they confer to the cell what is known as epigenetic plasticity [51]. Epigenetic plasticity, which will be explained later in Section 1.2.2, refers to the ability that these epigenetic changes may cause into the cell behaviour, making it more favourable or more reluctant to reprogramming, giving rise to plastic behaviour.

1.2.1 Epigenetic regulation of gene activity

When unpacked, the DNA molecule in higher eukaryotes would measure approximately two metres long and therefore it needs to be condensed in order to fit into the nucleus of cells, something the molecule does by folding into a supercoiled structure. The basic unit of DNA packageing is the nucleosome, in which, approximately 200 base pairs (bp) are wrapped around a core formed by 2 molecules each of 4 core histones (H2A, H2B, H3 and H4), an octamer, around which 147 bp of DNA are wrapped [42, 124] (see Fig.1.2 for a schematic representation). This protein-DNA polymer is known as chromatin, having in the nucleosomes its fundamental subunit. Chromatin is a massive macromolecular complex which has approximately six billion bases of DNA, wrapped about 30 million nucleosomes [51].

However, when packageing the DNA to fit into the confined space of the nucleus not all the information it encodes is physically available. Chromatin structure provides the solution to this problem by 'indexing' the information. Chromatin has not a uniform structure: it can be highly condensed (heterochromatin) or less compacted (euchromatin) [3]. These different structures, which can be identified by looking at the epigenetic marks [92], have clear implications on how genetic information may be used. When genetic information is located in a less compacted chromatin structure, chromatin is said to be active, because this genetic information will be available to initiate transcription. On the contrary, highly condensed chromatin is associated with inactive or repressive chromatin, which impedes the use of the genetic information it encodes [92].

Patterns of gene activity vary from one cell type to another. Hence, it is important to unravel the mechanisms that determine these patterns and how they are inherited by daughter cells. In other words, we need to understand why certain genes are maintained switched on and others are switched off in a particular cell type, and what is the cause for the activation/inactivation of these genes to vary if we analyse other cell types. In order to shed light on these concepts, the regulation of transcription needs to be described, since it consists the first and main step in the expression

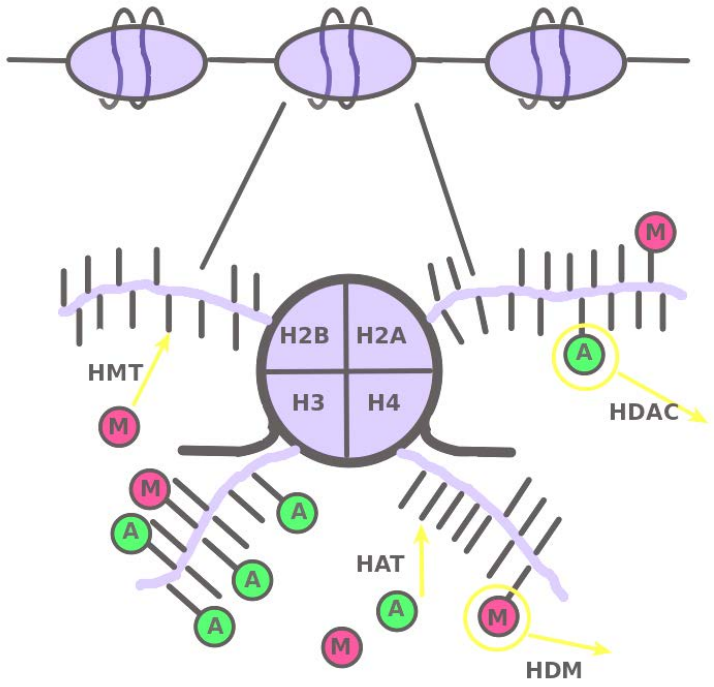


Figure 1.2: (Top) Schematic representation of the chromatin structure. Nucleosomes are shown in blue, around which the DNA is wrapped (in black). (Bottom) Zoom into the structure of a nucleosome, where DNA in black wraps the nucleosome. Shown in blue, one of the two molecules forming the nucleosome core, with 4 histones (H2A, H2B, H3 and H4) and their respective tails. The histone tail modifications, acetylation (A) and methylation (M) marks, determine whether the chromatin is active or repressive. Histone acetyltransferases (HAT), histone methyltransferases (HMT), histone deacetylases (HDAC) and histone demethylases (HDM) refer to the enzymes which add or remove these epigenetic marks.

of a gene [173].

The transcription of a gene starts when RNA polymerase binds the promoter region of the gene. Once bound, the RNA polymerase unwinds a short region of the DNA helix, i.e. it separates the double strand, and initiates RNA synthesis using one of the DNA strands as template. It is noteworthy that RNA polymerase, in order to bind to DNA, demands for the cooperation of transcription factors (TFs), which, together with the RNA polymerase, form an initiation complex at the promoter re-

gion of the gene, i.e. they form a 'transcribing machine'. TFs, thus, are needed for gene transcription [173].

TFs, also known as gene regulatory proteins, can either be specific for a particular gene or generic for the transcription of a large number of genes. Specific TFs can still bind promoter regions of other genes, but they will not trigger their transcription. On the contrary, when the appropriate TFs are bound to a particular region of the DNA, i.e. a promoter region of a determined gene, the corresponding gene can be turned on/off depending on the activatory/inhibitory effect of those TFs [173].

A common situation is depicted in Fig. 1.3, where the gene product itself acts as transcription factor. In this case, TFs are found in a dimer form. When dimerising, TFs can either have two protein molecules of the same type (homodimer) or different protein molecules (heterodimer). By regulating gene expression in this way, if the gene is on, as long as TFs are present, it will continue in this state.

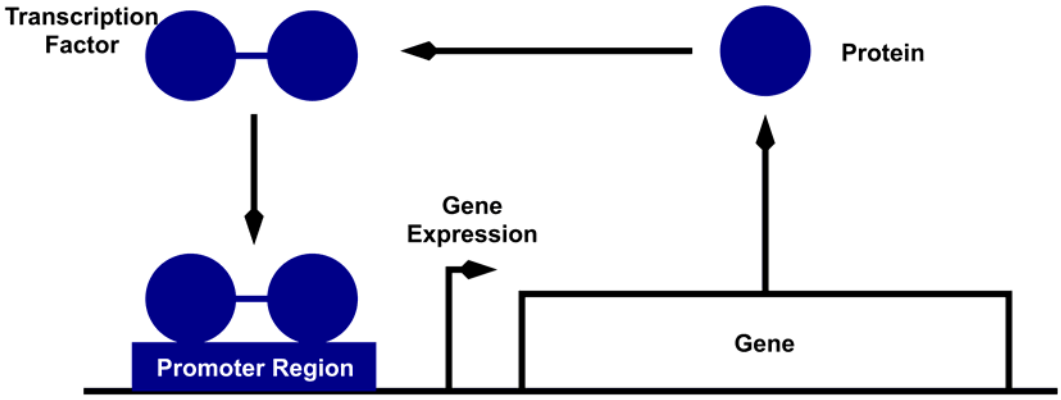


Figure 1.3: Schematic representation of a self-activating regulatory gene. As can be observed, regulation of gene transcription is performed by transcription factors. Protein, which is the gene product, upon dimerisation, acts as transcription factor for that gene that when bound to its promoter region, triggers gene expression. In this case, TF is assumed to be in homodimer form.

Nevertheless, for a gene to remain switched off over an extended period of time, such as the inactive X chromosome in female mammals, some other more reliable mechanism is needed. By analysing the female inactive X chromosome, it has been observed that its chromatin is packed in a much more compact structure (hete-

rochromatin) than the one corresponding to genes being transcribed [173, 117]. In this highly condensed state, the DNA is neither accessible to TFs nor to RNA polymerase and thus, transcription cannot occur, i.e. genes are transcriptionally inactive. Further studies have shown that chromatin structure can also serve to activate genes. Those genes that are switched on are located in a region where chromatin is packed in a more open structure, which allow these genes to be transcribed as TFs can access the DNA (see Fig. 1.4). Therefore, the promoter regions of active genes must be accessible to the transcriptional machinery, whereas inactive genes are located in inaccessible, more compact structures (heterochromatin) that prevent their inappropiate activity [51, 173]. In fact, for a gene to be maintained switched on, presence of TFs and accessible chromatin structure are both needed, because the solely presence of TFs is not enough for transcription to happen. For a schematic illustration of this situation, see Fig. 1.4.

Since any gene can assume different (on or off) trancriptional states depending on the cellular context, chromatin must be able to respond to appropriate cues and signals and change its state. The state of the chromatin depends both on the local chromatin state and on the global chromatin environment in the cell [51]. Some chemical alterations, such as epigenetic modifications, can change the packing state of the chromatin, having thus the ability to switch off some of the genes that are on, or activate some of those genes which are off. These epigenetic modifications, although reversible, are hereditary, which explains why daughter cells have the same gene expression pattern as their progenitors. Epigenetic modifications that disrupt the chromatin structure may come from genetic, metabolic or environmental stimuli, and they are performed by epigenetic enzymes [173].

These chemical modifications to the basic chromatin polymer consist on the addition of chemical marks to the histone proteins (covalent modifications) or directly to a base in the DNA. Such modifications have the capability of allowing or impeding expression of portions of the genome [3], thus deciding which information is made available. In addition to the information of the primary DNA sequence, much of the information regarding where and when to start transcription is stored in such modifications. Chromatin modifications include DNA methylation, or acetylation, methylation, phosphorylation, ubiquitination of the lysine (K) or the arginine (R) residues of the histones. The patterns of these modifications determine accessibility of the transcriptional machinery to the genome (see Fig. 1.4), that is, they are linked to active and silenced transcription [100, 92]. Generally, DNA methylation is related to absence of transcription, whereas histone acetylation is associated with

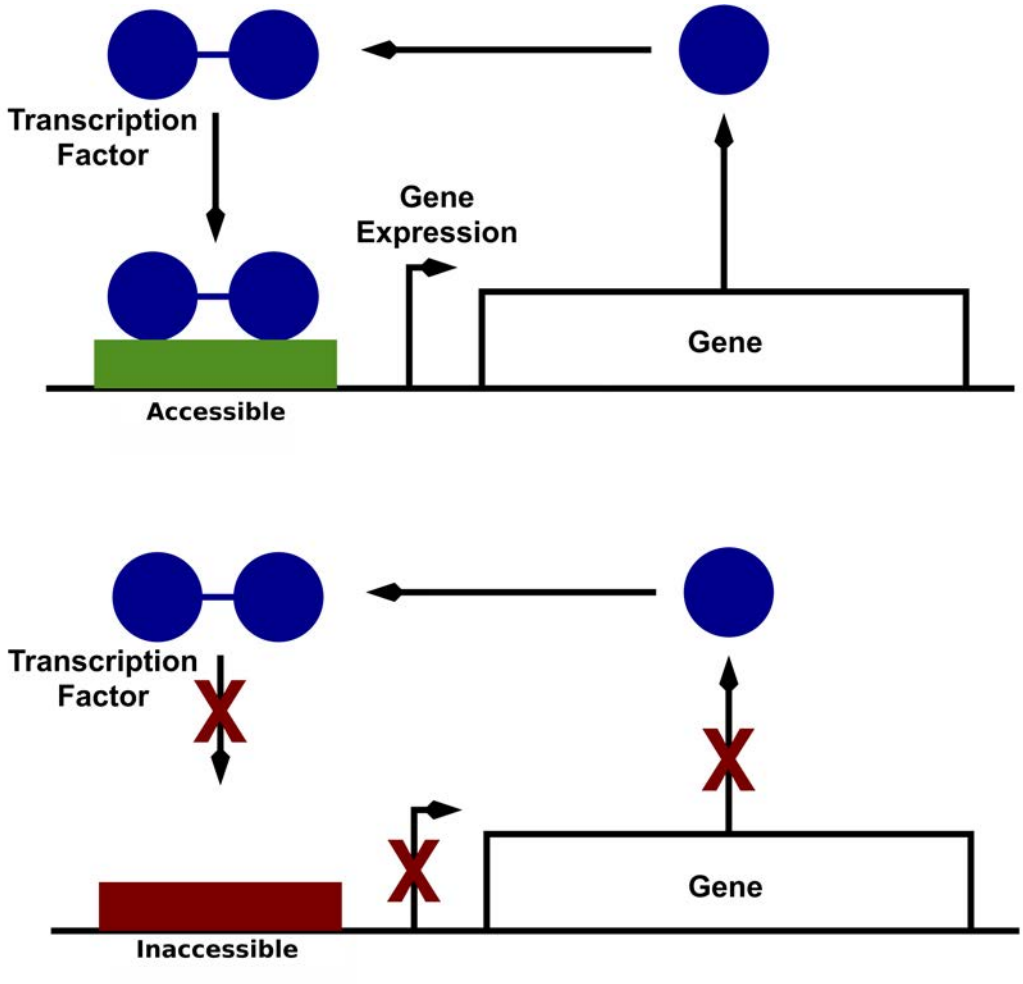


Figure 1.4: Schematic representation of a self-activating regulatory gene when taking into account chromatin structure. Regulation of gene transcription is performed by transcription factors when promoter region is accessible (euchromatin), top panel. If chromatin structure is densely packed (heterochromatin), then, even with the presence of the appropriate TFs, gene transcription cannot occur, because chromatin is inaccessible (bottom panel).

gene transcription [173].

Chromatin modifications can be carried out at different positions on the different histones, conferring a large information capacity on each nucleosome. Their addi-

tion or removal, particularly acetylation (A) and methylation (M), is performed by 4 classes of enzymes, as illustrated in Fig. 1.2: histone acetyltransferases (HATs), histone deacetylases (HDACs), histone methyltransferases (HMTs) and histone demethylases (HDMs). Such modifications can influence the activity of nearby genes, mainly because they affect the ability for certain TFs to bind [42]. The modifications directly to DNA are performed mainly by the DNA methyltransferase enzymes.

1.2.2 Epigenetic plasticity

A primary function of chromatin during development is to stabilise the different cell fates. In terms of the Waddington landscape, when a stem cell starts the differentiation process, chromatin affects the height of the walls that separate the valleys, which correspond to different possible cell fates (see Fig. 1.1 and Section 1.1.1). Once a fate is chosen, chromatin acts as a barrier that prevents cells from switching states. In ESCs, many genes are characterised as bivalent since they exhibit both positive (activating) and negative (inactivating) histone marks. As the cell differentiates, and thus commits to a specific lineage, the chromatin evolves towards more restrictive states (see Fig. 1.5 (b)). Such restrictive states are associated with heightened epigenetic barriers, which increase phenotypic robustness by stabilising the associated state of the gene regulatory system. Hence, chromatin progressively impedes changes of cellular state along with developmental progression [51].

Nevertheless, chromatin states can be altered by epigenetic dysregulations associated, for example, with cancer or age-related factors [51]. Such alterations can produce even more repressive states, thus locking cells in undifferentiated phenotypes (see Fig. 1.5 (b)) or permissive states. Such permissive states are characterised by lower epigenetic barriers, which facilitate stochastic transitions between phenotypes and in some extreme cases, to cell reprogramming (see Fig. 1.5 (c)). This phenomenon is known as epigenetic plasticity, as it confers cells the ability to move along Waddington landscape (see Fig. 1.1), not just downhill as in normal development, but also across it (cell fate transitions, known as transdifferentiation) and even uphill (cell reprogramming) [51].

1.2.2.1 Epigenetics and reprogramming efficiency

Reprogramming to iPSCs requires reprogramming factor binding ($OSKM$), transcription and chromatin state changes, since iPSCs are functionally indistinguishable

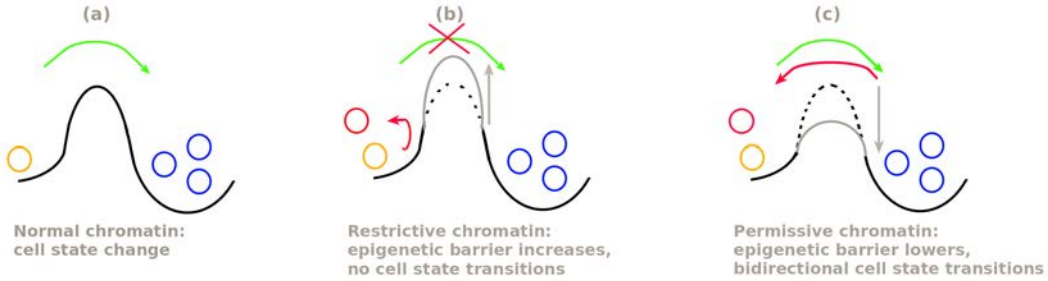


Figure 1.5: (a) Normal chromatin state, where cell adopts a new cellular state. (b) Restrictive chromatin, where the epigenetic barrier increases, impeding cell state transition and locking cells in a particular phenotype state. (c) Permissive chromatin, where the epigenetic barrier lowers, allowing bidirectional cell state transitions, that is, reprogramming and cell transdifferentiation. Both (b) and (c) schemes are examples of epigenetic plasticity.

from ESCs [124]. Consequently, reprogramming involves an epigenetic regulation, as it changes the chromatin state of a somatic cell to that of a pluripotent cell [154]. Widespread epigenetic modifications are the responsible for the changes in the chromatin state and thus, they are important controllers of reprogramming [124]. However, one of the main questions which is poorly understood is when the potential to reprogram is first established. As mentioned above, some authors believe that reprogramming is purely stochastic [70]. According to other views, there are cells that are predisposed to cell reprogramming [66] and, as showed in some studies, by perturbing the epigenetic state before $OSKM$ induction, the fraction of cells that successfully achieve reprogramming can be modulated [129].

According to these studies, variability in reprogramming ability is partly due to preexisting epigenetic heterogeneity (cell-to-cell variability), which can be tuned to alter the outcome of the reprogramming process. In fact, in order to rule out exogenous causes that could explain why some cells reprogram whereas others not, e.g. the existence of some signalling-cues from neighbouring cells, it has been tested whether a certain shared local environment would affect the future outcome of some cells with respect to reprogramming. As shown in the experiments carried out by Pour et al. [129], there is no significant evidence that the initial spatial location of the starting cells will affect the final fate.

These results suggest that the differences in the reprogramming outcome may be

of epigenetic origin. Furthermore, Pour et al. [129] show that the existing innate differences in nuclear state of the initial cells, i.e. different initial chromatin states, may be more permissive or more restrictive for transcription factors to bind the promoter regions [51, 129]. To check this, in [129], they perturb the methylation pattern prior to $OSKM$ induction. By altering the epigenetic marks in this manner, there is a change in the epigenome which posits cells more predisposed to reprogramming. More precisely, methylation patterns are altered by adding Lsd1 and Ezh2 inhibitors [129]. Lsd1 is a HDM which removes H3K4 mono- and di-methylation, and Ezh2 is a HMT that catalyses repressive H3K27 methylation, which has been linked to various cancers [51]. Hence, inhibiting either of them (Lsd1 or Ezh2) results in a more permissive chromatin state, reaching a 3-fold increase in the number of iPSCs obtained [129]. These results show that reprogramming is sensitive to heterogeneity of the epigenetically-regulated chromatin state.

As shown in several studies [150, 129, 66, 70], the molecular mechanisms during in vitro $OSKM$-driven reprogramming have been well characterised, but little is known about in vivo reprogramming. In particular, when studying in vitro reprogramming, its low efficiency has been linked to several barriers or bottlenecks, which are activated by cellular damage and which appear mostly in aged cells [115]. These cell-intrinsic barriers are mediated by tumour suppressor proteins, such as p53, which inhibit reprogramming in vitro [115]. These results on in vitro experiments, suggests a possible role for these tumour suppressors, ageing and cellular damage on in vivo reprogramming.

Recent groundbreaking work by Mosteiro et al. [115] has shed some light on in vivo reprogramming. They have shown that the expression of the Yamanaka factors ($OSKM$) in vivo stimulates reprogramming of a small fraction of cells, whilst causing also extensive damage to the other cells that do not reprogram, taking them to a state known as cellular senescence. Such cells are characterised both by their inability to proliferate and by the secretion of inflammatory cytokines, which altogether leads to tissue remodelling (senescent tissue). For example, cells without the tumour suppressor p53 cannot protect themselves from the effects of $OSKM$ overexpression which implies widespread DNA damage in the tissue, which leads to a large number of senescent cells and, consequently, high cytokine production and increased in vivo reprogramming. However, the absence of some other tumour suppressor, such as *Ink4A/ARF*, has been reported to have the opposite effect, that is, less senescence and less in vivo reprogramming, after $OSKM$ induction. Therefore, the positive correlation between senescent cells and ability to reprogram in vivo seems clear [115].

Such correlation is mediated by the cytokine-rich environment that senescent cells create [115]. Among the inflammatory cytokines produced by senescent cells, interleukin-6 (IL-6) seems to be crucial for senescence ability to trigger reprogramming. Its blocking, reduces both in vivo and in vitro reprogramming. On the contrary, IL-6 upregulation creates an environment which facilitates senescent neighbouring cells to undergo reprogramming, since it creates a permissive tissue environment by remodelling the extracellular matrix, which favours the reprogramming process (it lowers the existing barriers, see Fig. 1.5(c)). Similarly, ageing and tissue damage are two biological conditions related to higher levels of senescent cells which, in turn, implies that they also help reprogramming to occur. These relationships among ageing, tissue damage, senescent cells and reprogramming may be enlightening both for tissue repair and for understanding teratomas appearance, which are tumours arising from iPSCs, more frequent in ageing tissues [115].

1.2.2.2 Epigenetics and metabolism

It has been observed that DNA transcription is regulated by chromatin organization (see Section 1.2.1) and there is increased evidence that metabolic signals play important roles to determine chromatin structure. Recent efforts have been devoted to study the links between metabolism and epigenetic modifications of chromatin, both in normal scenarios and those appearing in some diseases such as cancer [100].

The capacity of the chromatin structure to affect cellular state transitions can differ as a function of metabolic conditions that change during ageing. Ageing is associated with profound changes in the epigenome involving large disturbances of the epigenetic landscape [18, 122]. Studies in model organisms have shown the remarkable plasticity of age-associated epigenetic modifications, due to an ageing-associated loss of epigenetic resilience [67, 14, 125]. Thus, whereas epigenetic alterations in DNA methylation and chromatin remodelling are considered highly conserved hallmarks of ageing [14, 138], the ability of cellular reprogramming-driven epigenetic remodelling to ameliorate age-associated phenotypes has been described recently. This finding unequivocally supports the causative role of epigenetic dysregulation as a driver of ageing [121]. However, it remains intriguing how ageing-related changes in cellular metabolism [62, 175, 24] might control the layers of epigenetic instructions that influence cell fate without involving changes in the DNA sequence [53].

The dynamics of epigenetic modifications are instrumental to connect the extracellular environment and nuclear transcription. More precisely, many signalling pathways in conjunction with transcription factors can remodel the epigenome, by recruiting some enzymes, such as histone modifying enzymes, that can modify chromatin. Nevertheless, there is also increasing evidence that the information about the cell metabolic state is also integrated into the regulation of epigenetics and hence, it is also connected to transcription. Therefore, the link between metabolism and transcription is not unexpected and, in fact, since most chromatin-modifying enzymes require substrates or cofactors that are intermediates of cell metabolism, the relationship appears as more or less evident. Hence, fluctuations of metabolite levels, such as acetyl-CoA or methyl groups, can have implications on the activity of chromatin-modifying enzymes and consequently, alter chromatin dynamics and transcription [100, 107, 109, 87, 148, 44, 113, 110].

Of special interest is understanding the molecular connections of cellular metabolism and epigenome, because in a large variety of diseases, such as cancer or type II diabetes, abnormalities in the metabolism and in the epigenome have been identified [100]. In particular, it may be the case that cellular ageing might result from stochastic translation of metabolic signals into cellular epigenetic states [53]. The reversible nature of epigenetic regulation of ageing is receiving increasing attention as it might offer a revolutionary strategy to delay or reverse diseases clustered in older individuals [89, 120]. Therefore, by increasing the knowledge about the relationship between epigenetics and metabolism, it may be easier to design some therapeutic strategies [100].

Over the past 20 years, various epigenetic enzymes have been identified and characterised. Such enzymes include the histone modifying enzymes (HMEs), which include HATs, HDACs, HMTs and HDMs, as well as DNA methyltransferases and DNA hydroxylases. HMEs add or remove epigenetic marks on the histone tails (see Fig. 1.2), whereas DNA methyltransferases or hydroxylases add the marks directly to the DNA chain. The activity of these chromatin-modifying enzymes is regulated by the concentration of their metabolic substrates or cofactors, such as acetyl-CoA, α-ketoglutarate (αKG) or the ratio NAD^+/NADH [100]. Such metabolic substrates and cofactors are capable of diffusing through nuclear pores thus providing a way for the cell to deliver metabolic information to nuclear transcription. We briefly review some of the better characterised interactions (as mentioned in [100]) between epigenetic modifiers and metabolic players in mammals.

Histone acetylation. Histone acetylation is catalised by HATs enzymes. In mammals, HATs transfer the acetyl group of the acetyl-CoA to the lysine residues of histones. Hence, high levels of acetyl-CoA promote histone acetylation, which helps to activate the expression of certain genes.
In mammalian cells, both enhanced glycolisis and glutamine metabolism in the tricarboxylic acid (TCA) cycle cause generation of acetyl-CoA. Therefore, metabolic alterations (such as those associated with cancer) can change the concentration of acetyl-CoA which can affect histone acetylation levels.

Histone deacetylation. HDACs enzymes catalise the removal of histone acetylation, and they can be classified into two main groups: the classical HDACs and the NAD^+-dependent deacetylases (the sirtuin family). The deacetylation reaction is energetically favourable and this is why sirtuins are a special family, because they deacetylase in an apparent wasteful way, since they depend on NAD^+, which hydrolises to become NADH. Sirtuins are thought to mediate the effect of caloric restriction on extending life span.

DNA and histone methylation. HMTs enzymes add methyl groups to lysine/arginine residues of histones, whilst DNA methyltransferases add methyl groups to DNA. The activity of these enzymes is structurally diverse and they possess high substrate specificities, which makes difficult to generalise their effect. Depending on the histone residue being methylated, the implications can be opposite: either being a silencing or an activating mark.

DNA and histone demethylation. Since a covalent methyl group is chemically stable, DNA and histone methylation were initially considered relatively static epigenetic marks. However, during embryonic development, extensive remodelling of cellular methylome occurs, which suggests that some enzymes that actively remove methylation marks have to exist.
One of the best characterised families of HDMs is the JHDM family (Jumonji-C domain containing HDM). The demethylation mechanism used by JHDMs involves αKG, which is a key metabolite of the TCA cycle. Moreover, αKG is also involved in mitochondria-cytosol transfer of NAD^+ and NADH. Hence, αKG presents a possible metabolic link between HDACs and HDMs.

1.2.2.3 Epigenetics and cancer

We have discussed evidence regarding the key role played by epigenetics on the regulation of gene transcription, as it can allow or restrict access of TFs to their associated promoter regions (see Section 1.2.1 and Fig. 1.4). Epigenetics is thus fundamental in the regulation of cellular processes, such as cell proliferation or cell differentiation. For this reason, when alterations in epigenetic mechanisms appear, they contribute to the initiation of certain disorders, particularly those associated with ageing and cancer. Some of the epigenetic disruptions related to cancer involve DNA methylation, which is epigenetically stable, and thus, preserved through mitosis. Hence, when cancer cells divide, the acquired epigenetic states may be maintanined in daughter cells, thus generating adaptive epiclones that drive malignant progression. This, thus, is an example of how epigenetic modifications can help cancer progression [51].

During normal development, chromatin structure is stable. A deviation from this normal behaviour, either being more restrictive or permissive, is a central element in tumorigenesis (see discussion in Section 1.2.2). Restrictive chromatin states are usually related to blocking differentiation programs, whilst chromatin permissive states, which are associated with epigenetic plasticity, allow malignant cells to activate alternative gene regulatory programs, which can result in cell fate transitions [51]. In this sense, cancer stem cells (CSCs) can be viewed as cancer cells that taking advantage of this epigenetic plasticity, can reprogram towards a less differentiated state.

Stem cells are essential in embryonic development, tissue homeostasis or renewal of certain tissues after injury. Adult stem cells have the capacity for self-renewal under basal conditions and during tissue repair. However, CSCs (also known as cancer initiation cells) have also been identified and they share some of the SC properties: they have a high capacity to self-renew, they can differentiate, migrate or activate anti-apoptopic pathways, if needed [135]. The cancer stem cell hypothesis postulates that, whilst they have capability for differentiation, CSCs have impaired stem cell self-renewal, that is, they divide symmetrically. In other words, CSCs lock cells in the stem cell state, impeding cell differentiation of one of the daughter cells. This fact implies an increase in the tumour stem cell population, which may be a key event for cancer progression [171].

CSCs are particularly relevant since they are resistant to therapy. Therefore, CSCs provide a mean for cancer to survive and regrow after treatment. Furthermore, when trying to diagnose cancer, the role of CSCs markers is key, and understanding

how these markers can be epigenetically modified may be cornerstone in order to design novel therapeutics with CSCs as their target [135]. The anti-stem cell based cancer therapy relies precisely upon identification of some CSCs markers, for which, small molecules targetting them are sought [135, 171].

Besides their role in CSCs, epigenetic modifications play a central role in the progression of cancer. Disruption of epigenetic mechanisms alters the activity of some genes which leads to malignant transformation. In general, these epigenetic alterations consist of global changes in DNA methylation, modifications in histone proteins and alterations in the expression of chromatin-modifying enzymes [51, 135]. Epigenetic dysregulation may be caused by genetic or metabolic stimuli which cause oncogene activation or tumour suppressor silencing. For example, a gain-of-function mutation of the polycomb repressor EZH2 (it becomes hyperactive) promotes chromatin restriction and hinders differentiation. EZH2 is a HMT that it has been found to be overexpressed in some lymphomas and melanomas. Metabolic enzyme mutations that disrupt the DNA methylation balance can cause tumour suppressor silencing. These stimuli, hence, may cause in a premalignant cell to promote tumour initiation or in an already malignant cell, to accelerate tumour evolution and adaptation. Such epigenetic modifications are not something isolated, since roughly 50% of human cancers have mutations in chromatin proteins, remarking the important role of epigenetic aberrations [51].

When specifically examining the epigenetic modifications found in cancer, many tumours exhibit genome-wide hypomethylation and hypermethylation, showing the importance of DNA methylation. A well established characteristic is hypermethylation of the CpG island in the promoter regions of many tumour supressor genes, such as p53 and BRCA1, which inactivates them. Regarding histone modifications, such as their acetylation or methylation, they are responsible for structural changes of the chromatin (see Section 1.2.2.2). Among these epigenetic modifications, histone acetylation, which is controlled by HATs and HDACs, is the one whose effects are better characterised, being a global positive mark of gene activity. Loss of histone acetylation can result in gene repression and HDACs are often found overexpressed in various types of cancer, which hinders the expression of some tumour supressor genes. Contrary to what happens with histone acetylation, there is not a clear interpretation for histone methylation, since its effect (i.e. silencing or activating) is site specific [135].

Purely epigenetic alterations may provide an explanation for tumours that arise

without (or just a few) genetic mutations. In other words, epigenetics may provide a rationale for tumours which lack of genetic signs [51]. This helps to gain a deeper characterisation of cancer and strengthens the consideration of epigenetics as a cancer hallmark.

Another important role for epigenetics in cancer appears in relation to the heterogeneity existing in cancer cell populations. Within the heterogeneity, there exist subpopulations of 'drug-tolerant' cells [139], which achieve resistance by nonmutational mechanisms. More precisely, they exhibit an altered chromatin state with reduced histone methylation, which makes them more resistant to treatment. Therefore, for this resistant subpopulations to persist, the presence of a HDM is needed [139]. By designing strategies able to modify the chromatin or the activity of HMEs, drug resistant subpopulations could disappear. Furthermore, a relationship between the 'drug-tolerant' cancer cell subpopulations and the CSCs is likely to exist [112].

1.2.2.4 Epigenetic alterations at the interface of ageing, metabolism and cancer

Whilst in Sections 1.2.2.2 and 1.2.2.3 epigenetic alterations appearing in metabolism and in cancer have been described separately, they are not independent. In the previous sections, we have given a general overview of epigenetics aspects of metabolism and cancer. Here, we provide a very brief summary of the interplay between epigenetics, ageing, metabolism and cancer.

1.2.2.4.1 Geroncogenesis: metabolic changes associated with ageing drive cancer

The correlation between ageing and cancer incidence rate is a well established empirical fact. The currently accepted explanation for such a correlation is subsumed under the multiple hit hypothesis or Knudson hypothesis [118, 91]. According to this theory, cancer is caused by successive accumulation of gene mutations, and thus more likely to appear at older age as this affords the time needed for such mutations to occur. Accordingly, more than 60% of all cancers occur in people aged over 65 [174]. However, a new paradigm is emerging due to recent experimental results according to which ageing interferes with the normal regulatory mechanisms regulating both differentiation and robustness of differentiated phenotypes. Crucial to these results is the effect of senescent cells. As mentioned in Section 1.2.2.1, Mosteiro et al. [115] have reported that activation of the senescence programme triggers secretion of a

number of signalling cues (mostly inflammatory signals, particularly IL-6), which alter the environment of ageing tissues so that (i) spontaneous reprogramming of somatic cells occurs, whereby they revert back to stem cell-like phenotypes, and (ii) stem cells are locked into self-renewal mode [108]. These two effects combined lead to lessened capability for aged tissues to regenerate and an increase of the self-renewing cell population, which directly correlates with heightened oncogenic risk.

Whilst Knudson's hypothesis provides a rationale for the heightened incidence of cancer with age, it fails to explain why cancer risk is greatly reduced by calorie restriction and physical exercise, letting epigenetics to take a central role and appear as a plausible option to its argumentation. It seems that the time taken to accumulate gene mutations is not the sole cause of the increased risk of cancer with age. Instead, decline in metabolic homeostasis and gene regulation that occurs normally with ageing may also be a key factor [174].

It is thought that during ageing the normal decline in oxidative metabolism constitutes an early hit that drives tumorigenesis. In particular, the sirtuins (see Section 1.2.2.2), which coordinate physiological responses to energetic demand and nutrient intake, seem to be highly related to these metabolic changes. Therefore, the ability to regulate the sirtuins may open a promising avenue to reverse the age-related metabolic changes that are believed to underlie tumorigenesis [174].

One of the latest identified hallmarks of cancer is the dysregulation of cellular energetic pathways [68]. As observed by the German biochemist Otto Warburg, cancer cells utilise glucose in a different way than normal cells [100, 174], commonly known as the Warburg effect. The main characteristic of the Warburg effect is that cells can reprogram carbon metabolism by reducing energy production from oxidative phosphorylation and upregulating glycolisis. This metabolic switch is advantageous to cancer cells because it allows the biosynthesis of macromolecules and organelles required for the rapid cell growth and division, both important cancer properties [106, 174].

The molecular mechanisms of such switch are not clearly understood. Some studies seem to point out to mutations in some metabolic regulators, as those seen in some rare genetic diseases [174]. In general, the most accepted view is that Warburg-like metabolic changes are genetic in origin, but an alternative hypothesis is that the metabolic switch is due to epigenetic changes whose likelihood increases with age but which appear earlier in cancer cell lineages. According to this idea, the Warburg-

like metabolic state will appear in normal tissues and this sets the cell in a system where later mutations can lead tumorigenesis. This hypothesis, called geroncogenesis, states that the driver of tumorigenesis is the ageing-induced metabolic decline, in other words, the ageing-induced dysregulation of mitochondrial metabolism [174].

The age-related metabolic changes can be viewed as an early hit that pushes cells first towards a reprogramming of metabolism and then to a cancer-like state, which is accelerated by sedentary lifestyle and an increased calorie consumption. Geroncogenesis provides a framework to explain why calorie restriction can delay the appearance of the Warburg-like metabolic state (the change in the metabolism) and hence reduce the subsequent chance to develop cancer [111]. Recently, it has been shown that the age-related metabolic decline is associated with epigenetic alterations, which highlights the importance of the epigenetics in this metabolic shift. In particular, this connection has been identified through the sirtuin family [174], as we detail in the following paragraphs.

The decline in metabolism during ageing is due to a loss in activity of longevity regulators that are critical for the maintenance of cellular homeostasis. Sirtuins, a seven-member enzymatic family of nicotinamide adenine dinucleotide NAD^+-dependent deacetylases, are mostly in charge of the control of such regulators. In particular, the founding member of this family is *Sir2*, which is a yeast transcriptional silencing protein that delays ageing in response to calorie intake. In mammals, they have different roles regulating metabolism, DNA repair, cyrcadyan rhytms and ageing. Precisely, its activity decline during ageing (especially SIRT1, SIRT3 and SIRT6) has been implicated in age-related diseases [174].

The reason for sirtuin activity reduction is a decrease in NAD^+ levels, which occurs with ageing, and which is accelerated by obesity and counteracted by calorie restriction and physical activity. Therefore, it is clear that maintaining sirtuin activity at normal levels, helps to hinder tumorigenesis and slows tumour growth, because a change towards a glycolytic (Warburg-like) metabolism is prevented [174].

One of the major limiting factors when trying to fight against cancer is that it mainly consists of mutations (hits in the cancer's progression traditional view), which are irreversible. Nevertheless, the recently found relationship between epigenetics and cancer has brought hope into its treatment, mostly due to the fact that epigenetics is reversible. More precisely, if age-related metabolic changes are an early driver of tumorigenesis, then, molecules which could prevent or reverse these metabolic ageing

may be appropriate for cancer treatment. Being still more precise, small molecules that activate sirtuins may emerge as a satisfactory way to induce a youthful metabolic state, which will make for cancer harder to appear [174].

1.2.2.4.2 The role of oncometabolites in metabolic reprogramming and epigenetics in cancer

In section 1.2.2.4.1 we have discussed metabolic reprogramming as a hallmark of cancer [68]. However, the issue of whether metabolic rewiring in cancer serves just as a way to fulfill the proliferative demands of the tumour cells, or rather it has other implications in driving cancer development, playing an important role, remains still up in the air.

Recent evidence supports the notion that the switch in metabolism plays other roles apart from supplying building blocks to produce new cells [106]. In particular, cancer-associated mutations in metabolic enzymes, such as those found in isocitrate dehydrogenase 1 (IDH1) and isocitrate dehydrogenase 2 (IDH2) have been identified, which are found in low-grade gliomas, acute myeloid leukemias and lymphomas [100, 51].

IDH1 and IDH2 activities are $NADP^+$ dependent, and they interconvert isocitrate and αKG in cytosol and mitochondria, respectively. The mutant IDHs produce 2-hydroxyglutarate (2HG) from αKG, which modulates the activities of enzymes dependent on αKG, mainly inhibiting them. This implies that the activity of JHDMs is altered and hence, these mutations found in cancer have also implications in the epigenetic marks [100]. In fact, cells expressing mutant IDH generate an oncometabolite which leads to DNA hypermethylation. Therefore, IDH mutations found in tumorigenic pathways involve a remodelling of the cellular epigenetic state. However, the most remarkable consequence of this epigenome modification is that the epigenetic changes induced by mutant IDH cause a blockage of cell differentiation: they lock the differentiation genes in a silent state, which facilitates cancer development through accumulation of undifferentiated cells capable of self-renewal, since DNA methylation is stable and thus preserved through mitosis [51, 100].

To summarise, it can be stated that during the differentiation process of progenitor cells, JHDM inhibition by 2HG causes defective histone demethylation and blocks the accessibility of differentiation-related genes. Since 2HG inhibits JHDM, as a consequence, DNA hypermethylates, which permanently locks the differentiation

genes in a silent state. Remarkably, it has been observed that many tumours exhibit a differentiation block [100, 51].

1.2.3 Reinterpreting Waddington epigenetic landscape

According to the picture proposed by Waddington in [162], cell differentiation is depicted as sliding down through the epigenetic landscape, where the valleys correspond to different cell types, and their separation, the walls, are due to epigenetic differences existing between the different cell types (between the valleys), which restrict cell identity: the deeper the well, the more robust the cell state [167, 51].

For decades after Waddington put forward his landscape concept, it was believed that the transitions represented by the landscape were irreversible. The work of Yamanaka [150] established the feasibility of reprogramming the landscape. These results call for the need of reinterpreting the landscape. The way the landscape has to be thought of is as being a dynamical landscape, that is, the height of the walls and valleys, or the slope of the landscape, can change over time. By changing the height of the epigenetic barriers, cell reprogramming can also be explained. Epigenetic plasticity has been identified as a mechanism able to alter the barriers of the landscape [51]. By considering Waddington landscape in this manner, both cell reprogramming and transitions between different states are possible. This, hence, includes as possible ingredients to the landscape the effect of metabolic changes, ageing or diseases such as cancer, which are known factors to alter the epigenetics and, consequently, the height of the walls separating the different states.

1.3 Background on mathematical modelling

In this Section, we provide a brief summary of previous modelling work on epigenetic and gene regulation relevant to this thesis.

1.3.1 Mathematical models of epigenetic regulation

Recent advances in the experimental determination of the mechanisms of epigenetic regulation (ER) have triggered an interest in developing mathematical models capable of reducing their intrinsic complexity to its essential components regarding ER of gene expression [146, 110, 16, 31, 15, 131] and epigenetic memory [42, 37, 6, 144, 16, 31, 145, 131]. References [31, 131] provide comprehensive reviews.

In order to contextualise the model discussed in this thesis which will be presented later, we summarise here the current state of the art in ER modelling.

Models of ER were originally formulated in order to shed light onto the mechanisms of epigenetic memory. Since DNA during the cell cycle is duplicated and, therefore, the epigenetic marks diluted, early ER models were aimed at explaining how epigenetic-regulatory states remain stable upon cell division and transmitted to daughter cells. Such models must satisfy two essential properties, namely, they must be bistable, i.e. each steady state corresponds to an alternative epigenetic state, and the basin of attraction of such states must allow that large perturbations of the ER systems undergoing DNA replication should not change the epigenetic state, allowing, thus, mitotic heritability [37].

Dodd et al. [42] developed the first of such ER models. The authors considered a region of DNA consisting of N nucleosomes, each assumed to be in one of three states, namely, unmodified (U), methylated (M) or acetylated (A). Modifying (HMTs and HATs) and de-modifying (HDMs and HDACs) enzymes carry out nucleosome modifications and removal of marks. A crucial ingredient of the model by Dodd et al. [42] is that histone-modifying enzymes are recruited by modified nucleosomes, thereby providing the necessary positive feedback for the system to be bistable. However, recruitment based on next-neighbours interactions is not enough to produce robust bistability; on the contrary, the coiled structure of DNA appears to impose long-range interactions.

The model by Dodd et al. [42] has been modified and extended in several ways [144]. Sneppen and Dodd have applied the same ideas [145] to modelling the patterns of epigenetic regulation in CpG islands [99]. These authors found that nucleosome dynamics (movement and insertion) is essential to explain experimental data; the key mechanisms appeared to be exclusion of nucleosome occupation in regions of unmethylated DNA and strong correlation between DNA methylation and nucleosome position. Another interesting feature of the model developed by Sneppen and Dodd [144] is that medium-length correlations are provided by the size of nucleosomes, which allows relaxing the requirement for recruited demethylation.

Further, Angel et al. [6] have proposed an ER model to explain quantitative epigenetic control associated with the phenomenon of vernalisation, i.e. the perception and epigenetic memory of a period of cold temperatures to initiate flowering later. Considering that *Arabidopsis* vernalisation is mediated by polycomb repressive com-

plex 2 (PRC2)-based silencing of the floral repressor flowering locus C (*FLC*), Angel et al. [6] developed an ER model in which the essential features were long-range recruitment of positive and negative marks, random insertion of unmodified nucleosomes, as well as DNA replication phenomena diluting epigenetic marks by a 1/2 factor. The ER model by Angel et al. [6] is capable of reproducing not only the patterns of *FLC* silencing following a period of exposition to cold temperatures but also the quantitative dependence with respect to the duration of the exposition to low temperatures. It is noteworthy that, for simplicity, the models [6, 42, 144, 145] do not explicitly take into consideration the protein complexes mediating the epigenetic modifications; rather, they assume that modifications themselves directly affect other modification sites.

Besides the issue of maintaining stable epigenetic memory, recent efforts have been devoted to the study of the regulation of epigenetic modifications by transcription factors [146, 15]. Sneppen et al. [146], based on the experimental observation that TFs can recruit histone-modifying enzymes, have proposed a model where ER is coupled to transcription factors. Based on their previous model of epigenetic modification [42], they assume that TF binding to DNA within a nucleosome further activates the positive modification of the nucleosome (i.e. the U to A transition), with a given rate σ, which is assumed to be proportional to TF concentration and consequently allows to produce ultrasensitivity of gene activation. A similar approach has been recently proposed by Berry et al. [15]. An essential feature of the latter model is the proposed feedback between transcription and epigenetic chromatin modification: activation of transcription depends on the balance between positive and negative modifications, and, in turn, each passage of RNA polymerase II, which is modelled as a discrete event, causes demethylation (see [15] for details). The model by Berry et al. generates robust bistable behaviour capable of sustaining epigenetic memory. Berry et al. argue that this a consequence of the model assumption that transcription-induced removal of methylation marks is performed uniformly over the DNA region, thus providing enough cooperativity for the system to function as epigenetic memory storage. Furthermore, upon exhaustive parameter sensitivity analysis of their model, they find that slow dynamics of chromatin modification buffers the system against noisy transcriptional regulation proceeding in a bursty fashion. An important feature that distinguishes this model from its predecessors is the assumption of next-neighbour recruitment as opposed to long-distance recruitment.

All of the above-mentioned models are of a mechanistic character. Bintu et al. [16] have recently proposed a phenomenological ER model capable of explaining ex-

perimental data obtained by using a reporter gene that expresses a fluorescent protein with induced recruitment of a number of epigenetic-modifying enzymes. This model considers active, reversible silent, and irreversible silent states. Such states represent the active or silent epigenetic states themselves without embodying the activating and silencing mechanisms involved. This model is able to reproduce the data obtained from their experiments and predict the rates of transition between states.

1.3.2 Mathematical models of gene regulatory networks (GRNs)

The gene regulatory network (GRN) concept was firstly introduced by Monod and Jacob, when in the 1960s they discovered that genes can regulate each other's expression. This, in turn, inspired Kauffman [86] who, more than 40 years ago, built the first computational model of a gene network. Since the 1990s, mathematical models of GRNs have been used for a large range of purposes such as graphical representation of regulatory interactions found in high-throughout experiments, as toy models in mathematical modelling, or in order to study some biological questions such as cell differentiation [78].

As Monod and Jacob observed, genes do not act independently, and GRNs account for their interactions. More precisely, it can be said that GRNs are the network representation of the mutual influence between the expression of different genes through molecular regulatory interactions encoded in the genome. In particular, it is the role of epigenetics to store information about the relationships in a GRN.

It is noteworthy that the usual approach (see e.g. [78]) considers a GRN as a time-invariant architecture, as the links represent the same gene regulatory interaction at all time. However, an alternative view considers that GRNs are not fixed networks, i.e. the regulatory interactions may change from inhibiting to activating or vice versa depending on cell type, for example [132].

We consider that the structure of the GRN is static, being encoded in the genome: each interaction is molecular in nature and its specificity and effect are determined. According to this view, the structure of the GRN changes through gene mutations. However, since the time scale associated with these changes is much larger than the organism lifespan, we do not take them into account.

Cell regulatory systems exhibit stochastic behaviour. In particular, gene expres-

sion is known to be stochastic due to fluctuations in the number of TFs, as well as the inherent stochasticity in gene state switching (on or off), depending on the epigenetic regulation. Such randomness causes that a single genotype can have different sustained phenotypes: depending on the gene expression levels, which are affected by chromatin states, the cell fate may be one or another. Therefore, epigenetic regulation has consequences on the GRN behaviour.

As stated in Sections 1.1.1 and 1.2, it is known that one genotype can give rise to different stable phenotypes, so there is not a 1:1 correspondence between genotype and phenotype, as neo-Darwinism has traditionally assumed. GRNs can help to shed light on this concept. GRNs and gene expression noise can, in the absence of genetic mutations, generate a diversity of inheritable phenotypic states, which then can be used as starting point for natural selection. In other words, they provide a non-genetic source of diversity, i.e. a non-genetic phenotypic variability. Therefore, the stochastic dynamics of the GRNs should be taken into account in what would be an improved neo-Darwinian theory. In fact, if the stochastic dynamics was properly incorporated, the fundament of this theory would be broaden, rather than threaten [78].

Cells of multicellular organisms have a genome-scale network of $\mathcal{O}(10^5)$ genes. Although having the same genome-wide regulatory network, cells are of different types, which correspond to different stable patterns of gene expression within the GRN [79]. The influence of expression between genes (repressing or activating), impedes the genome to freely choose any possible gene expression pattern. Some of them are much more stable than others, with some of these patterns being that unstable that they will not be realised ever, because this would go against the interrelationships established by the GRN. So, when deciding which pattern the GRN expresses, several constraints limit the large amount of possibilities that otherwise would be available [78].

Mathematically, each possible state that the GRN can adopt is represented by a point $S(t) \in \mathbb{R}_+^N$, whose components are given by the level of expression of each gene at time t, with N denoting the total number of genes in the GRN. When the dynamics of the GRN changes because of different level of gene expression, then, S moves to another state S' in the N-dimensional state space. This change of state can start because of some random perturbation that moves the system away from its stable configuration S. Upon such perturbation, the GRN either relaxes back to its state prior to the perturbation or, rather, it looks for another stable pattern

where it will settle down [78]. A suitable mathematical approach to describe the GRNs dynamics is offered by dynamical systems. This description accounts for the stable cell states, S, as stable steady states (stable stationary points) of the associated dynamical system [167]. Therefore, trying to predict the long-term behaviour of the dynamical system becomes a biologically interesting question, since it will correspond to a particular cell fate S that the system will adopt.

The problem of differentiation, i.e. the switch from SC into a more specialised cell type, can be subsumed under this paradigm by using genome-wide GRNs, which highlight the existence of high-dimensional attractor states, corresponding to the final cell fates [79]. Since cell differentiation gives rise to many different cell types, the genome-wide GRN exhibits multistability, where each possible cell fate represents a stable steady state. Hence, cell differentiation dynamics can be seen as an ordered process, which stops in a stable state corresponding to a particular phenotype. This has been shown, for instance, in the dynamics of neutrophil differentiation, where a profiling of a 2773-dimensional gene expression state space illustrates that different state space trajectories, which seem to diverge at the first stages of differentiation, finally converge to a common state, the differentiated one [79].

Experimental data regarding neutrophil differentiation [79] raise the question of whether the time course of thousand of genes in GRNs exhibit a globally coherent pattern of attraction to a common state. Although theoretically it makes perfect sense [79], validating this hypothesis experimentally is hard. Available data, generally, represents the topology of the GRNs, and less frequently, their dynamics. The lack of this information is because it is really complicated to experimentally measure it, especially when GRNs get larger. However, some insight has been gained when analysing smaller GRNs. Some of the common properties that these GRNs seem to exhibit include sparseness of interactions and preferential use of certain functions for the interactions between the genes (inhibition or activation, mainly) [79].

The problem of reprogramming (or transdifferentiation) can also be understood in these terms. Cells are initially at one state S, but the GRN can sustain other steady states S' (as mentioned, the system has multistability, which corresponds to all the possible phenotypes). Hence, a change in the gene expression pattern, which may be caused by epigenetic alterations or by ectopic expression of certain genes, may force cells to change state. In particular, ER plays a key role in the switch of phenotype, since it can silence or activate the expression of a certain gene (or genes) and this may settle the cell in a position wehere another stable pattern (another cell

fate) must be adopted. Being able to quantify the statistics of these phenotypical transitions has drawn researchers attention, since it can help to understand how some cells may survive under certain changing or unfavourable conditions [58]. Furthermore, such quantification is also important to determine on which time scale these transitions occur. Thomas et al. [155] show how different timescales may help to the appearance of phenotypic switching in GRNs and they present a new modelling approach that can deal with all these different timescales.

Of particular interest is computing how long it takes for a GRN to switch a gene (or a set of them) from its active to its inactive form, or vice versa. Several approaches to quantify this transition have been used, which are based on the saddle-crossing rate formula [58, 127, 182], which measures the barrier the saddle-point offers in order to reach the other stable state. Such transitions between metastable states occur by crossing the mean-field separatrix through a saddle-point. Between any two of these stable states, there is always an optimal transition path which goes through the saddle point[2].

In general, the transition rate between phenotypic states is defined as the multiplicative inverse of the mean-first-passage time (MFPT) starting from one phenotypic state and reaching the other [58]. In order to have a quantitative measure of the transition rates between states, there is at our disposal several tools. When the system of study is described by the master equation (ME), the Wentzel-Kramers-Brillouin-Jeffreys (WKBJ) theory, more commonly known just as WKB approximation, provides a framework for this computation [19, 47, 75]. Alternatively, when the system is described using the chemical Langevin equation (CLE), the Freidlin-Wentzell (F-W) large deviation theory [54] is used. Both methodologies provide a mathematical derivation for computing the transition rate from one phenotypic state A to another phenotypic state B, which has an exponential form, with exponent minus the system size (or, equivalently, with the inverse of noise intensity). Notice that since it is a transition rate, the larger the system size (or the smaller the noise), the longer it takes for the system to switch, since the transition rate decreases. Such transitions are related to jumps over the epigenetic barriers separating Waddington wells. Epigenetic plasticity, which is the responsible for lowering or heightening the barriers, may thus accelerate or slow down the phenotypic transitions.

Stochastic multistable systems spend most of the time fluctuating around one of the metastable states, although rare events (which are exponentially suppressed with

[2]In particular, a first order saddle point, i.e. with just one component unstable [167]

the size of the system) can occur, which results into transitions between states. The fact that the fixed-point attractors of the mean-field description are stable under stochastic perturbations implies that the phenotypes which those attractors represent are quite robust. Noticeably, the exponential form of the transition rate between mean-field stable states implies that the system is robust to perturbations, i.e. the phenotypes are stable against intrinsic stochasticity [58]. Thus, phenotypic transitions will only happen rarely due to stochastic noise. Intuitively, these transitions between phenotypes are more efficient if the perturbation is applied along an optimal fluctuational path, rather than in any other direction.

Although the F-W theory is formulated in terms of transition between two (mean-field) attractors, it has been extended to deal with more than two attractors [182], something needed when dealing with larger GRNs because they exhibit multistability with usually more than 2 stable steady states. This formulation, in turn, allows to compute relative stabilities of the N attractors the system may have, which implies that once the system switches, the other remaining attractors of the GRN system are not equally likely to be taken [182]. This situation is relevant regarding cells attempting to reprogram to a different cell fate; some transitions between cell fates are more realistic than others [64]. For this formulation to be possible, a global potential function must be defined where each state configuration S is associated with a potential energy [78, 167]. For a review of existing techniques in this area, we refer to [182].

F-W theory and its extensions allow us to explain transitions between cell attractors corresponding to different phenotypes that happen in normal development, but also in disease, when there is a remodelling of the epigenetic landscape. This is something that the deterministic dynamical systems approach usually fails to describe because it focus its attention on the linear stability, but it does not give the dynamics on the transition between states timescale, which is one of the main characteristics of the behaviour of complex stochastic systems [182, 78].

Mathematical models of GRNs have been used to model different biological situations, in particular in the context of cell differentiation. One of them is the interaction between GATA1 and PU.1, where the dominance of one gene or the other represents one of the earliest and most fundamental decisions during haematopoietic development [64]. They have an antagonistic relationship where, for simplicity, it is assumed that expression of PU.1 leads to monocytic cells, whilst expression of GATA1 yields to erythroid cells. The modelling procedure used in [64] progresses in a gradual way by first assuming mutual inhibition, and then, adding self-activation

to the 2-gene regulatory network. This model is reminiscient of a simpler genetic switch controlling the choice between lysogenic and lytic pathways in phage lambda, by the cross-antagonistic and autoregulatory transcriptional regulators Cro and CI. This genetic switch has been tested experimentally using a synthetic genetic network [72].

In general, GRNs with mutual inhibition have been used as a modelling framework for binary cell fate choices, broadly used, but not limited, to lineage specification or lineage reprogramming experiments [64]. More precisely, these GRNs are useful as they steer the system towards one of the states, as well as stopping the regulators of the other possible cell states. A clear well-known example, and one of the simplest GRNs, is the genetic toggle-switch [127, 58]. Usually, this system exhibits bistable behaviour, where two stable states (the on and off states) coexist with an unstable state.

Regarding cell differentiation, if one has the Waddington landscape in mind, one may imagine a particular GRN that would represent the full developmental potential of the genome, since it would have an attractor for each possible phenotype, that is, each possible cell type [78, 64]. The intermediate steps in cell differentiation would be those states where cross-antagonist genes are both coexpressed. These states are, thus, still multipotent progenitors, because they can still give rise to different phenotypes.

Modelling differentiation processes as GRNs where coexpression of mutually inhibiting genes leads to eventual monoexpression of one of those genes, and, consequently, in a well-defined final phenotype, may explain why direct reprogramming may be achieved by inducing ectopically certain factors. Inducing this expression provokes that one stable state solution S is destabilized, the outcome of which may be a different phenotype. Examples of this bifurcation cascade through resolving cross-antagonistic pairs have been modelled in the HSCs, in the T-cells development or when deciding on the predominance of either Nanog or GATA6, which is responsible for deciding whether ESCs maintain their identity (they are kept as SCs), or they differentiate into endoderm [64].

This mathematical approach for cell differentiaton allows us to explain many other features. For example, it helps to explain both why the order of factor inducing may affect the final outcome of cell reprogramming and why it is possible to reach a final state through different paths. Furthermore, it also implies that transitions

between different cell fates are easier to obtain by those states where a gene regulatory switch (branching point) is closer than in those systems where different switches must be overcome. This also means that the number of factors which need to be induced will also depend somehow on the number of cross-antagonistic branching points separating those states [64].

However, when modelling cell fate switch using GRNs there is the possibility of direct reprogramming. In this case, after inducing certain factors (which cause epigenetic alterations), the reprogramming path takes advantage of a change in the height of the walls separating the states, and thus, it crosses the ridge that divides the two-lineage committed territories without reactivating progenitor cells [64]. In fact, "direct" jumps of the system are unusual but feasible, due to the fact that the barrier height separating the basins of the initial and final states can only be finite and therefore, with a perturbation large enough, this should be affordable [78]. Direct reprogramming reinforces the idea that cell differentiation and cell reprogramming follow different paths (as mentioned in Section 1.1.2).

Although it may seem that GRNs have no room for mutations, this is not true. In GRNs, mutations alter the phase space, either by creating or destroying an attractor, or by changing the basin of attraction of at least one state. The first case would allow the system to move to places where it has never been before, to move the system from a place where it had settle down or would modulate the accessibility to a given state, that is, how likely is that state to be reachable [78].

Hence, mutations can cause a gradual change, for instance, increasing/decreasing the contour of the basin of attraction of a state, or a direct effect, when creating a state, for example. The latter case can help to explain some abrupt changes in evolution, by allowing the GRN to settle down in an, until then, impossible state. Since this is a newly created state, evolutionary progress can be triggered because this state could be optimal with respect to other states, since it could be a local minimum more stable. In conclusion, GRNs can give a plausible explanation about how small mutations can create large changes in phenotype, something which has been hard to explain with the neo-Darwinian approach [78].

We have reviewed the use of GRNs to model gene interactions which give rise to the appearance of different phenotypes. Beyond this issue, GRNs can also be used to model cell-cell interactions. Normally, both dynamics, i.e. cell-cell and gene-gene interactions, are analysed using dynamical systems by means of (ordinary, stochas-

tic) differential equations, as mentioned above. In both cases, the appearance of Hill functions is recurrent, and the Hill exponent deserves a special mention. Depending on its value, the type of behaviour the system being modelled can adopt may be completely different. For instance, the appearance of bistability depends on this [105].

Pattern formation can be explained by studying models of intercellular signalling (cell-to-cell interactions). The idea of cell-cell dynamics models is to have the same GRN in each cell and add some relationships between the GRNs of different cells, for example, lateral inhibition between cells which are adjacent. When coupling both models, that is, the one for gene to gene interactions and the one for cell-cell communication, then, interesting results are found in terms of expression dynamics, gene expression attractors, as well as in terms of spatial patterning. A clear example of this coupling is the one described by de Back et al. in [105] for endocrine cells in the pancreas, where results regarding mutlistability of spatial patterns are obtained, something largely missing with other modelling frameworks.

1.4 Aims, objectives and thesis structure

The overarching aim of this thesis is the formulation and analysis of a new multiscale model of gene regulatory networks (GRNs) which account for the effects of epigenetic regulation of gene expression. Specifically, we will focus on GRNs relevant to reprogramming of somatic cells. By careful modelling and analysis of the epigenetic regulatory systems acting upon the GRN, we study the influence of epigenetic plasticity on cell reprogramming in health and disease. We focus on models of epigenetic regulation based on enzyme-catalysed modifications of histones. Our models allow us to shed some light on possible epigenetic mechanisms involved in healthy and pathological cell reprogramming during cancer and ageing.

Specific objectives of this thesis are:

1. Formulating a first model of epigenetic regulation to analyse the effects of epigenetic heterogeneity on the predisposition to cell reprogramming. Within this model, we define two 'epi-phenotypes', one resilient to reprogramming and another one which allows for reprogramming. We generate an ensemble of such ER systems to analyse the features that characterise each phenotype and their

robustness. This is developed in Chapter 2.

2. Formulating and analysing a stochastic multiscale model of a coupled ER-GRN system. We develop a method for stochastic model reduction based on multiple time scale asymptotics. Such reduction allows us to formulate an efficient numerical simulation method. We apply our formulation to investigate the effects of epigenetic plasticity and heterogeneity on the robustness of differentiated phenotypes and cell reprogramming. This is the content of Chapter 3.

3. Formulating and analysing a second model of ER, which is an extension of the model introduced in Chapter 2, that allows us to characterise two types of cellular reprogramming, namely, healthy reprogramming and pathological reprogramming. The former is a physiological type of reprogramming, which is reversible, and contributes to the regenerative capabilities of tissues. The latter is an irreversible form of reprogramming, where reprogrammed cells are not allowed to (re-)differentiate. This model allows for definitions of epi-phenotypes that are compatible with both forms of reprogramming, the physiological and the pathological. We also formulate epigenetic strategies able to move the ER system from the pathological scenario to the physiological one. This new model is developed in Chapter 4.

Finally, in Chapter 5, we discuss both our findings and new research avenues for future work.

Chapter 2

Epigenetic regulation model of cell fate reprogramming in ageing and disease

2.1 Summary

We begin our investigation by formulating a model of epigenetic regulation (ER). We further define an epi-state, epi-phenotype, associated with a somatic, fully-differentiated cell and describe our method, using approximate Bayesian computation (ABC), for generating an ensemble of ER systems compatible with such an epi-state. We show that, in agreement with recent experimental results [129], heterogeneity within the ensemble generates a sub-population where reprogramming is favoured. Further, and also in agreement with Pour et al. [129], we show that such preexisting ER heterogeneity can be harnessed by tuning the activity of ER enzymes (and their cofactors) to alter the cellular response to reprogramming. Our aim in this chapter is to study and to understand the relationship between epigenetic landscapes, cofactor fluctuations and cell state transitions (see Fig. 2.1 for a schematic representation). This chapter is based on our published work [53].

2.2 Model formulation and analysis

In this Section, we provide the details regarding our stochastic model of epigenetic regulation of gene expression which is based on [110]. Our model belongs to a family which considers that single unmodified (U) loci can be modified so as to acquire

positive (A) or negative (M) marks. A positive feedback mechanism is introduced whereby M marks help to both add more M marks and remove A marks from neighbouring loci. The positive marks are assumed to be under the effects of a similar positive reinforcement mechanism [42, 131].

2.2.1 Stochastic model of epigenetic regulation

The stochastic model of epigenetic regulation is formulated in terms of the associated Chemical Master Equation (CME), which, in general, is given by:

$$\frac{\partial P(\mathbf{X},t)}{\partial t} = \sum_i \big(W_i(\mathbf{X}-\mathbf{r}_i)P(\mathbf{X}-\mathbf{r}_i,t) - W_i(\mathbf{X})P(\mathbf{X},t)\big) \tag{2.2.1}$$

where $\mathbf{X} = (X_1, \ldots, X_n)$ is the vector containing the number of molecules of each molecular species at time t, $W_i(\mathbf{X})$ is the transition rate corresponding to reaction channel i and $\mathbf{r}_i$ is a vector whose entries denote the change in the number of molecules of each molecular species when reaction channel i fires up, i.e., $P\big(\mathbf{X}(t+\Delta t) = \mathbf{X}(t) + \mathbf{r}_i | \mathbf{X}(t)\big) = W_i(\mathbf{X})\Delta t$.

Our model (see Table 2.2) is based on the stochastic models for epigenetic regulation proposed by Dodd et al. [42] and Menéndez et al. [111]. The model originally formulated in [42] considers nucleosome modification as the basic mechanism for epigenetic regulation. Nucleosomes are assumed to be in one of three states, methylated (M), unmodified (U), and acetylated (A), and the dynamics of the model is given in terms of the transition rates between these three states.

Dodd et al. [42] consider that direct transitions between M and A are very unlikely. Instead, they assume that transitions occur in a linear sequence given by:

$$M \rightleftharpoons U \rightleftharpoons A,$$

i.e. methylated nucleosomes can only undergo loss of the corresponding methyl group to enter the unmodified state which, then, by means of the intervention of the corresponding histone-modification enzyme, can acquire an acetyl group, and vice versa. They further put forward the hypothesis that such nucleosome modifications are of two types, namely, recruited and unrecruited:

- Recruited modification refers to a positive feed-back mechanism where change in the modification status of nucleosome is facilitated by the presence of other modified nucleosomes (i.e. by the presence of other methylated or acetylated

nucleosomes). Mathematically, this is expressed through a non-linear dependence on the number of M-nucleosomes and A-nucleosomes of the corresponding transition rates (see Table 2.2).

- Unrecruited modification refers to nucleosome modifications whose probability is independent of the modification status of the other nucleosomes.

More specifically, the reactions involved in our model are:

1. HDM-mediated demethylation:

$$M + HDM \leftrightarrows C_M \rightarrow U + HDM$$

   HDM-mediated demethylation can be both unrecruited, where the rates associated with each reaction are constant (see Table 2.2, reactions 1 to 3), and recruited, where all the associated rates are taken to be proportional to the number of A-nucleosomes (see Table 2.2, reactions 4 to 6).

2. Methylation:

$$U \rightarrow M$$

   Methylation can also be unrecruited (Table 2.2, reaction 7) or recruited, in which case the associated rate is proportional to the number of methylated nucleosomes (Table 2.2, reaction 8).

3. HDAC-mediated deacetylation:

$$A + HDAC \leftrightarrows C_A \rightarrow U + HDAC$$

   HDAC-mediated deacetylation can be both unrecruited (see Table 2.2, reactions 9 to 11), or recruited, where all the associated rates are proportional to the number of M-nucleosomes (see Table 2.2, reactions 12 to 14).

4. Acetylation:

$$U \rightarrow A$$

   As with the previous processes, acetylation can also be unrecruited (Table 2.2, reaction 15) or recruited, in which case the transition rate is proportional to the number of acetylated nucleosomes (Table 2.2, reaction 16).

Table 2.1: Random variables

Variable	Description
X_1	Number of unmodified nucleosomes (U-nucleosome)
X_2	Number of methylated nucleosomes (M-nucleosome)
X_3	Number of acetylated nucleosomes (A-nucleosome)
X_4	Number of HDM enzyme molecules
X_5	Number of methylated nucleosome-HDM enzyme complexes
X_6	Number of HDAC enzyme molecules
X_7	Number of acetylated nucleosome-HDAC enzyme complexes

We consider the scenario where both histone hypomethylation (low level of negative (M) marks) and hyperacetylation (high level of positive (A) marks) allow for genes to be expressed, insofar the associated transcription factors are present [100]. On the contrary, we associate histone hypermethylation and hypoacetylation with silent states where genes are not expressed even in the presence of the appropriate transcription factors. In this chapter, we focus on the conditions for bistability to arise and the robustness of the associated *open* and *closed* states particularly in connection with the abundance or activity of HDMs and HDACs.

The transition rates (Table 2.2) associated with each of these reactions are modelled using the law of mass action [59]. This model is an extension of the one formulated by Menéndez et al. [110], since we account explicitly for the enzymatic activity of histone demethylases and histone deacetylases. In our modelling approach, we treat removal of M and A marks by representing them using the Michaelis-Menten model of enzymatic catalysis. This allows us to assess how varying HDM and HDAC abundance can directly affect the behaviour of the ER system, which was not possible in previous models [110]. For simplicity, the activity of HMs and HACs activity is not represented explicitly, thus keeping the model more tractable. By doing so, we are able to add HDAC activity to our analysis, which was absent from [110], and to study the combined effects of impaired HDM and HDAC activities. We also use the current model to perform an exhaustive exploration of the feasible space of parameters, thereby allowing us to produce meaningful predictions as to the effect of ageing and oncometabolic transformation on the robustness of differentiated phenotypes. Our aim thus is to analyse the effects of varying the concentration of these enzymes (HDMs and HDACs) as well as possible synergies between them.

Table 2.2: Random processes and their transition rates. Reaction numbers (R1, R2, R3 or R4) correspond to the enumeration in Section 2.2.1. X_i are as in Table 2.1. (Notation: unr. (r.) denotes unrecruited (recruited))

Transition rate	r	Event
$W_1(x) = k_1 X_2 X_4$	$r_1 = (0, -1, 0, -1, +1, 0, 0)$	M-nucleosome-HDM enzyme complex formation; R1, unr.
$W_2(x) = k_2 X_5$	$r_2 = (0, +1, 0, +1, -1, 0, 0)$	M-nucleosome-HDM enzyme complex splits; R1, unr.
$W_3(x) = k_3 X_5$	$r_3 = (+1, 0, 0, +1, -1, 0, 0)$	Demethylation and HDM enzyme release; R1, unr.
$W_4(x) = k_4 X_2 X_3 X_4$	$r_4 = (0, -1, 0, -1, +1, 0, 0)$	M-nucleosome-HDM enzyme complex formation; R1, r.
$W_5(x) = k_5 X_3 X_5$	$r_5 = (0, +1, 0, +1, -1, 0, 0)$	M-nucleosome-HDM enzyme complex splits; R1, r.
$W_6(x) = k_6 X_3 X_5$	$r_6 = (+1, 0, 0, +1, -1, 0, 0)$	Demethylation and HDM enzyme release; R1, r.
$W_7(x) = k_7 X_1$	$r_7 = (-1, +1, 0, 0, 0, 0, 0)$	Methylation; R2, unr.
$W_8(x) = k_8 X_1 X_2$	$r_8 = (-1, +1, 0, 0, 0, 0, 0)$	Methylation; R2, r.
$W_9(x) = k_9 X_3 X_6$	$r_9 = (0, 0, -1, 0, 0, -1, +1)$	A-nucleosome-HDAC enzyme complex formation; R3, unr.
$W_{10}(x) = k_{10} X_7$	$r_{10} = (0, 0, +1, 0, 0, +1, -1)$	A-nucleosome-HDAC enzyme complex splits; R3, unr.
$W_{11}(x) = k_{11} X_7$	$r_{11} = (+1, 0, 0, 0, 0, +1, -1)$	Deacetylation and HDAC enzyme release; R3, unr.
$W_{12}(x) = k_{12} X_3 X_2 X_6$	$r_{12} = (0, 0, -1, 0, 0, -1, +1)$	A-nucleosome-HDAC enzyme complex formation; R3, r.
$W_{13}(x) = k_{13} X_7 X_2$	$r_{13} = (0, 0, +1, 0, 0, +1, -1)$	A-nucleosome-HDAC enzyme complex splits; R3, r.
$W_{14}(x) = k_{14} X_7 X_2$	$r_{14} = (+1, 0, 0, 0, 0, +1, -1)$	Deacetylation and HDAC enzyme release; R3, r.
$W_{15}(x) = k_{15} X_1$	$r_{15} = (-1, 0, +1, 0, 0, 0, 0)$	Acetylation; R4, unr.
$W_{16}(x) = k_{16} X_1 X_3$	$r_{16} = (-1, 0, +1, 0, 0, 0, 0)$	Acetylation; R4, r.

Specifically, we focus our analysis on plastic behaviour of the epigenetic regulatory states when the activity of histone-modifying enzymes (HMEs) is down-regulated against the background of heterogeneity due to variability in the pool of cofactors for chromatin-modifying enzymes. We proceed by first defining a base-line scenario in which the associated epigenetic regulatory system is such that, for average values of HDM and HDAC activities, the differentiation-promoting gene ER is open and the pluripotency-promoting gene ER is closed. We refer to this scenario as the (normal) differentiated cell epi-state. We then proceed to generate an ensemble of ER systems that satisfy the requirements imposed by this base-line scenario; the necessary variability to generate this ensemble is provided by heterogeneity in abundance of epigenetic cofactors. Analysis of this ensemble reveals that the requirements of the base line scenario restrict the values of a few parameters only, leaving ample flexibility to fix the rest of them. This behaviour is typical of the so-called *sloppy models* [34], where available data constrains a limited number of parameters (or parameter combinations), the system being robust to the choice of a large number of model parameters. In our case, this feature is absolutely essential since, nested within this heterogeneous ensemble of ER systems, there exists a sub-ensemble of plastic ER systems, i.e. ER systems which allow for cell-fate switching.

2.2.2 Mean-field limit and quasi-steady state approximation

In order to gain some insight into the behaviour of the stochastic ER model, we analyse its mean-field limit regarding time scale separation and the quasi-steady state approximation. For a full account of the technicalities we refer to [1, 40].

The mean-field equations, which describe the time evolution of the ensemble average of the variables X_i, associated with the stochastic system with rates given in Table 2.2 are:

$$\frac{dQ_i}{dt} = \sum_{j=1}^{16} r_{j,i} W_j(\mathbf{Q}) \tag{2.2.2}$$

where $\mathbf{Q}$ is a vector whose entries, Q_i, are defined as $Q_i \equiv \langle X_i \rangle$ (i.e., its average value). In order to proceed further, we assume that the variables describing the system are divided into two groups according to their characteristic scales. More specifically, we consider the situation where the subset of chemical species X_i, with $i = 1, 2, 3$, scales as $X_i = Sx_i$, where $x_i = O(1)$, whilst the remaining species are such that X_i, with $i = 4, 5, 6, 7$, scale as $X_i = Ex_i$, where $x_i = O(1)$. Key to our

approach is the further assumption that S and E must be such that $\epsilon = \frac{E}{S} \ll 1$. The averaged variables, Q_i, are similarly divided into two groups: slow variables, i.e., $Q_i = Sq_i$ $(i = 1, 2, 3)$, and fast variables, i.e., $Q_i = Eq_i$ $(i = 4, 5, 6, 7)$. Note that this assumption, also known as the Briggs-Haldane approximation, is standard in enzyme kinetics (see Keener and Sneyd, [88]).

Under this rescaling, we define the following scale transformation for the transition rates in Table 2.2: $W_j(\mathbf{Q}) = k_4 S^2 E \omega_j(\mathbf{q})$. We further rescale the time variable so that a dimensionless variable, τ, is defined as $\tau = k_4 SEt$. It is now straightforward to verify that, upon rescaling, the mean-field equations become:

$$\frac{dq_i}{d\tau} = \sum_{j=1}^{16} r_{j,i}\omega_j(\mathbf{q}),\ i = 1, 2, 3, \tag{2.2.3}$$

$$\epsilon\frac{dq_i}{d\tau} = \sum_{j=1}^{16} r_{j,i}\omega_j(\mathbf{q}),\ i = 4, 5, 6, 7. \tag{2.2.4}$$

with $\epsilon = E/S$.

Eqs. (2.2.3) and (2.2.4) show that if $\epsilon = E/S \ll 1$ holds, then, the mean field equations (Eqs. (2.2.3) and (2.2.4)) naturally display multiple scale structure, which we will exploit to simplify our analysis by means of a quasi-steady state approximation (QSSA) [137].

Multi-scale asymptotic analysis [76] relies upon the construction of two approximations. One of these approximations is constructed so that it is valid for longer times and it is referred to as the *outer solution*. The second one, i.e. the *inner solution*, approximates the behaviour of the system at shorter times. These two regimes must satisfy appropriate matching conditions which ensure that both solutions produce a uniformly valid approximation. The outer solution is usually obtained in terms of the quasi-steady state approximation [88], which describes the dynamics of the system once it has settled down onto the associated invariant manifold. Regarding Eqs. (2.2.3)-(2.2.4), the QSS approximation consists on assuming that $\epsilon\frac{dq_i}{d\tau} \simeq 0$ which leads to a differential-algebraic system of equations which provides us with the QSSA.

In order to obtain the QSSA, we follow the classical approach in enzyme kinetics [76, 22, 88] and assume that the slow variables correspond to the number of U-, M-, and A-nucleosomes, i.e. the *substrates*, whilst the enzymes HDM and HDAC and the

associated complexes are assumed to be the fast variables. Under this assumption, the resulting system displays multi-scale structure which we exploit to analyse the system [1, 40].

The long time behaviour of the system is thus determined by its QSSA:

$$\frac{dq_1}{d\tau} = e_{HDM}\frac{(\kappa_1+q_3)(\kappa_3+\kappa_6 q_3)q_2}{(\kappa_2+\kappa_3)+(\kappa_1+q_3)q_2+(\kappa_5+\kappa_6)q_3} + \\ +e_{HDAC}\frac{(\kappa_9+\kappa_{12}q_2)(\kappa_{11}+\kappa_{14}q_2)q_3}{(\kappa_{10}+\kappa_{11})+(\kappa_9+\kappa_{12}q_2)q_3+(\kappa_{13}+\kappa_{14})q_2} - (\kappa_8 q_2+\kappa_7+\kappa_{16}q_3+\kappa_{15})q_1 \quad (2.2.5)$$

$$\frac{dq_2}{d\tau} = -e_{HDM}\frac{(\kappa_1+q_3)(\kappa_3+\kappa_6 q_3)q_2}{(\kappa_2+\kappa_3)+(\kappa_1+q_3)q_2+(\kappa_5+\kappa_6)q_3} + (\kappa_8 q_2+\kappa_7)q_1 \quad (2.2.6)$$

$$\frac{dq_3}{d\tau} = -e_{HDAC}\frac{(\kappa_9+\kappa_{12}q_2)(\kappa_{11}+\kappa_{14}q_2)q_3}{(\kappa_{10}+\kappa_{11})+(\kappa_9+\kappa_{12}q_2)q_3+(\kappa_{13}+\kappa_{14})q_2} + (\kappa_{16}q_3+\kappa_{15})q_1 \quad (2.2.7)$$

$$q_4 = e_{HDM}\frac{\kappa_2+\kappa_3+(\kappa_5+\kappa_6)q_3}{(\kappa_2+\kappa_3)+(\kappa_1+q_3)q_2+(\kappa_5+\kappa_6)q_3} \quad (2.2.8)$$

$$q_5 = e_{HDM}\frac{(\kappa_1+q_3)q_2}{(\kappa_2+\kappa_3)+(\kappa_1+q_3)q_2+(\kappa_5+\kappa_6)q_3} \quad (2.2.9)$$

$$q_6 = e_{HDAC}\frac{\kappa_{10}+\kappa_{11}+(\kappa_{13}+\kappa_{14})q_2}{(\kappa_{10}+\kappa_{11})+(\kappa_9+\kappa_{12}q_2)q_3+(\kappa_{13}+\kappa_{14})q_2} \quad (2.2.10)$$

$$q_7 = e_{HDAC}\frac{(\kappa_9+\kappa_{12}q_2)q_3}{(\kappa_{10}+\kappa_{11})+(\kappa_9+\kappa_{12}q_2)q_3+(\kappa_{13}+\kappa_{14})q_2} \quad (2.2.11)$$

where the re-scaled parameters κ_j are defined in Table 2.3, and the conservation laws

$$q_4(\tau)+q_5(\tau) = e_{HDM}, \quad (2.2.12)$$

$$q_6(\tau)+q_7(\tau) = e_{HDAC}, \quad (2.2.13)$$

hold. These conservation laws account for the fact that the total number of enzyme molecules, i.e. the enzyme molecules in their free form and those forming a complex must be constant. Hence, the quantities e_{HDM} and e_{HDAC} are defined as $e_{HDM} = \frac{z_0}{E}$ and $e_{HDAC} = \frac{v_0}{E}$, respectively, where z_0 and v_0 are the numbers of HDM and HDAC enzyme molecules, respectively. E is the characteristic scale (i.e. average) of abundance of the histone-modifying enzymes which, for simplicity, has been taken to have the same value for both HDMs and HDACs. This result opens interesting avenues to investigate, since both oncometabolic transformation and ageing appear to reduce the number of both types of enzymes [100, 101, 175, 125, 115, 33, 48]. Therefore, our theory allows us in a natural manner to explore the effects of these anomalies on the stability of epigenetic regulatory states because the values of e_{HDM} and e_{HDAC} are undetermined, so they can be used as parameters.

Table 2.3: Mean-field limit dimensionless parameters

Dimensionless parameters
$\epsilon = E/S$, $\kappa_1 = k_1/(k_4 S)$, $\kappa_2 = k_2/(k_4 S^2)$, $\kappa_3 = k_3/(k_4 S^2)$
$\kappa_5 = k_5/(k_4 S)$, $\kappa_6 = k_6/(k_4 S)$, $\kappa_7 = k_7/(k_4 SE)$, $\kappa_8 = k_8/(k_4 E)$
$\kappa_9 = k_9/(k_4 S)$, $\kappa_{10} = k_{10}/(k_4 S^2)$, $\kappa_{11} = k_{11}/(k_4 S^2)$, $\kappa_{12} = k_{12}/(k_4)$
$\kappa_{13} = k_{13}/(k_4 S)$, $\kappa_{14} = k_{14}/(k_4 S)$, $\kappa_{15} = k_{15}/(k_4 SE)$, $\kappa_{16} = k_{16}/(k_4 E)$

2.2.3 Parameter values and ensemble generation

2.2.3.1 Viability conditions and reference parameter values.

We start this discussion by defining the somatic cell epi-state in terms of appropriate viability conditions. In order to define such viability conditions unambigously, we restrict our discussion to the context of the gene regulatory network used in [110] (see also Fig. 2.1), i.e. a network of mutually repressive differentiation genes and mutually reinforcing pluripotency genes, with further mutual inhibition between differentiation and pluripotency genes [102]. For simplicity, as shown in Fig. 2.1, we consider a network of two genes, one promoting differentiation and another promoting pluripotency. Such systems, despite their simplicity, have proved themselves useful to study differentiation in a number of contexts [77, 105]. Within such a context, the phenotype of a normal somatic, differentiated cell demands that those genes promoting pluripotent behaviour and/or proliferation should be silent, whereas genes promoting differentiation and quiescent behaviour should be active.

Within such a GRN context, we consider that our ER systems are composed of two replicas of the stochastic epigenetic regulation model, as described in Section 2.2.1, each associated with a different gene in the GRN (see Fig. 2.1). Each of these ER systems is characterised by a different set of kinetic rates, κ_j, so that their behaviour can be tuned to the demands of the viability conditions (to be specified below). For the remainder of this chapter, an *open* epigenetic state refers to a steady state of the system where $q_1 \simeq q_2 \simeq 0$ and $q_3 \simeq 1$ (predominance of positive marks). A *closed* or *silent* epigenetic state is associated with $q_1 \simeq 0$, $q_2 \simeq 1$ and $q_3 \simeq 0$ at equilibrium (predominance of negative marks).

The biological rational for these definitions, based on recent experimental evidence, is as follows. Post-translational modification (PTM) of individual histones, such as acetylation and methylation, plays pivotal roles in the epigenetic regulation

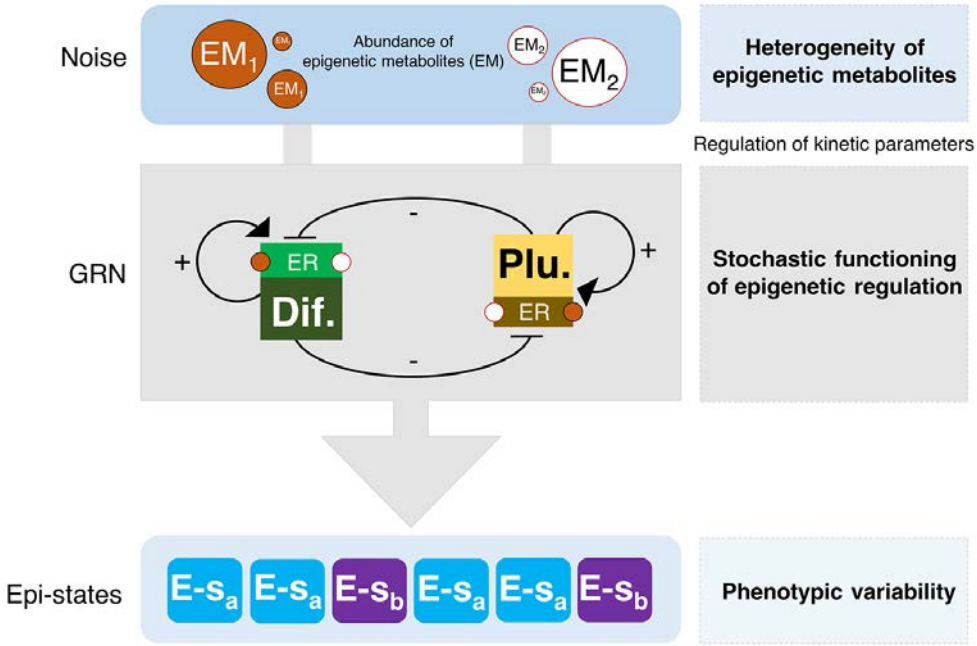


Figure 2.1: A stochastic model of ageing metabolism-regulated cell fate. Schematic representation of the minimal gene regulatory network (GRN) considered in our stochastic model of epigenetic regulation (ER), consisting of a coupled pluripotency and differentiation modules. The heterogeneity of epigenetic metabolites (EM), which operates as regulator of the kinetic parameters (k_j in our model) promoting/impeding the functioning of histone modifiers, stochastically drives phenotypic variability (epi-states). Arrows denote activation and blunt-ended lines denote inhibitory interactions.

of gene expression through chromatin structure changes. Histone acetylation is generally associated with a chromatin structure that is open and therefore accessible to transcription factors and, therefore, gene activation [168, 11, 122]. Histone methylation is linked to either active or repressed genes, depending on the residue that is being modified (e.g., H3K4me3 mark is associated with active promoters whereas H3K27me3 and H3K9me2/3 are associated with repressed regulatory regions).

Although it is likely that the sum of numerous PTMs within regulatory regions determine the transcriptional state of a specific set of genes, for practical reasons epigenetic studies usually involve profiling of one or a couple of well-established hi-

stone modifications. Nevertheless, the silent/closed chromatin state is associated with low levels of acetylation and high levels of certain methylated sites. Our model acknowledges not only that; during ageing, the abundance and activity of enzymes in charge of adding and removing histone marks, such as HDMs and HDACs, changes, mainly because these chromatin-modifying enzymes for both activating and repressing histone marks require certain cofactors (such as metabolites) which change their abundance. This adds a level of complexity which is considered in the model.

For each component of the ER system (differentiation and pluripotency epigenetic regulation), we have set the parameters κ_j so that they satisfy the following general viability conditions, namely, (i) when $e_{HDM} = e_{HDAC} = 1$ (normal or average enzymatic concentration), the regulatory system is mono-stable (stable open chromatin state for the differentiation ER, stable closed chromatin state for the pluripotency ER, as shown in Fig. 2.4), and (ii) for $e_{HDM} < 1$, $e_{HDAC} < 1$ both the differentiation and pluripotency ER exhibit a bistable regime. These conditions correspond to a base-line scenario where we are considering the epigenetic regulation of a gene regulatory system which, in a normal, terminally differentiated cell, involves that differentiation-promoting genes should be open and accessible to the transcriptional machinery of the cell, whereas pluripotency-promoting genes are epigentically silenced [110]. Reference (default) parameter values satisfying the viability conditions for the different scenarios described later on are given in Section A.1.3, Tables A.5 and A.6.

2.2.3.2 Ensemble generation.

Beyond the behaviour of the ER system for the reference parameter sets (Tables A.5 & A.6 and Tables A.7 & A.8 in Section A.1.3), we have generated an ensemble of ER systems to analyse the robustness of the different scenarios we analyse later on in Section 2.3. Such ensemble is generated using approximate Bayesian computation (ABC) [156].

Approximate Bayesian Computation (ABC) methods have been devised to tackle those inference problems for which the estimation of the likelihood function is computationally too demanding. Let $\theta = (k_1, \ldots, k_n)$ be the vector whose components are the parameters to be estimated and x be the data. The general aim is to approximate the so-called posterior distribution, $\pi(\theta|x)$, i.e. the conditional probability of θ given the data, from a prior distribution of the parameters, $\pi(\theta)$. In general, the posterior is given by $\pi(\theta|x) \sim f(x|\theta)\pi(\theta)$, where $f(x|\theta)$ is the likelihood function.

All ABC methods follow the same generic procedure:

- Sample a candidate sequence of parameters, θ, from the proposed prior distribution, $\pi(\theta)$.
- Sample or simulate a data set x from the model represented by the conditional probability density $f(x|\theta)$.
- Compare the simulated data set, x, to the experimental data, x_0, according to some distance function, $d(x, x_0)$. If $d(x, x_0) \leq \epsilon$, where ϵ is an a priori prescribed tolerance, then θ is accepted.

The result of this algorithm is a sample of parameters from a distribution $\pi(\theta|d(x, x_0) \leq \epsilon)$. The above procedure has, in general, the obvious shortcoming that it is often difficult to determine an appropriate distance function between full data sets. To overcome this obstacle, one often uses a distance between (a set of) summary statistics, $S(x)$ and $S(x_0)$, of the datasets, such as the mean and/or the variance.

In our case, we have used the following version of the ABC rejection sampler method [156], taking $\theta = (k_1, \ldots, k_{16})$:

1. Sample θ^* from $\pi(\theta) = \prod_j \pi_j(k_j)$ where $\pi_j(k_j) = U(0, 6.5 \cdot 10^6)$ for k_j, $j = 2, 3, 7, 10$ and 15, and $\pi_j(k_j) = U(0, 5 \cdot 10^4)$, otherwise, for the differentiation-promoting gene, and $\pi_j(k_j) = U(0, 6.5 \cdot 10^6)$ for k_j, $i = 2, 3, 7, 10, 11$ and 15, and $\pi_j(k_j) = U(0, 5 \cdot 10^4)$, otherwise, for the pluripotency-promoting gene.
2. Simulate data set, x^*, from the Master Equation with transition rates as per Table 2.2 using Gillespie's stochastic simulation algorithm. We generate 10 realisations and collect data at times t_i, $i = 1, \ldots, 25$, corresponding to the raw data time points.
3. For each time point, t_i, we compute two summary statistics: the mean over the 10 realisations, $\overline{x^*}(t_i)$, and the associated standard deviation, $\sigma^*(t_i)$.
4. If $\sum_{i=1}^{25} \left(\overline{x^*}(t_i) - \overline{x_0}(t_i)\right)^2 \leq \epsilon_1$ and $\sum_{i=1}^{25} \left(\sigma^*(t_i) - \sigma_0(t_i)\right)^2 \leq \epsilon_2$ hold, θ^* is accepted, where x_0 denotes the experimental data.
5. Go back to Step 1.

We run this algorithm until the number of accepted parameter sets reaches 10000. This method has been used to generate an ensemble of differentiation epigenetic regulation systems (see Fig. 2.2) and an ensemble of pluripotency epigenetic regulation systems (shown in Fig. 2.3).

Our approach is as follows. For each mode of epigenetic regulation (i.e. differentiation or pluripotency ER), we generate simulated data (denoted as "raw data" in Fig. 2.4) using the stochastic simulation algorithm on the model defined by the transition rates Table 2.2. This simulated data will play the role of the raw data, x_0, to which we wish to fit our model. We consider two different data sets x_{0_d} and x_{0_p}, corresponding to the differentiation gene (reaction rates from Table A.5 in Section A.1.3) and the pluripotency gene (reaction rates from Table A.6 in Section A.1.3), respectively. Each data set consists of 10 realisations and 25 time points per realisation. For each time point, t_i, we consider two summary statistics: the mean over realisations, $\bar{x}(t_i)$, and the associated standard deviation, $\sigma(t_i)$. We then run the ABC rejection sampler method until we reach an ensemble of 10000 parameter sets for each epigenetic regulatory system, which fit the simulated data, x_0, within the prescribed tolerances for the mean and standard deviation. Fig. 2.4(a) & (b) shows results comparing the reference (raw simulated) data to a sub-ensemble average, the 200 sets that best fit the raw data (full posterior distributions are shown in Fig. 2.2, differentiation-promoting gene, and Fig. 2.3, pluripotency-promoting gene).

The above procedure provides us with an ensemble of parameter sets that are compatible with the raw data, i.e. such that they fit the data within the prescribed tolerances. The heterogeneity associated with the variability within this ensemble has a clear biological interpretation. The rates k_j are associated with the activity of the different enzymes that carry out the epigenetic-regulatory modifications (HDMs, HDACs, as well as, histone methylases (HMs) and histone acetylases (HACs)), so that variation in these parameters can be traced back to heterogeneity in the availability of cofactors, many of them of metabolic origin such as NAD+ or acetyl-CoA, which are necessary for these enzymes to perform their function (as illustrated in Fig. 2.1).

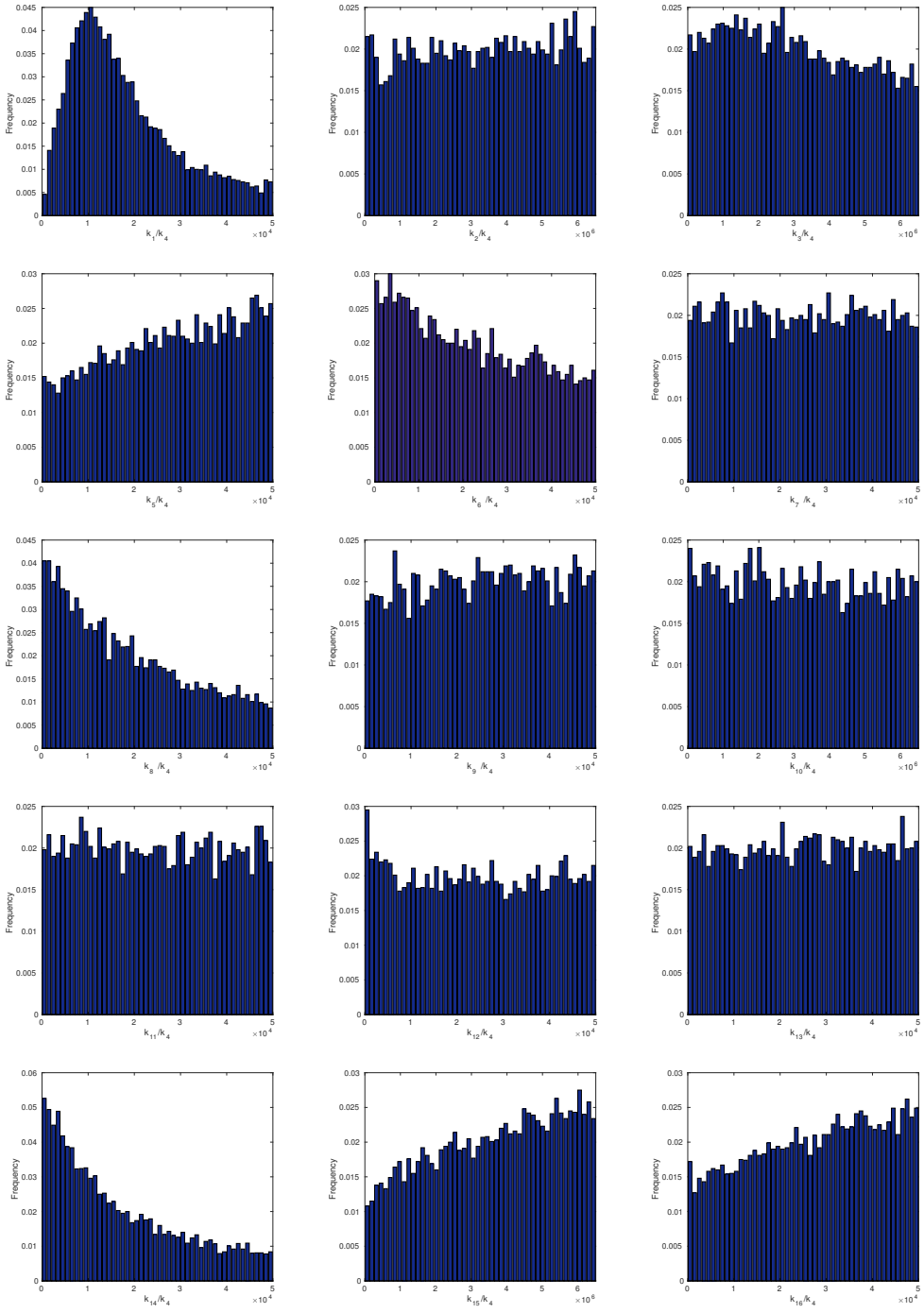


Figure 2.2: Differentiation gene. $\epsilon_1 = 0.6$ and $\epsilon_2 = 1$. $z_0 = v_0 = E = 5$ and $S = 250$.

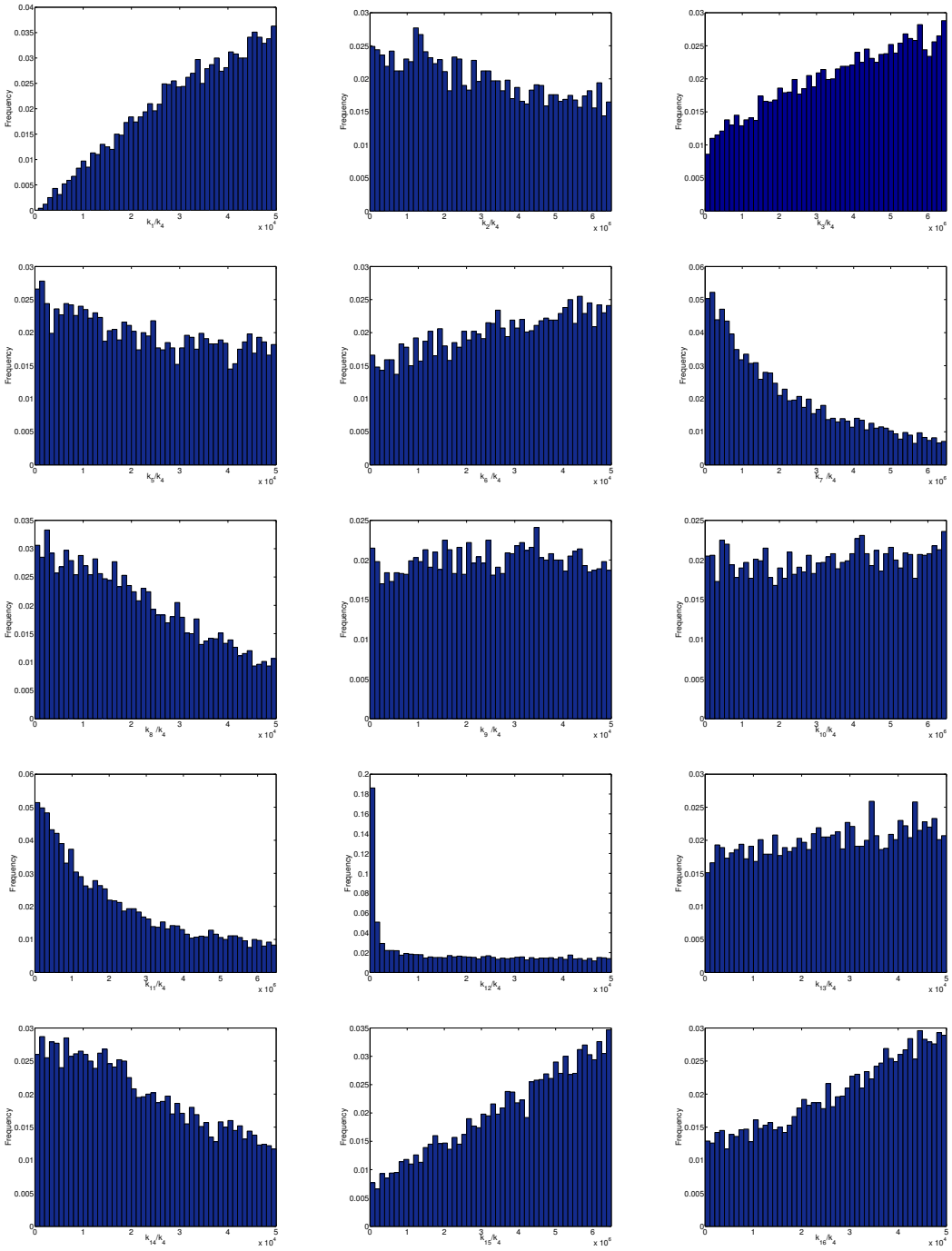
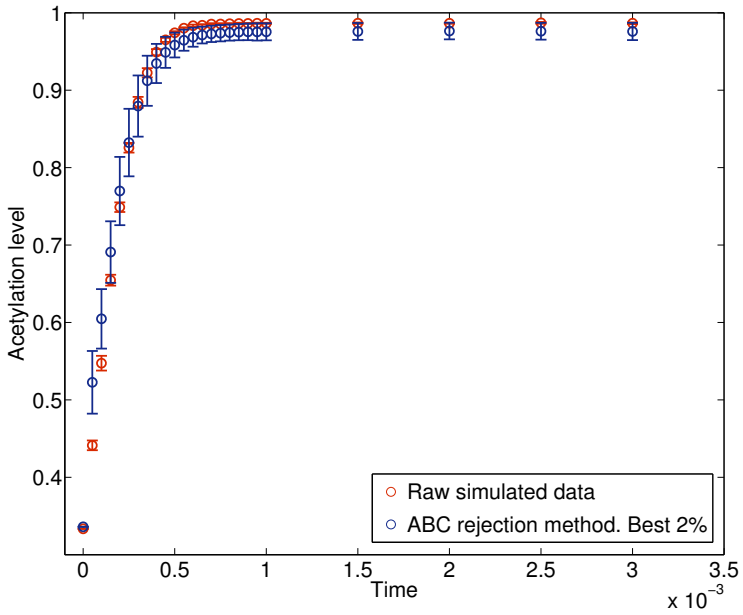


Figure 2.3: Pluripotency gene. $\epsilon_1 = 0.4$ and $\epsilon_2 = 1$. $z_0 = v_0 = E = 5$ and $S = 250$.

(a)



(b)

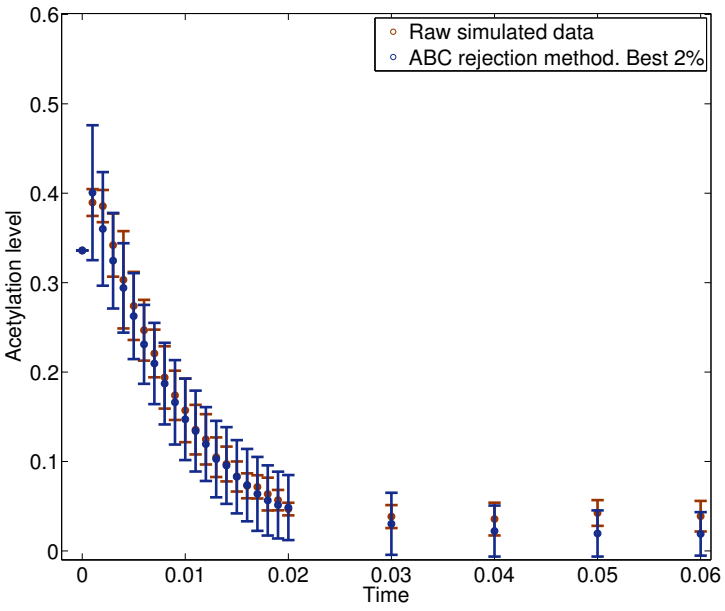


Figure 2.4: Caption on the following page.

Figure 2.4: (Previous page.) Plot (a) shows results regarding the parametric sensitivity analysis of the epigenetic regulatory system for the differentiation-regulating gene. Plot (a) shows the comparison between the raw simulated data and the ABC ensemble average, limited to the 200 ABC parameter sets that best fit the data. Plot (b), idem for the pluripotency-regulating gene. Raw simulated data is generated by using the SSA on the model defined by the rates shown in Table 2.2 with parameter values given in Tables A.5 and A.6 in Section A.1.3, respectively.

The rationale for this procedure is as follows. All the parameter sets within the ensemble are such that for normal or average concentration of HMEs exhibit the behaviour of the base-line scenario, i.e. mono-stable open(closed) chromatin state for the differentiation(pluripotency) ER system. However, and guided by the results reported by Pour et al. [129], we will show that heterogeneity within such ensemble allows to define a sub-population of ER systems with elevated reprogramming potential.

Furthermore, the generated kinetic rate constants are dimensionless, i.e. they are relative to a global scale associated to k_4 (see Table 2.3). Such feature implies that there is an undetermined time scale in our system associated with the (inverse of the) rate constant k_4. This additional degree of freedom can be used to fit our model of epigenetic (de-)activation to particular data. Furthermore, the global time scale corresponding to the differentiation ER regulation (i.e. de-silencing dynamics, Fig. 2.4(a)) does not need to coincide with the global time scale associated with the pluripotency ER system (i.e. silencing dynamics, Figure 2.4(b)). Therefore, our model has the capability of reproducing different systems characterised by different time scales as previously shown by Bintu et al. [16].

2.3 Results

We explore the behaviour of our system as the number of HDMs and HDACs vary relative to their average abundance against the background of variability provided by our ABC-ensemble approach.

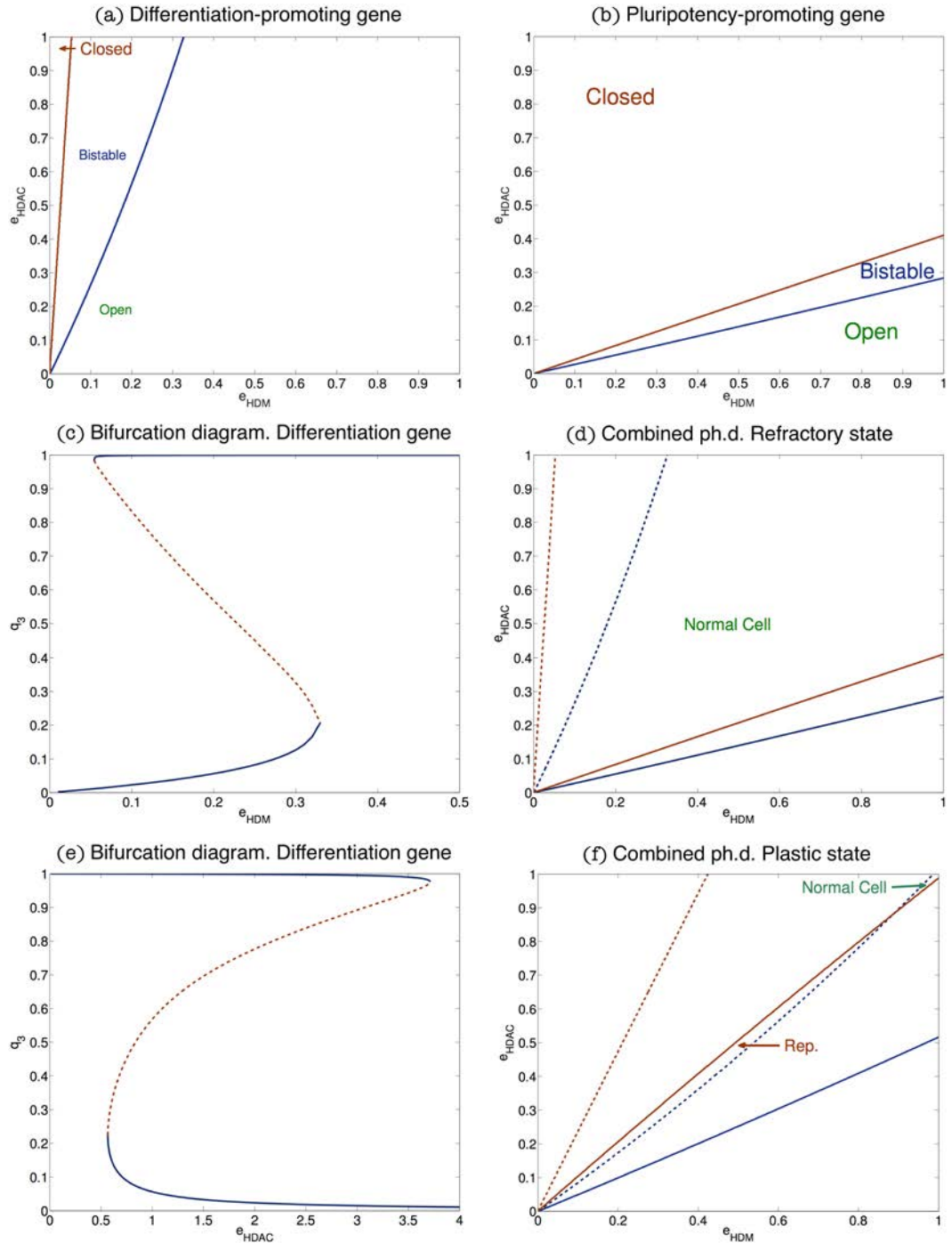


Figure 2.5: Caption on the following page.

Figure 2.5: (Previous page). Plots (a) and (b) show the phase diagrams associated with the QSS approximation for the differentiation and pluripotency promoting genes, respectively. We examine the stability properties of the QSSA as when e_{HDM} and e_{HDAC} are varied. The system exhibits bistability in the region between the red and blue lines. In the region above the red line the only stable steady state is the closed state. By contrast, in the region below the blue line only the open steady state is stable. Parameters values are given in Table A.5 in Section A.1.3 for the differentiation-promoting gene and Table A.6 in Section A.1.3 for the pluripotency-promoting gene. Plots (d) and (f) show the combined phase diagram for both the differentiation-promoting and the pluripotency-promoting models of epigenetic regulation for two clinically relevant cases. In both plots, solid (dashed) lines correspond to the stability limits of the pluripotency(differentiation)-promoting gene. In plot (d), the region between the solid red line and the dashed blue line is associated with *normal cell* behaviour, i.e. open differentiation-promoting gene and silenced pluripotency-promoting gene, whereas in Plot (f), the region marked as *Rep.* is associated with epigenetic regulation configurations which facilitate cell reprogramming. Plot (d) shows a *refractory* epigenetic scenario and Plot (f) depicts a *plastic* scenario. Parameter values for Plot (d) as per Table A.5 in Section A.1.3 (dashed lines) and Table A.6 in Section A.1.3 (solid lines). Parameter values for Plot (f) are given in Table A.7 in Section A.1.3, and Table A.8 in Section A.1.3. Plots (c) and (e) show two bifurcation diagrams, i.e. two sections of Plot (a), corresponding to the differentiation-promoting gene, of the QSS approximation. Plot (c) corresponds to fixing $e_{HDAC} = 1$ and letting HDM activity to vary. Plot (e) examines the bifurcation properties of the system for $e_{HDM} = 0.2$ as HDAC concentration changes.

2.3.1 Variation in the abundance of HDM and HDAC drives epigenetic switch

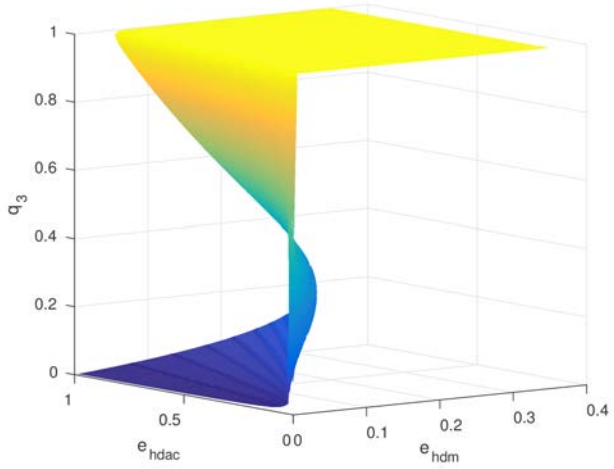
We first focus on a bifurcation analysis of the mean-field QSSA equations (2.2.5)-(2.2.11), to investigate the qualitative behaviour of the ER system as the relative abundances of HDMs and HDACs are varied. Results are shown in Figs. 2.5(a) and (b). In particular, the phase diagram of both ER systems obtained by varying the parameters e_{HDM} and e_{HDAC}. Both these diagrams display three differentiated regions: one in which the only stable steady-state is the one associated with a silenced gene, another one in which the only stable steady-state is the corresponding to an open gene, and a third one where the system is bistable. Fig. 2.5(a) is associated with the differentiation-promoting gene, and Fig. 2.5(b) corresponds to the

pluripotency-promoting gene (parameters given in Table A.5, Table A.6 in Section A.1.3, respectively). Since the qualitative behaviour of the differentiation-promoting gene and the pluripotency-promoting gene is the same, without loss of generality, we will focus on the description of the behaviour of the differentiation-promoting gene. In order to clarify the three regions (open, closed and bistable) displayed in Fig. 2.5(a), Fig. 2.6(a) shows a 3D plot, where the vertical axis shows the level of positive marks (q_3). This plot shows that the system dysplays bistable behaviour: depending on the parameter values e_{HDM} and e_{HDAC}, the system may be both in the open state (high levels of q_3, top of the plot), or in the closed state. Fig. 2.6(b) displays the projection on the xy-plane of the plot shown in Fig. 2.6(a), where we can clearly identify the three regions described in Fig. 2.5(a).

A more detailed picture of the situation illustrated in Fig. 2.5(a) and Fig. 2.6 is given in Fig. 2.5 (c), which shows the bifurcation diagram where e_{HDM}, i.e. HDM concentration, is taken as the control parameter, whilst keeping e_{HDAC} constant. In particular we show the steady state value of q_3, i.e. the variable associated with positive marks, as a function of HDM concentration. This allows to distinguish the three regions displayed in Fig. 2.5(a). We observe that a decrease in HDM abundance makes the corresponding gene inaccessible to the transcription machinery, i,e., it closes the chromatin (corresponding to the closed region, Fig. 2.5 (a)). As HDM concentration recovers, the system enters a bistable regime where both the active and silent states coexist (region marked as bistable in Fig. 2.5(a)). Further increase of the demethylase concentration drives the system through a saddle-node bifurcation, where the closed chromatin state and the saddle point collide resulting in mutual cancellation. Beyond this saddle-node bifurcation, the only stable steady-state is the active state, corresponding to open chromation (region labelled as open in Fig. 2.5(a)). It is noteworthy that these results are in agreement with the oncometabolic transformation scenario associated with IDH mutations proposed by Thompson and co-workers [101, 100] in which downregulation of HDM activity locks differentiation genes into a silenced state which favours reprogramming of differentiated somatic cells into a pluripotent cell [110]. The association between IDH mutations and cancer progression has been well established in the case of glioblastomas and acute myelogenous leukaemia [32, 12, 9, 81].

In Fig. 2.5(e), we show the bifurcation diagram associated with fixing e_{HDM} and varying $e_{HDAC}$of the differentiation-regulating gene. Reduced HDAC concentration recovers the base-line state where the epigenetic regulatory machinery is set to the open state. As HDAC concentration recovers, the system enters a bistable regime in

(a)



(b)

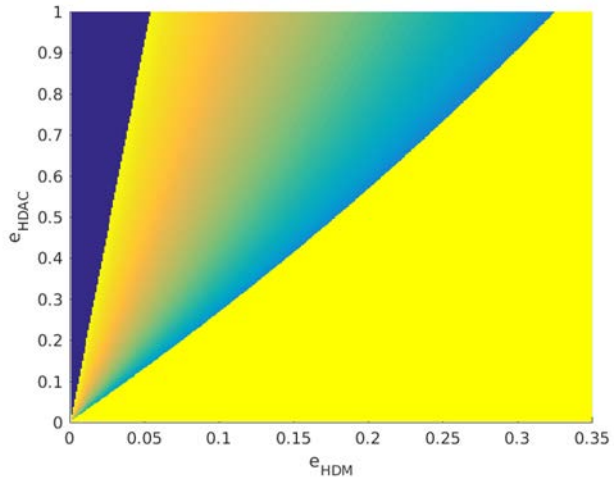


Figure 2.6: Plot (a) shows a 3D plot, where the x-axis represents e_{HDM}, y-axis represents e_{HDAC} and the z-axis represents the steady state value of positive marks, q_3. Depending on the q_3 value, the system can be open (high value of q_3), closed (low value of q_3) or bistable (region where the two states coexist, together with an unstable state). Plot (b) represents a projection of the plot shown in (a) on the xy-plane. In this plot, we can again identify the three regions: closed (left region), bistable (middle region) and open (right region). These regions can be easily understood by matching the color of each region to the ones shown in Plot (a), which, in turn, can be related to levels of q_3.

which both the active and silent states coexist. Further increase in HDAC activity locks the system into the close chromatin state so that the gene is silenced. This implies that reduced HDAC activity may help to rescue differentiation-regulating genes from the effects of IDH mutation.

In these plots (Fig. 2.5 and Fig. 2.6) we can see that variation of the activity of HDM and HDAC enzymes, can alter the stability of the epigenetic regulatory state. In Section A.1.1 of the Appendix, we have verified the accuracy of the predictions of the bifurcation analysis by means of direct simulation of the stochastic model (Table 2.2) using the stochastic simulation algorithm, commonly known as Gillespie's algorithm [59, 60].

2.3.2 Mean-field analysis of the stochastic epigenetic regulation model: refractory vs plastic scenario

We now proceed to analyse in more detail the implications of the bifurcation analysis regarding robustness of the somatic cell epi-state. Fig. 2.5(d) shows the combined phase diagram of both modes of epigenetic regulation (differentiation- and pluripotency-promoting). In this combined diagram, the region between the solid red line and the dashed blue line represents the part of the phase space where the differentiation genes are open and the pluripotency genes are closed (region marked as Normal Cell in Fig. 2.5(d)). This sub-space is therefore associated with normal, differentiated somatic cells. As it has been previously shown in [110], efficient reprogramming requires both closed differentiation genes and open pluripotency genes. Such situation is not viable under the scenario shown in Fig. 2.5(d) because these two conditions cannot hold simultaneously, which we therefore dubb as the *refractory scenario*.

By contrast, Fig. 2.5(f) corresponds to a *plastic scenario*, where, under appropriate conditions, cells become poised for reprogramming. The main difference with the refractory scenario is the intersection between the bistability regions of both the differentiation regulator and the pluripotency gene. In Fig. 2.5(f), the regime where both bistability regions overlap is the one between the red solid line and the blue dashed line (region marked as Rep. in Fig. 2.5 (f)). Within this region, since both genes are in the bistable epigenetic regulatory regime, it is possible to find the differentiation gene in its closed state and the pluripotency gene in the open state. Such situation makes reprogramming much more likely to occur [110] and therefore we identify this feature of the phase space with plastic behaviour. Therefore, by

driving the ER system into this region by means of down-regulation of both HDM and HDAC activity, cells become epigenetically poised to undergo reprogramming. This is consistent with evidence according to which both oncometabolic transformation (e.g. IDH mutation leading to down-regulation of JHDM activity [100, 101]) and ageing (e.g. down-regulation of SIRT6 [175, 125, 115]) induce loss of HDM and HDAC activity thus facilitating reprogramming.

These results are consistent with the results of Pour et al. [129] regarding the existence of a sub-population with higher reprogramming potential which is, at least in part, due to preexisting epigenetic heterogeneity.

2.3.3 Heterogeneity and robustness of the refractory and plastic scenarios

Once we have established the possibility of two epigenetic-regulatory scenarios, we need to assess their robustness regarding variations in the kinetic rates, k_j, as well as to determine whether there are key parameters which exhibit signicant biases related to either type of behaviour. In order to study these issues, we have carried out an exhaustive parameter sensitivity analysis. We have generated an ensemble of differentiation and an ensemble of pluripotency ERs characterised by the corresponding parameter sets $\theta = (k_j,\, j = 1, \ldots, 16)$ (see Table 2.2) compatible with simulated data for the epigenetic regulation systems, as explained in Section 2.2.3.2.

In each ensemble, we first identify those sets θ that satisfy the viability conditions within the tolerances set in the ABC-method (as described in Section 2.2.3.1). We then proceed to classify the elements within the viable subensemble according to whether they exhibit refractory or plastic behaviour. Once we have thus classified the ER systems in each ensemble, we compare the empirical cumulative distribution functions (CDFs) of each k_j. The rationale for such an investigation is that the requirement upon system behaviour should reflect on statistically significant biases of the CDFs of key parameters. In particular, we consider that a specific scenario is sensitive to parameter k_j if its CDF is significantly different from that of a uniform distribution [50]. Throughout the section, we have imposed a level of confidence of 95 %.

We first consider the differentiation ER system. In particular, we focus on the sub-ensemble of the 400 parameter sets that best fit the raw data. Within such sub-ensemble, we proceed to evaluate the robustness of the different scenarios we

are considering. We first analyse the base-line scenario for the epigenetic regulation of a differentiation-regulated gene, namely, (i) when $e_{HDM} = e_{HDAC} = 1$, the regulatory system is mono-stable (only the open chromatin state is stable), and (ii) for $e_{HDM} < 1$, $e_{HDAC} < 1$ there exists a region of bistability. Out of all the parameter sets of the considered sub-ensemble, only 94 fulfill these requirements (93 sets from the ensemble generated and the one created corresponding to Table A.7 in Section A.1.3). We refer to these as the *viable set* or *viable subensemble*. The remaining 307 are bistable at $e_{HDM} = e_{HDAC} = 1$, and they will be referred to as the *non-viable set*. In Fig. 2.7, we present the corresponding CDF of each k_j within both sets.

Regarding the viable set, we first seek to assess which kinetic constants have distributions which deviate in a statistically significant manner from the uniform distribution [50]. Such parameters are deemed to be the essential ones for the ER system to exhibit the behaviour associated with the viable set [50]. We perform this analysis by means of the Kolmogorov-Smirnov (KS) test [35, 29], which we use to compare our samples with the uniform distribution. According to such analysis, the kinetic constants k_1, k_3, k_6, k_7, k_{12}, k_{14}, and k_{16} are not uniformly distributed (p-values are reported in Table A.1 in Section A.1.2). Thus, only the values of these kinetic constants are constrained by the requirement that the system is mono-stable for $e_{HDM} = e_{HDAC} = 1$.

Nested within the viable subsensemble, there are parameter sets which exhibit plastic behaviour, as illustrated in the example shown in Fig. 2.5(f), i.e. with a region where the bistable regions of the differentiation and pluripotency ER systems overlap, which allows for facillitated reprogramming. We thus continue by studying the plastic subset regarding both its frequency within the viable subset and further restrictions imposed on parameter variability. We first check the number of plastic parameter sets within the viable set relative to the pluripotency-gene ER system defined by Table A.8 in Section A.1.3. Somehow unexpectedly, the plastic scenario is rare, but not exceptional: amongst the 94 parameter sets that we have identified as viable, 10 exhibit plasticity (see Fig. 2.7 for their CDFs).

Further restrictions on parametric heterogeneity imposed by the plastic scenario are analysed regarding the variation of the CDFs of kinetic constants when compared to those associated with the whole viable subset. The results of KS analysis performed on the data shown in Fig. 2.7 show that only the distributions of k_1 (associated with recruited demethylation), k_9 (unrecruited deacetylation), and k_{14} (recruited deacetylation) are significantly modified by the plasticity requirement (p-

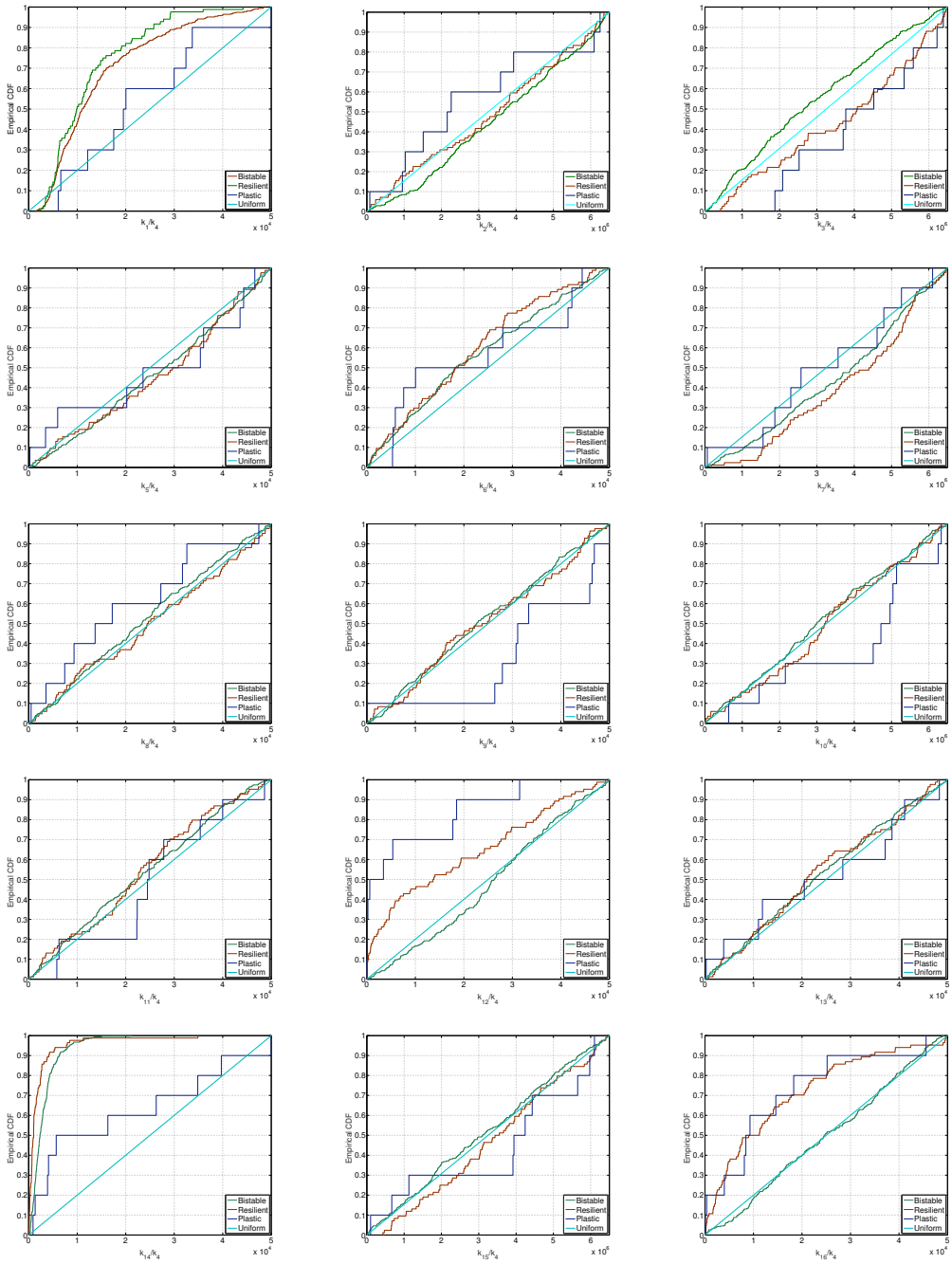


Figure 2.7: Caption on the following page.

Figure 2.7: This figure shows the cumulative distribution function(CDF) for a sample consisting of the 401 differentiation gene ER parameter sets generated by ABC which best fit the synthetic data shown in Fig. 2.4(a), i.e. SSA simulated data for the default stochastic ER differentiation system (see Table A.5 in Section A.1.3). Out of these 401 parameter sets, 94 satisfy the constraints associated with the differentiation epiphenotype. Amongst these, 10 are found to show plastic behaviour. The remaining 307 parameter sets generate bistability at $e_{HDM} = e_{HDAC} = 1$. Colour code: blue and red lines correpond to the CDF of the plastic and refractory differentiation epiphenotypes, respectively. Green lines correspond to the CDF of the parameters that generate bistability at $e_{HDM} = e_{HDAC} = 1$. Cyan lines correspond to the CDF of a uniform distribution, which we add for reference.

values reported in Table A.3 in A.1.2).

From a more mechanistic perspective, we observe that, within the plastic set, the mass of the CDFs of k_1, k_9 and k_{14} is displaced towards the large-value end of their intervals with respect to their behaviour within the full viable set. In other words, k_1, k_9 and k_{14} tend to be larger for plastic ER systems than for non-plastic, viable ER systems. In essence, we observe that ER systems exhibiting plastic behaviour tend to have increased activity in the enzymes performing histone deacetylation. This is consistent with recent evidence that ageing decreases histone acetylation and promotes reprogramming [175, 125, 115].

The same analysis has been conducted regarding the ensemble of parameter values generated for the pluripotency gene ER system (full posterior distribution in Fig. 2.3). The results of this analysis are shown in Fig. 2.8. Detailed analysis using the KS test of the ensemble viable pluripotency ER systems shows that k_3, k_8, k_{12}, k_{14}, k_{15}, and k_{16} are significantly constrained by the requirements of such scenario (i.e. their CDF departs significantly from the uniform distribution, as shown by the p-values from Table A.2 in A.1.2). We then move on to investigate further restrictions within the plastic set when compared to the viable pluripotency subset. We observe that only the CDFs associated with k_2 and k_6 are significantly different (p-values reported in Table A.4 in A.1.2). In both cases, values of k_2 and k_6 associated with plasticity are larger than in the general viable population. Both parameters are associated with demethylation activity.

Our ensemble analysis thus provides a rationale for the coupling between varia-

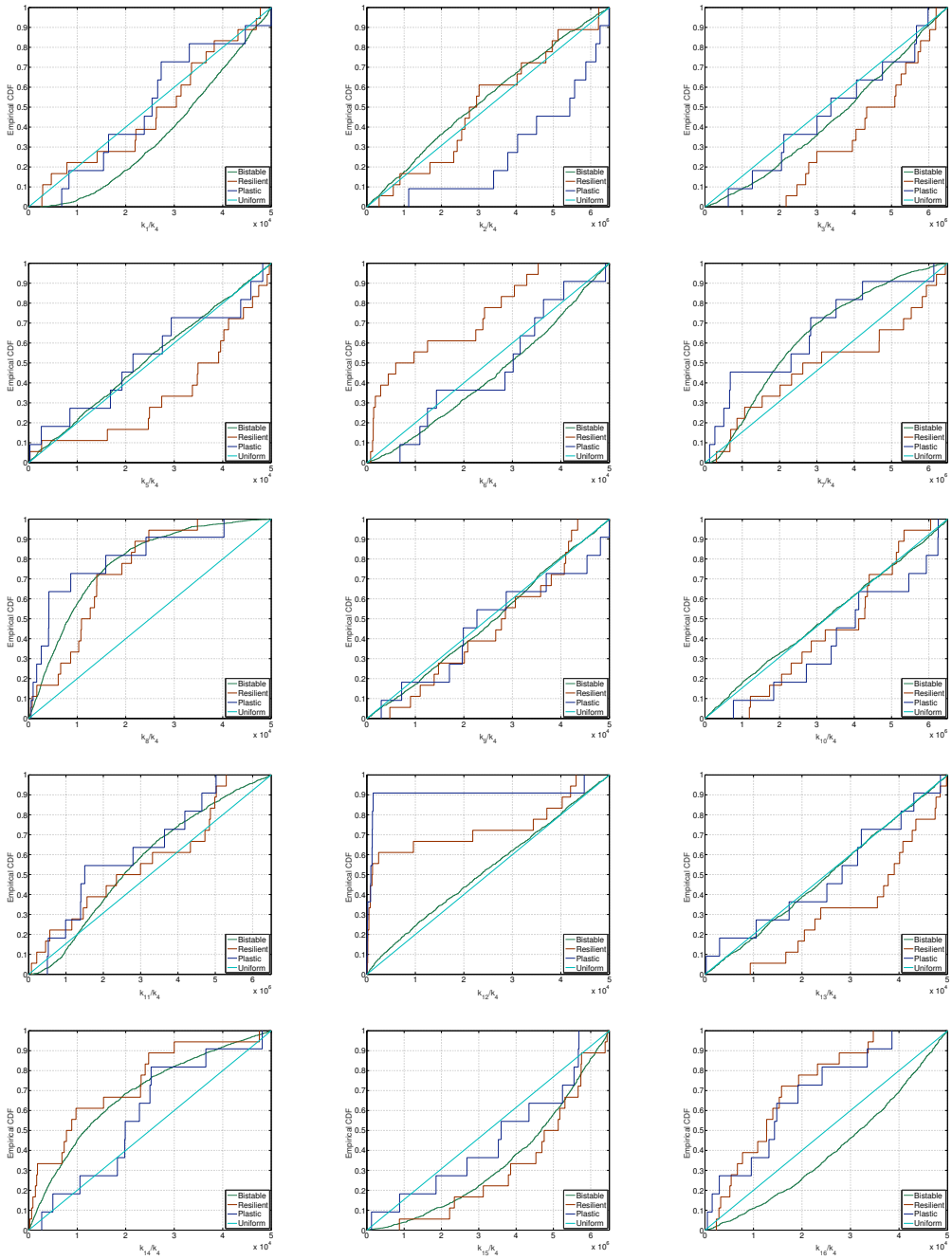


Figure 2.8: Caption on the following page.

Figure 2.8: This figure shows the cumulative distribution function (CDF) for a sample consisting of the 1401 pluripotency gene ER parameter sets generated by ABC which best fit the synthetic data shown in Fig. 2.4(b), i.e. SSA simulated data for the default stochastic ER pluripotency system (see Table A.6 in A.1.3). Out of these 1401 parameter sets, 29 satisfy the constraints associated with the pluripotency epiphenotype. Amongst these, 11 are found to show plastic behaviour. Another 1367 parameter sets generate bistability at $e_{HDM} = e_{HDAC} = 1$. The remaining 5 parameter sets are bistable at $e_{HDM} = e_{HDAC} = 1$ but they are rejected since their steady states do not correspond to open/closed situations. Colour code: blue and red lines correpond to the CDF of the plastic and refractory pluripotency epiphenotypes, respectively. Green lines correspond to the CDF of the parameters that generate bistability at $e_{HDM} = e_{HDAC} = 1$. Cyan lines correspond to the CDF of a uniform distribution, which we add for reference.

tions in the size of the pool of epigenetic cofactors and increased reprogramming in a heterogeneous cell population. A notable case in point is provided by metabolic changes during ageing: those cells where key metabolites such as acetyl-CoA and NAD+ are less abundant lose acetylation capability (in our model, this is reflected through the dependence of histone-modifying enzyme activity on the concentration of these cofactors), leading to cells poised for reprogramming.

This analysis allows us to design a strategy to interfere with the epigenetic regulatory system, regarding the ability to either drive the system away from plastic behaviour or to drive it to the plasticity scenario, while keeping it functional (i.e. within the restrictions of the base-line scenario). An example illustrating the effectiveness of this strategy is shown in Fig. 2.9. Consider the viable set of the ER differentiation-promoting gene, Fig. 2.7, which is neutral with respect to the value of k_9: k_9 remains uniformly distributed within the viable subset. By contrast, when plasticity is required, the admissible values of k_9 accumulate mostly towards the large-value end. This suggests that decreasing the value of k_9 might be a viable strategy to restore resilience. To check this, we consider the parameter set, $\theta = k_j/k_4,\ j = 1, \ldots, 16$, that gives rise to the plastic behaviour depicted in Fig. 2.5(f) (Table A.7 in Section A.1.3, for the differentiation-promoting gene). We then analyse the effect of modifying the value of k_9 for the differentiation-promoting gene on system behaviour. The new parameter set, $\theta' = k'_j/k_4,\ j = 1, \ldots, 16$, is such that $k'_9 = k_9/4$ and $k'_j = k_j$ for all $j \neq 9$ (k_j values as per Table A.7 in Section A.1.3). Parameter values for the pluripotency gene remain unchanged (as per Table

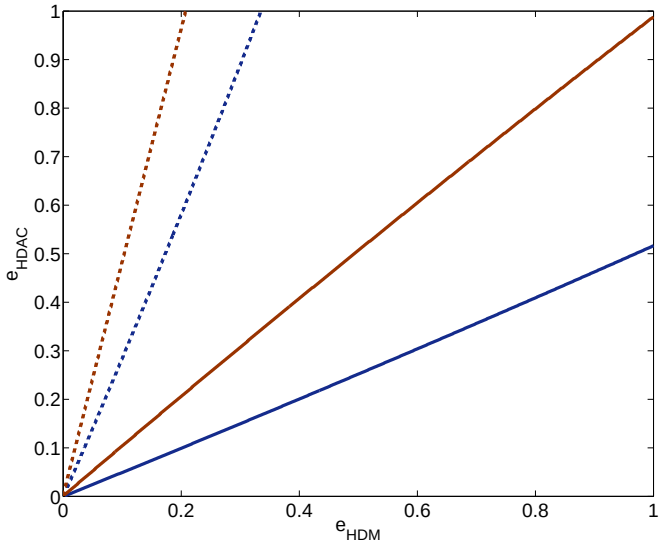
A.8 in Section A.1.3). The corresponding phase space is shown in Fig. 2.9(a). We observe that by reducing deacetylase activity in this fashion, the ER system reverts to resilient behaviour. This suggests that, by regulating the abundance of cofactors associated with (de)acetylation, we can drive the system off the plastic regime into the base-line behaviour.

Similarly, we can seek for complex, combined strategies to increase the robustness of plastic behaviour. An example of such strategy is shown in Fig. 2.9(b). Based on the results of the KS test for the differentiation-promoting gene, we observe that deacetylation-related rates k_9 and k_{14} are significantly increased in plastic scenarios. Taking parameter sets from a resilient scenario (Tables A.5 & A.8 in Section A.1.3, which lead to a combined phase diagram qualitatively similar to that shown in Fig. 2.5(d)) and modifying k_9 and k_{14} for the differentiation-promoting gene so that $k'_9 = 3k_9$ and $k'_{14} = 3k_{14}$ while keeping all the others at the same value, the resulting ER system corresponds to a plastic system. Futhermore, this combined strategy results in more robust plasticity (as compared to e.g. the case shown in Fig. 2.5(f)), as measured by the area of the phase space region where reprogramming is feasible. These results indicate that by combining the strategies suggested by the statistical analysis of the plastic sub-ensemble, we can find conditions for optimal conditions to achieve robust reprogramming. This, in turn, highlights the importance of cofactor levels, since as it has been shown in Fig. 2.9, depending on its availability, the same ER system can be driven to the plastic or resilient state.

These results are also consistent with the results of Pour et al. [129], who reported that variations in the activity of HMEs could be used to marshal the variability associated with epigenetic heterogeneity to alter the cellular response to reprogramming stimuli.

These strategies require close attention to be payed to the correlations between parameters. Parameters in complex systems biology models exhibit strong correlations which confer the system with essential properties such as *sloppiness*, which refers to the property exhibited by many multi-parameter models arising in systems biology, whereby the system's behaviour is insensitive to changes in parameter values except along a small number of parameter combinations [34]. In order to quantify such correlations, we have used hierarchical clustering. The results are shown in Fig. 2.10(a) & 2.10(b) for the base-line and the plastic scenarios of the differentiation-regulating ER system, respectively. Not unexpectedly, we observe that, with respect to the base-line scenario, correlations substantially change when the plastic scenario

(a)



(b)

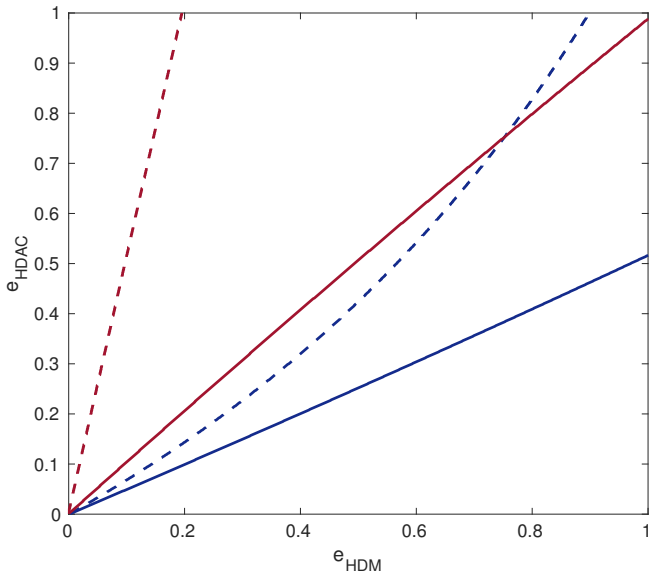


Figure 2.9: Caption on the following page.

Figure 2.9: (a)This plot shows results regarding restoration of base-line behaviour by removal of plasticity by restoring acetylation activity. It shows the phase space corresponding to the ER system composed of a differentiation-promoting gene with parameter set given by θ' with $k'_9 = k_9/4$ (see text for details) and a pluripotency-gene with parameters given by Table A.8 in Section A.1.3. This result demonstrates that by reducing deacetylation activity, we can drive the system off plastic behaviour and restore the normal situation as described by the base-line scenario. (b) This plot shows results regarding the appearance of the plastic behaviour by increasing deacetylation activity. Parameter values for the differentiation-promoting gene are given by θ' with $k'_9 = 3k_9$ and $k'_{14} = 3k_{14}$ (see text for details) and for the pluripotency-promoting gene are given by Table A.8 in Section A.1.3. This result shows an strategy to drive the system to the plastic scenario and hence, indicates how to obtain favourable scenarios for reprogramming.

is considered. Although the strategies illustrated in the results shown in Fig. 2.9 changed one or two parameters alone independently of all the others, more general situations will require to closely monitor these correlations to understand which combinations of parameters are relevant to control the system's behaviour [34].

2.4 Discussion

In this Chapter, we have provided computational evidence for the role of stochastic translation of epigenetic cofactors into resilient/plastic cell states via ER systems as a mechanistic facilitator of cellular ageing, and its reversal. When changes in levels of such cofactors operate as regulators of the kinetic parameters associated with chromatin-modifying enzymes such as HDMs and HDACs, the ensemble of ER configurations reveals the occurrence of cell-to-cell phenotypic variability in terms of different epi-states (see Fig. 2.11). This model provides a rationale for the responsiveness of cellular phenotypes to metabolic signals, as metabolic pools serve as epigenetic cofactors. The metabolic control of epigenetic landscapes and cell state transitions might therefore operate as a common hub capable of facilitating the pathogenesis of ageing-related diseases including cancer.

Several layers of molecular communication exist between cell metabolism and chromatin remodelling [180, 83, 113, 90, 56, 140]. A first layer of metabolo-epigenetic

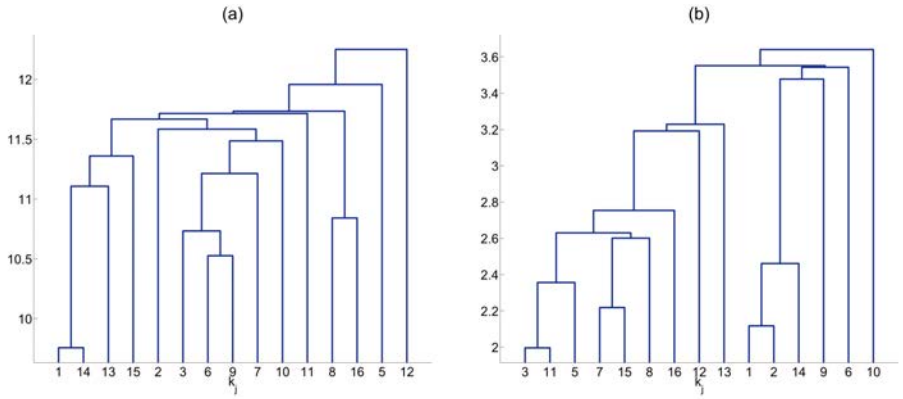


Figure 2.10: Plots showing hierarchical clustering analysis for the parameter sets that satisfy the base-line scenario (plot (a)) and the plastic scenario (plot (b)) of the differentiation-regulating ER system, respectively.

regulation includes metabolites/nutrient-responsive TF-dependent transcriptional regulation of chromatin regulators (HMT, HAT, etc.), which can lead to global changes on chromatin structure. Second, metabolites can modulate chromatin modifications at specific genomic loci by affecting the activity/localisation of proteins that recruit or regulate chromatin-modifying enzymes during, for example, transcriptional activation phenomena. Third, chromatin-modifying enzymes employ many metabolites as donor substrates and cofactors, and changes in levels of these *bona fide* epigenetic metabolites can in turn lead to changes not only in the global status of chromatin modifications but also to gene specific regulation under different metabolic conditions.

Our mathematical model only incorporates the third such layer through cofactor-induced heterogeneity. Because any metabolic input has the potential to affect various epigenetic marks via its effects on transcription, our model ignores metabolic regulation of TF activity. In contrast to other metabolically-regulated enzymatic activities such as phosphorylation in which the substrate (ATP) is present in cellular concentrations far greater than the enzyme K_m values, i.e., the concentration of metabolite at half maximum velocity of enzyme-mediated reaction, the physiological cellular concentrations of donors and cofactors that are employed by histone-modifying enzymes (e.g., organic ketoacids such as the demethylase cofactor α-ketoglutarate for HDMs or the NAD+ deacetylase cofactor for HDACs) are close to HDM and HDAC K_m values [113, 51]; consequently, based solely on the intrinsic biochemical character-

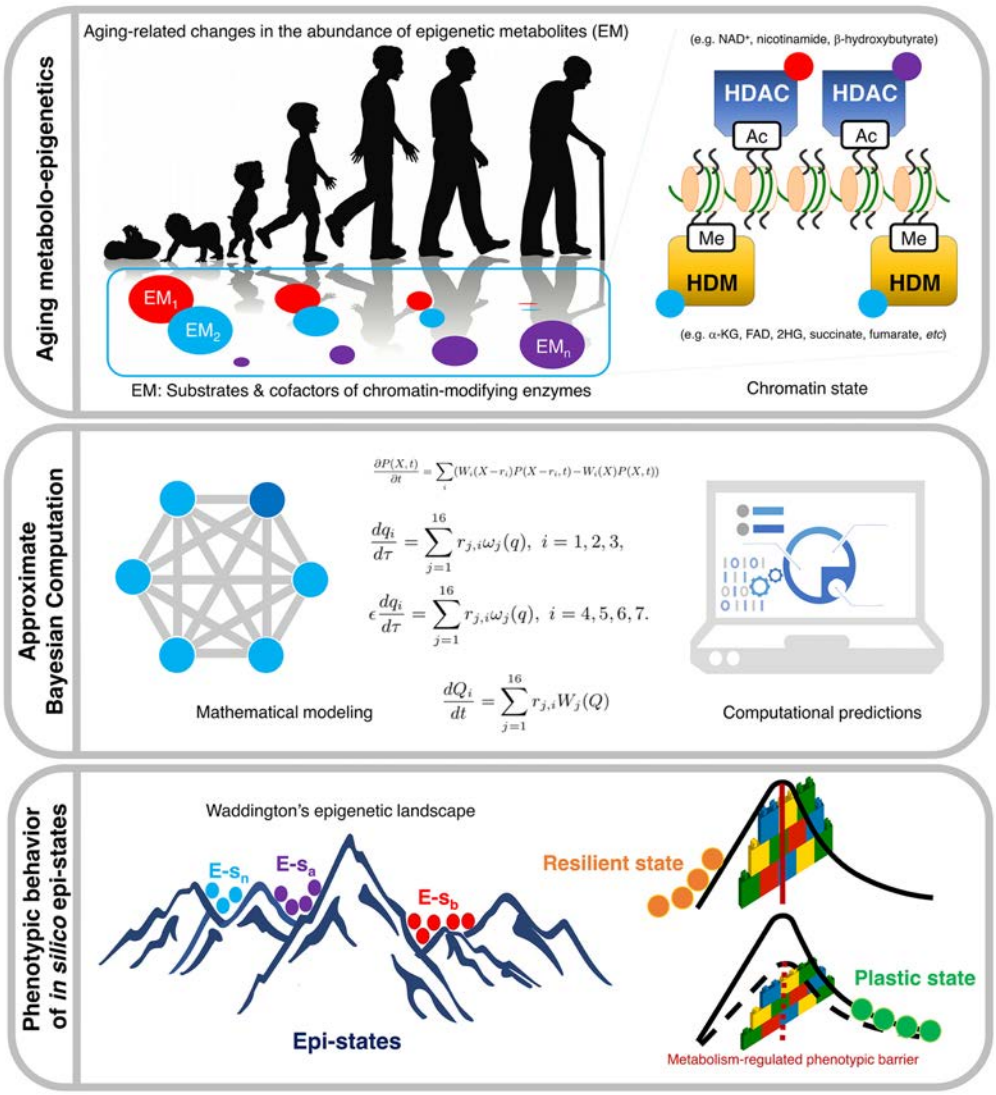


Figure 2.11: Caption on the following page.

Figure 2.11: (Previous page). Epigenetic regulation of cell fate reprogramming in ageing and disease: A predictive computational model. Cell reprogramming, a process that allows differentiated cells to re-acquire stem-like properties, is increasingly considered a critical phenomenon in tissue regeneration, ageing, and cancer. In light of the importance of metabolism in controlling cell fate, we designated a computational model capable of predicting the likelihood of cell reprogramming in response to changes in ageing-related epigenetic metabolites (EM). Our first-in-class Approximate Bayesian Computation (ABC) approach integrates the biochemical basis of ageing-driven metabolite interaction with chromatin-modifying enzymes to predict how ageing-driven metabolic reprogramming could alter cell state transitions via reorganisation of chromatin marks without affecting the shape of the Waddingtonian epigenomic landscape. Our predictive mathematical model improves our understanding of how pathological processes that involve changes in cell plasticity, such as tissue repair and cancer, might be accelerated or attenuated by means of metabolic reprogramming-driven changes on the height of phenotypic transitioning barriers.

istics of chromatin-modifying enzymes such as HDMs and HDACs, small fluctuations in the concentrations of such metabolites could significantly alter HDM and HDAC activities, either increasing or decreasing their respective histone-modifying activities. This layer of metabolo-epigenetic regulation is commonly viewed as a direct link from cell metabolism to chromatin-modification status, which could be mathematically modelled and tested as has been confirmed in our current model (see Fig. 2.11).

Evidence accumulates demonstrating that differing metabolomes can be found in distinct cell states, thus suggesting that changes in metabolism can impact and probably specify cell fate via alteration of the chromatin landscape [142, 141, 152, 147, 52, 136]. Yet, there is a scarcity of examples showing that metabolic changes can restructure the epigenetic landscape and lead to different cell states regardless of other global changes in cell physiology occurring in response to this variation in metabolite levels. Our findings support the notion that changes in the abundances of certain metabolites would alter specific chromatin marks, thereby determining both the stability of cell types and the probability of transitioning from one epi-state to another [130]. Our model infers that such a change in metabolite level would be sufficient to either impede or allow cell epi-state transitions by regulating the height of the phenotypic barriers in the context of Waddingtons landscape (Fig. 2.11). However, we should acknowledge that the necessary involvement of cellular metabolism on the structure of the epigenetic landscape will require the experimental coupling of

defined metabolic conditions with epigenome editing systems (e.g., CRISPR-Cas9) capable of targeting specific histone post-transcriptional modifications playing important roles in chromatin structure [74].

Our ensemble approach provides mechanistic support to the notion that emergence of the cellular and molecular hallmarks of ageing including cancer might result from a metabolically driven loss of epigenetic resilience. Flavahan et al. [51] have recently proposed that non-genetic stimuli including ageing and metabolic insults can induce either overly restrictive chromatin states, which can block tumor-suppression and/or differentiation programs, or overly permissive/plastic chromatin states, which might allow normal and cancer cells to stochastically activate oncogenic programs and/or nonphysiologic cell fate transitions. Our ensemble approach provides a framework that supports heterogeneity of epigenetic states as an engine that facilitates cancer hallmarks and other ageing diseases. On the one hand, the ability of resilient states to maintain large epigenetic barriers refractory to non-physiologic cell fate transitions might explain why the NAD+-dependent HDAC/sirtuin pathway is one of the few mechanisms described to mediate the correction or resetting of the abnormal chromatin state of ageing cells induced by calorie restriction, the most robust life span-extending and cancer preventing regimen [122, 96, 82, 49]. On the other hand, the ability of plastic states to lower epigenetic barriers, and increase the sensitivity of primed cells to undergo reprogramming-like events leading to loss of cell identity is consistent with the ability of certain metabolites to promote oncogenesis by epigenetically blocking the HDM-regulated acquisition of differentiation markers [178, 114, 119, 110].

The traditional view of cancer formation (i.e., the Knudson model [91]) exclusively involves the binary acquisition and accumulation of genetic alterations as the principal driver mechanism for the age-dependency of multistage cancer development. Our ensemble approach suggests an alternative, namely, that oncogenic chromatin aberrations might also occur via purely epigenetic stimuli, giving thus a plausible explanation for those cancers appearing without genetic mutations. Our model shows that, nested within the ensemble of ER systems, those that prime cells for reprogramming exhibit properties associated with age-induced epigenetic dysregulation [39, 108].

Furthermore, our model suggests the possibility of an ageing-progressive evolution from the resilient to the plastic cell state. As discussed, the plastic scenario depicted in Fig. 2.5(f) is characterised by reduced acetylation levels, which have

been linked to ageing [33, 133, 48, 125]. Therefore, it could occur that cells initially bearing a resilient epi-phenotype may switch to a plastic one as ageing progresses. This idea is in agreement with the results from Mosteiro et al. [115], where senescent cells, which appear within ageing tissues, are more likely to undergo reprogramming, i.e. they are in the plastic scenario.

Aging-responsive ER reprogramming might thus operate in a more progressive and graded manner to increase cancer susceptibility without the need to induce genetic mutations. Our ensemble model is mechanistically consistent with the fact that those cancers in which the sole presence of epigenetic metabolites (e.g., oncometabolites) suffices to stabilise undifferentiated cellular states by preventing demethylation of genes implicated in differentiation have accelerated models of oncogenesis. Accordingly, patients with glioma with gain-of-function isocitrate dehydrogenases (IDH) mutations generating the HDM oncometabolite/inhibitor 2-hydroxyglutarate, which establishes a hypermethylator phenotype that stabilizes undifferentiated cellular states that may be targetable and expanded later by transforming mutations, are, on average, several years younger that those with wild-type IDH gliomas [17, 12, 30, 128, 41]. Compared with sporadic forms, familial paragangliomas due to mutations in the succinate dehydrogenase complex and the consequent accumulation of the HDM inhibitor succinate, which also establishes a hypermethylator phenotype and the epigenetic silencing of key differentiation genes [95, 177], tend to present at a younger age [98, 28]. Moreover, a mosaic pattern of IDH mutation-bearing cells has been suggested to explain the occurrence of diverse and multiple tumors in some pediatric disorders [4, 123]. Whereas the epigenetic signature of adult somatic cells must be partially and acutely erased to adopt a more plastic epigenome, such cellular plasticity, which might occur via metabolically driven epigenetic activation of promoter regions of pluripotency genes, could impose a chronic, locked gain of stem cell-like states disabled for reparative differentiation. In this sense, it has been coined the term metabostemness, which refers to the metabolic control of the epitranscriptional orchestration of reprogramming that redirects normal and tumor cells towards less-differentiated cellular states, stem cell-like states in most cases [107].

The existence of metabolism-permissive resilient and plastic epigenetic landscapes might have predictive power on the susceptibility of a cell to lose its normal cellular identity through reprogramming-like resetting phenomena. The beneficial or deleterious decision paths during the maintenance of cell and tissue homeostasis might be closely related to the ability of epigenetic landscapes to modulate the intrinsic responsiveness to reprogramming cellular identity. The incapability of finishing

cellular reprogramming, or at least to increase cellular epigenetic plasticity, might impede tissue self-repair in response to injury, stress, and disease, thus driving the observed ageing phenotypes. Accordingly, the infliction of chronic injury and the ageing phenotype have been shown to render tissues highly permissive to in vivo reprogramming [115] while the cyclic, transient expression of reprogramming factors has recently been shown to increase lifespan in a murine model of premature ageing via remodeling of the chromatin landscape [121]. Because our model suggests that the fine-tuning of metabolic epigenetic cofactors might direct plastic epigenetic states to re-enter into epigenetic resilience, and vice versa, it would be relevant to experimentally evaluate whether specific metabolic interventions might either mimic transient reprogramming and revert some age-associated features without promoting complete undifferentiation, or prevent the occurrence of unrestricted/uncontrolled plasticity in chronically injured tissues such as those occurring in ageing and cancer.

In summary, by integrating the ability of chromatin epigenetic modifiers to function as sensors of cellular metabolism, our ensemble model provides computational support to the notion that a metabolism-responsive loss of epigenetic resilience might mechanistically facilitate cellular ageing. The stochastic translation of metabolic signals into resilient/plastic cell states via ER systems might be viewed as a metabolo-epigenetic dimension that not only facilitates cellular ageing, but that also offers new therapeutic and behavioural avenues for its reversal. Our findings strongly suggest that the development of predictive mathematical models and computational simulation platforms capable of operatively integrate the metabolic control of epigenetic resilience and plasticity and its combination with confirmatory lab-based testing might accelerate the discovery of new strategies for metabolically correcting the aberrant chromatin structure that affects cellular identity and epi-state transitions in ageing and ageing-related diseases.

Chapter 3

Unlocking the pluripotent phenotype: A multiscale model of the epigenetic regulation of cell fate and plasticity

3.1 Summary

In this Chapter, we introduce a stochastic model of combined epigenetic regulation (ER)-gene regulatory network (GRN) to study the effects of epigenetic plasticity, caused by the ER heterogeneity, on cell-fate determination. By adding ER to the picture, the model presented in this Chapter extends previous approaches where phenotypes are associated with the attractors of complex gene regulatory systems and their robustness, as well as with the resilience of such attractors in the presence of intrinsic noise, environmental fluctuations, and other disturbances [79, 155, 181, 58, 170, 127, 97]. Furthermore, based on the existence of multiple scales, we are able to formulate a method for stochastic model reduction, from which we derive an efficient hybrid simulation scheme that allows us to deal with the combined ER-GRN model. Our analysis of the coupled system reveals a regime of tristability in which pluripotent stem-like and differentiated steady-states coexist with a third indecisive state. Crucially, ER heterogeneity of differentiation genes is for the most part responsible for conferring abnormal robustness to pluripotent stem-like states. We then formulate epigenetic heterogeneity-based strategies capable of unlocking and facilitating the transit from differentiation-refractory (pluripotent stem-like) to differentiation-primed epistates. The application of the hybrid numerical method val-

idated the likelihood of such switching involving solely kinetic changes in epigenetic factors. Our strategies allow to unlock persistent states of pathological pluripotency, suggesting that epigenetic heterogeneity regulates the mechanisms and kinetics of phenotypic robustness of cell fate reprogramming.

3.2 Model formulation and analysis

In this Chapter, we aim to study a ER-GRN model which can describe cell differentiation and cell reprogramming. One of the simplest GRNs which allows to do this consists of two genes, one promoting differentiation, and the other promoting pluripotency (see Fig. 3.1(a)). Nevertheless, in this Section (Section 3.2), we formulate our model and we analyse it considering the most generic case, i.e. we assume to have an arbitrary number of genes N_G. By doing so, our theoretical analysis can be further applied to any ER-GRN model, which implies a wide applicability of the derived formulation. However, when possible, we try to relate the theory developed to our particular ER-GRN so as to keep track of our case study.

3.2.1 General description of the stochastic model of an epigenetically-regulated gene network

Consider a gene regulatory network composed of N_G self-activating genes which can repress each other. In particular, we consider that the gene product of each of these genes forms homodimers, which act as a transcription factor (TF) for its own gene by binding to its own promoter. Furthermore, each gene within the network has a number of inhibitors, which operate via competitive inhibition: the homodimers of protein j bind to the promoter of gene i, and by doing so they impede access of the TF to the promoter of gene i. In Fig. 3.1(a), an illustrative scheme of the simplified case of two mutually inhibiting genes, one promoting pluripotency (blue) and one promoting differentiation (green), is shown. The regulation topology of the network can be represented using a weighted adjacency matrix B. B is a $N_G \times N_G$ matrix, whose elements, $b_{ij} > 0$, are the binding rates of homodimers of protein j to the promoter of gene i (see Fig. 3.1(a)). Moreover, the expression of gene i is induced at a constant basal production rate, $\hat{R}_i$, independent of the regulatory mechanism described above. Proteins (TF monomers) of type i are synthesised at a rate proportional to the number of bound promoter sites with rate constant k_{i1} and degraded with degradation rate k_{i2} (see Fig. 3.1(a), Tables 3.1 and 3.2).

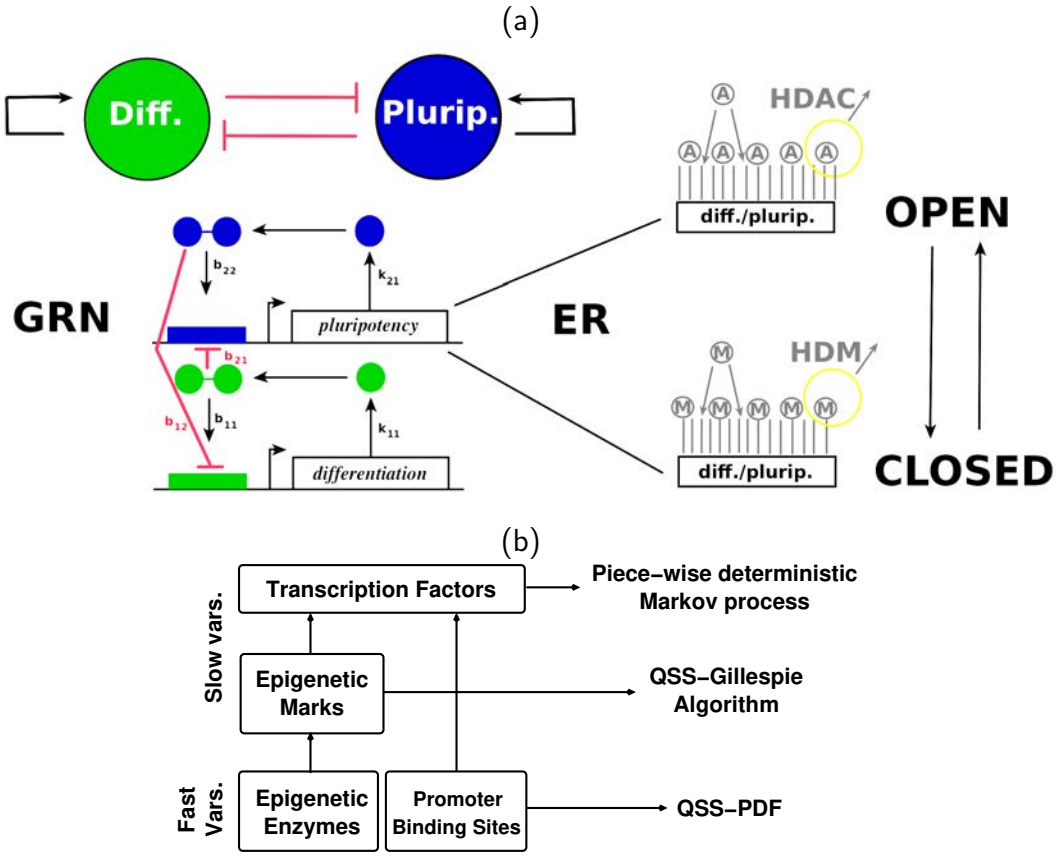


Figure 3.1: Schematic reprentation of the ER-GRN model and its multiscale reduction. (a): Gene regulatory network (GRN) of two self-activating, mutually-inhibitory genes with epigenetic regulation. In the GRN model, the gene product (denoted by X_i in Table 3.2) is its own transcription factor which, upon dimerisation, binds the promoter region of the gene thus triggering gene transcription. The transition rates corresponding to this GRN are given in Table 3.2. For simplicity, we use an effective model in which the formation of the dimer and binding to the promoter region is taken into account in a single reaction, and the resulting number of promoter sites bound by two transcription factors is denoted by X_{ij} (see Table 3.2). Furthermore, depending on whether the epigenetic state is open (i.e. predominantly acetylated (A)) or closed (i.e. predominantly methylated (M)) the promoter region of the gene is accessible or inaccessible to the transcription factor, respectively. (b): Schematic representation of the time separation structure of the multiscale method developed to simulate the ER-GRN system. See Section 3.2.2 for more details.

Table 3.1: Brief description of the variables and parameters involved in the symmetric model of gene regulatory network with competitive binding inhibition, with i denoting any gene of the GRN, i.e. $i = 1, \ldots, N_G$. For simplicity, we will assume that $\mathcal{O}(E) = \mathcal{O}(Y)$.

Variable	Description
N_G	Number of genes
S	Characteristic scale (average) for protein number
E	Characteristic scale (average) for binding site number in the promoter regions
Y	Characteristic scale (average) for binding site number for epigenetic marks
Z	Characteristic scale (average) for epigenetic enzyme number
e_i	Number of binding sites in the promoter region of gene i
e_{HDM}	Number of HDM molecules
e_{HDAC}	Number of HDAC molecules
$\hat{R}_i$	Basal rate of induction of gene i
k_{i1}	Rate of transcription of gene i
k_{i2}	Degradation rate of protein of type i
b_{ij}	Binding rate of protein j homodimers onto the promoter region of gene i
u_{ij}	Unbinding rate of protein j homodimers from the promoter region of gene i
c_{ij}	Kinetic rate of the jth reaction from the ER system of gene i (see Table 3.3)
X_i	Number of transcription factor monomers of type i
X_{ij}	Number of sites of the promoter region of gene i bound to a protein j homodimer
Y_{i1}	Number of unmodified(U) nucleosomes in gene i (ER model)
Y_{i2}	Number of methylated(M) nucleosomes in gene i (ER model)
Y_{i3}	Number of acetylated(A) nucleosomes in gene i (ER model)
Y_{i4}	Number of free HDM enzyme molecules associated with ER system of gene i
Y_{i5}	Number of methylated nucleosome-HDM enzyme complexes in gene i (ER model)
Y_{i6}	Number of free HDAC enzyme molecules associated with ER system of gene i
Y_{i7}	Number of acetylated nucleosome-HDAC enzyme complexes in gene i (ER model)
$x_i = \frac{X_i}{S}$	Re-scaled number of X_i (slow GRN variables)
$x_{ij} = \frac{X_{ij}}{E}$	Re-scaled number of X_{ij} (fast GRN variables)
$y_{ij} = \frac{Y_{ij}}{Y}$	Re-scaled ER slow variables, $j = 1, 2, 3$, for gene i
$y_{ij} = \frac{Y_{ij}}{Z}$	Re-scaled ER fast variables, $j = 4, 5, 6, 7$, for gene i

In addition to TF regulation, we further consider that each gene is under epigenetic regulation (ER). ER controls gene transcription by modulating access of TFs to the promoter regions of the genes (see Fig. 1.4). In other words, in our model, ER is associated with an upstream drive that regulates gene expression [36]. Such epigenetic control is often related to alternative covalent modifications of histones. To address the high complexity of ER, we focus on a simpler stochastic model of ER. the one presented in Section 2.2.1. Just as a recall, our model belongs to a

Table 3.2: Transition rates associated with the stochastic dynamics of the GRN. Note that the rate of binding of homodimers to the promoter region of gene i is modulated by the level of acetylation of gene i, Y_{i3}: if Y_{i3} is above the threshold Y_0, gene i is *open*, i.e. the promoter is accessible to homodimers and TFs. By contrast, if the gene's acetylation levels decay, gene i is *silenced* and the promoter is inaccessible to gene-transcription regulatory dimers.

Transition rate	Event $(i, j = 1, \ldots, N_G)$
$W_{i1}(X) = \hat{R}_1 + k_{i1} X_{ii}$	$X_i \to X_i + 1$
$W_{i2}(X) = k_{i2} X_i$	$X_i \to X_i - 1$
$W_{i3}(X) = b_{ij} H(Y_{i3} - Y_0) \left(e_i - \sum_{k=1}^{N_G} X_{ik} \right) X_j (X_j - 1)$	$X_j \to X_j - 2$, $X_{ij} \to X_{ij} + 1$
$W_{i4}(X) = u_{ij} X_{ij}$	$X_j \to X_j + 2$, $X_{ij} \to X_{ij} - 1$

wider class of models which consider that single unmodified (U) chromatin loci can be modified so as to acquire positive (A) or negative (M) marks. These positive and negative marks involve covalent modification of histones. Of such modifications we consider methylation (associated with negative marks) and acetylation (associated with positive marks) [42]. An illustrative example on how epigenetic modifications, acetylation and methylation, alter the accessibility of TFs to the promoter regions of the genes is shown in Fig. 3.1(a), where the open and closed state denote the accessibility or inaccessibility, respectively, of TFs to the promoter region of a determined gene. Both modifications are mediated by associated enzymes: histone methyltransferases (HMTs) and demethylases (HDMs), and histone acetyltransferases (HACTs) and deacetylases (HDACs). For simplicity, and similarly to Chapter 2, we only explicitly account for HDM and HDAC activity (see Fig. 3.1(a)). In our model, a positive feedback mechanism is introduced whereby M marks help to both add more M marks and remove A marks from neighbouring loci (recruited mechanism). The positive marks are assumed to be under the effects of a similar positive reinforcement mechanism [42, 131]. A full description of the details of the ER model are given in Section 2.2.1. Table 3.3 provides the transition rates for the ER model for a GRN of an arbitrary number of genes.

Under suitable conditions, determined by the activity and abundance of histone-modifying enzymes and co-factors, the positive reinforcement mechanism produces robust bistable behaviour. In this bistable regime, the two possible ER stable states are: an *open* epigenetic state where the levels of positive (negative) marks are elevated (downregulated). In this case, the promoter of the gene is accessible to TFs and

Table 3.3: This table shows the transition rates associated with the stochastic dynamics of the epigenetic regulatory system of gene i. The random variables Y_{ik} are defined in Table 3.1. The different modification reactions are assumed to be of two types, recruited and unrecruited. Details regarding the assumptions between this distinction as well as a full description of the formulation of the stochastic epigenetic regulation model are given in Section 2.2.1. See Table 2.2 for the interpretation of each transition rate ($W_j = V_{ij}$, $j = 1, \ldots, 16$).

Transition rate	Reaction change vector
$V_{i1} = c_{i1} Y_{i2} Y_{i4}$	$r_{E_{i_1}} = (0, -1, 0, -1, +1, 0, 0)$
$V_{i2} = c_{i2} Y_{i5}$	$r_{E_{i_2}} = (0, +1, 0, +1, -1, 0, 0)$
$V_{i3} = c_{i3} Y_{i5}$	$r_{E_{i_3}} = (+1, 0, 0, +1, -1, 0, 0)$
$V_{i4} = c_{i4} Y_{i2} Y_{i3} Y_{i4}$	$r_{E_{i_4}} = (0, -1, 0, -1, +1, 0, 0)$
$V_{i5} = c_{i5} Y_{i3} Y_{i5}$	$r_{E_{i_5}} = (0, +1, 0, +1, -1, 0, 0)$
$V_{i6} = c_{i6} Y_{i3} Y_{i5}$	$r_{E_{i_6}} = (+1, 0, 0, +1, -1, 0, 0)$
$V_{i7} = c_{i7} Y_{i1}$	$r_{E_{i_7}} = (-1, +1, 0, 0, 0, 0, 0)$
$V_{i8} = c_{i8} Y_{i1} Y_{i2}$	$r_{E_{i_8}} = (-1, +1, 0, 0, 0, 0, 0)$
$V_{i9} = c_{i9} Y_{i3} Y_{i6}$	$r_{E_{i_9}} = (0, 0, -1, 0, 0, -1, +1)$
$V_{i10} = c_{i10} Y_{i7}$	$r_{E_{i_{10}}} = (0, 0, +1, 0, 0, +1, -1)$
$V_{i11} = c_{i11} Y_{i7}$	$r_{E_{i_{11}}} = (+1, 0, 0, 0, 0, +1, -1)$
$V_{i12} = c_{i12} Y_{i3} Y_{i2} Y_{i6}$	$r_{E_{i_{12}}} = (0, 0, -1, 0, 0, -1, +1)$
$V_{i13} = c_{i13} Y_{i7} Y_{i2}$	$r_{E_{i_{13}}} = (0, 0, +1, 0, 0, +1, -1)$
$V_{i14} = c_{i14} Y_{i7} Y_{i2}$	$r_{E_{i_{14}}} = (+1, 0, 0, 0, 0, +1, -1)$
$V_{i15} = c_{i15} Y_{i1}$	$r_{E_{i_{15}}} = (-1, 0, +1, 0, 0, 0, 0)$
$V_{i16} = c_{i16} Y_{i1} Y_{i3}$	$r_{E_{i_{16}}} = (-1, 0, +1, 0, 0, 0, 0)$

transcription can occur. By contrast, in the absence (abundance) of positive (negative) marks the gene is considered to be *silenced*, as TFs cannot reach the promoter.

An essential part of the stochastic dynamics of the ER system is the noise-induced transitions between the open and silenced states. Escape from steady states is a well-established phenomenon (see e.g. [57]) and thoroughly analysed within the theory of rate processes [69] and large deviation theory [54, 158, 127]. As we will illustrate below, these noise-induced dynamics are essential to classify the epiphenotypes of somatic cells [53] and stem cells and unravel the mechanisms of reprogramming and locking.

3.2.2 Multi-scale analysis and model reduction

The system that results from coupling the ER and GRN models becomes rather cumbersome and computationally intractable as the GRN grows. For this reason, in order to analyse the behaviour of the resulting stochastic model, we take advantage of intrinsic separation of time scales [25, 26, 10, 5, 84, 1, 85, 40]. We exploit this time scale separation to reduce our model by performing stochastic quasi-steady state approximations (QSSA) by means of asymptotic analysis of the stochastic ER-GRN system as established in [10, 5, 84, 85] (see Fig. 3.1(b)). Especifically, we assume that the characteristic scale for the number of TF monomers (S), the number of promoter binding sites (E), the number of ER modification sites (Y), and the number of ER enzymes (Z), are such that $S \gg E, Y \gg Z$ and $\mathcal{O}(E) = \mathcal{O}(Y)$ (see Table 3.1 for the definition of these variables). Note that the assumption $Y \gg Z$ is exactly the Briggs-Haldane hypothesis for enzyme kinetics [88] since the ER modification sites are the substrates for the ER enzymes (see Section A.2.1.1). The multiscale analysis is carried out in detail in Section A.2.1.1-A.2.1.4. We show that, upon appropriate assumptions regarding the characteristic scales of the different molecular species, our model exhibits a hierarchy of time scales, which allows to simplify the model and its computational simulation.

The resulting reduced stochastic model is such that, since $S \gg E$ and $Y \gg Z$, the number of bound-to-promoter TFs and ER enzyme-substrate complexes are fast variables that can be sampled from their quasi-equilibrium distribution with respect (or conditioned to) their associated slow variables. TFs and ER modification sites (ER substrates) are slow variables whose dynamics, which dominate the long time behaviour of the system, are given by their associated stochastic dynamics with the fast variables sampled from their quasi-steady state approximation (QSSA) probability density functions (PDFs). The assumption that $S \gg Y$ allows for further simplification of the model, as it allows to take the limit of $S \gg 1$ in the stochastic equations for the TFs monomers which leads to a *piece-wise deterministic* Markov description: the dynamics of the number of TFs monomers is given by an ODE which is perturbed at discrete times by a noise source [84].

We present a summarised version of the asymptotic model reduction. Details of this analysis are provided in Section A.2.1.2. The starting point of our analysis is the so-called Poisson representation of the stochastic process, which is equivalent to the Master Equation, [5]:

$$X_i(t) = X_i(0) + \sum_{k=1}^{R_G} r_{G_{ik}} \mathcal{P}\left(\int_0^t W_{ik}(\mathbf{X}(s); \mathbf{Y}_i(s))ds\right) \tag{3.2.1}$$

$$X_{ij}(t) = X_{ij}(0) + \sum_{k=1}^{R_G} r_{G_{ijk}} \mathcal{P}\left(\int_0^t W_{ik}(\mathbf{X}(s); \mathbf{Y}_i(s))ds\right) \tag{3.2.2}$$

$$Y_{ij}(t) = Y_{ij}(0) + \sum_{k=1}^{R_E} r_{E_{ijk}} \mathcal{P}\left(\int_0^t V_{ik}(\mathbf{Y}_i(s))ds\right) \tag{3.2.3}$$

where X_i denotes the product of gene i, X_{ij} refers to the number of dimers of type j bound to the promoter region of gene i and Y_{ij} corresponds to the number of molecular species of type j within the ER model of gene i (see Table 3.1). We also use the notation $\mathbf{X} = (X_1, \ldots, X_{N_G}, X_{11}, \ldots, X_{N_G N_G})$ and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i7})$. $\mathcal{P}(\lambda) \sim \text{Poisson}(\lambda)$, i.e. $\mathcal{P}(\lambda)$ is a random number sampled from a Poisson distribution with parameter λ [5], R_G and R_E denote the total number of reactions in the GRN model (see Table 3.2) and in the ER model (Table 3.3), respectively, with W_{ik} and V_{ik} denoting the transition rates corresponding to the GRN model and the ER model (see Table 3.2 and 3.3, respectively). The stoichiometries $r_{G_{ik}}$, $r_{G_{ijk}}$ and $r_{E_{ijk}}$ denote the change in number of molecules that reaction k has on X_i, X_{ij} and Y_{ij}, respectively. Eqs. (3.2.1) and (3.2.2) are associated with the the stochastic dynamics of the GRN (see Table 3.2), which are regulated by the ER part of the model. Eq. (3.2.3) describes the dynamics of the ER system, which drives the dynamics of the GRN (see Tables 3.2 and 3.3).

Under the appropriate conditions, separation of time scales can be made explicit by re-scaling the random variables and the transition rates. Based on previous works [1, 40, 53], we propose the following rescaling:

$$\begin{aligned}
&X_i = Sx_i,\ X_{ij} = Ex_{ij},\\
&Y_{ij} = Yy_{ij},\ \text{for } j = 1,2,3,\ Y_{ij} = Zy_{ij},\ \text{for } j = 4,5,6,7\\
&W_{ik}(\mathbf{X}; \mathbf{Y}_i) = b_{11} E S^2 w_k(\mathbf{x}; \mathbf{y}_i),\ V_{ik}(\mathbf{Y}_i) = c_{14} Z Y^2 v_{ik}(\mathbf{y}_i).
\end{aligned} \tag{3.2.4}$$

In Eq. (3.2.4), the scale factors S, E, Y, and Z, as described in Table 3.1, are the characteristic number of protein transcripts, promoter region binding sites, histone modification sites, and epigenetic enzymes (HDMs and HDACs), respectively. For

simplicity, we assume that these scales are the same for all the genes involved in the GRN. As mentioned, we assume that $S \gg E \simeq Y \gg Z$. We further define a re-scaled (dimensionless) time: $\tau = b_{11}ESt$. After using Eq. 3.2.4 as well as the re-scaled time, the resulting re-scaled parameters for the GRN model (Table 3.2) and the ER model (Table 3.3) are given in Table 3.4.

Table 3.4: Re-scaled GRN and ER parameters (see Section A.2.1.2)

Rescaled variables	Dimensionless parameters
$\tau = b_{11}EST$	$\epsilon_1 = \frac{E}{S}$ $\quad R = \frac{\hat{R}}{b_{11}ES^2}$
$x_i \equiv q_i = \frac{X_i}{S}$	$\omega_{i1} = \frac{k_{i1}}{b_{11}S^2}$ $\quad \omega_{i2} = \frac{k_{i2}}{b_{11}ES}$
$x_{ij} \equiv q_{ij} = \frac{X_{ij}}{E}$	$\beta_{ij} = \frac{b_{ij}}{b_{11}}$ $\quad \delta_{ij} = \frac{u_{ij}}{b_{11}S^2}$
$y_{ij} = Y_{ij}/Y$, $j = 1, 2, 3$	$\kappa_{i1} = \frac{c_{i1}}{c_{14}E}$, $\kappa_{i2} = \frac{c_{i2}}{c_{14}E^2}$, $\kappa_{i3} = \frac{c_{i3}}{c_{14}E^2}$
$y_{ij} = Y_{ij}/Z$, $j = 4, \ldots, 7$	$\kappa_{i5} = \frac{c_{i5}}{c_{14}E}$, $\kappa_{i6} = \frac{c_{i6}}{c_{14}E}$, $\kappa_{i7} = \frac{c_{i7}}{c_{14}EZ}$
	$\kappa_{i8} = \frac{c_{i8}}{c_{14}Z}$, $\kappa_{i9} = \frac{c_{i9}}{c_{14}E}$, $\kappa_{i10} = \frac{c_{i10}}{c_{14}E^2}$
	$\kappa_{i11} = \frac{c_{i11}}{c_{14}E^2}$, $\kappa_{i12} = \frac{c_{i12}}{c_{14}}$, $\kappa_{i13} = \frac{c_{i13}}{c_{14}E}$
	$\kappa_{i14} = \frac{c_{i14}}{c_{14}E}$, $\kappa_{i15} = \frac{c_{i15}}{c_{14}EZ}$, $\kappa_{i16} = \frac{c_{i16}}{c_{14}Z}$

By using Eq. (3.2.4) in Eqs. (3.2.1)-(3.2.3), we obtain:

$$x_i(\tau) = x_i(0) + \sum_{k=1}^{R_G} r_{G_{ik}} \frac{1}{S} \mathcal{P}\left(S \int_0^\tau w_k(\mathbf{x}(\sigma); \mathbf{y}_i(\sigma)) d\sigma \right) \text{ (slow)} \tag{3.2.5}$$

$$x_{ij}(\tau) = x_{ij}(0) + \sum_{k=1}^{R_G} r_{G_{ijk}} \frac{1}{E} \mathcal{P}\left(E \frac{1}{\epsilon_1} \int_0^\tau w_k(\mathbf{x}(\sigma); \mathbf{y}_i(\sigma)) d\sigma \right) \text{ (fast)} \tag{3.2.6}$$

$$y_{ij}(\tau) = y_{ij}(0) + \sum_{k=1}^{R_E} r_{E_{ijk}} \frac{1}{Y} \mathcal{P}\left(Y \frac{1}{\epsilon_2} \int_0^\tau v_{ik}(\mathbf{y}_i(\sigma)) d\sigma \right), \; j = 1, 2, 3 \text{ (slow)} \tag{3.2.7}$$

$$y_{il}(\tau) = y_{il}(0) + \sum_{k=1}^{R_E} r_{E_{ilk}} \frac{1}{Z} \mathcal{P}\left(Z \frac{1}{\epsilon_2} \frac{1}{\epsilon_3} \int_0^\tau v_{ik}(\mathbf{y}_i(\sigma)) d\sigma \right), \; l = 4, 5, 6, 7 \text{ (fast)} \tag{3.2.8}$$

where $\epsilon_1 = \frac{E}{S} \ll 1$, $\epsilon_2 = \frac{b_{11}S}{c_{14}Z}$, and $\epsilon_3 = \frac{Z}{Y} \ll 1$, with $\epsilon_1 < \epsilon_3$. We have no direct information to estimate the order of magnitude of ϵ_2. Thus, without loss of generality

we will assume that $\epsilon_2 = \mathcal{O}(1)$.

The scaling hypothesis $S \gg E \simeq Y \gg Z$ allows for a series of successive approximations which enables us to reduce the model Eqs. (3.2.5)-(3.2.8) into a much less computationally demanding system. First, provided that both $\epsilon_1 \ll 1$ and $\epsilon_3 \ll 1$, we can assume that the (rescaled) rates associated with the fast variables GRN-ER dynamics (Eqs. (3.2.6) and (3.2.8)) are much larger than those corresponding to their slow counterparts (Eqs. (3.2.5) and (3.2.7)). Under these conditions, the stochastic dynamics of the fast variables reaches their (quasi-)steady states while the slow variables are effectively frozen [25, 26, 10, 143].

We proceed with the asymptotic model reduction by first addressing the QSSA PDFs of the fast variables (see *Inner solution* below). We then move on to study the QSS approximation of the slow variables, in particular, the large-S asymptotics of the protein concentration dynamics (see *Outer solution* below).

3.2.2.0.1 Inner solution

The inner solution corresponds to the relaxation dynamics of the fast variables onto their quasi-equilibrium state, while the slow variables remain unchanged. The solution of the inner dynamics allows us to determine the QSSA PDFs of the fast variables conditioned to fixed values of the slow variables.

We proceed by considering the following rescaling of the time variable $T = \epsilon_1^{-1}\tau$. Upon such rescaling, it is straightforward that all the rates of the reactions affecting the slow variables (Eqs. (3.2.5) and (3.2.7)) are now $\mathcal{O}(\epsilon_1)$, which implies that the slow variables, x_i and y_{ij} (for $j = 1, 2, 3$), can be considered to remain frozen whilst the fast variables reach their quasi-equilibrium distribution according to the dynamics:

$$x_{ij}(T) = x_{ij}(0) + \sum_{k=1}^{R_G} r_{G_{ijk}} \frac{1}{E} \mathcal{P}\left(E \int_0^T w_k(\mathbf{x}(\sigma); \mathbf{y}_i(\sigma)) d\sigma \right) \tag{3.2.9}$$

$$y_{il}(T) = y_{il}(0) + \sum_{k=1}^{R_E} r_{E_{ilk}} \frac{1}{Z} \mathcal{P}\left(Z \frac{c_{14}\epsilon_1}{b_{11}\epsilon_3} \int_0^T v_{ik}(\mathbf{y}_i(\sigma)) d\sigma \right), \tag{3.2.10}$$

where $l = 4, 5, 6, 7$, $\epsilon_2 = \mathcal{O}(1)$ and the slow variables, x_i and y_{ij}, $j = 1, 2, 3$, are considered to stay constant.

Consider the (inner) dynamics of the number of bound sites within the promoter regions, $X_{ij}(T)$, Eq. (3.2.9). Provided that the ER of gene i remains in the open state, i.e. $y_{i2} \ll 1$ and $y_{i3} \sim \mathcal{O}(1)$ (abundance of positive marks), the resulting stochastic dynamics describes how the binding sites switch between bound-to-TF dimer to unbound-to-TF dimer at constant rates (since the number of the different TF molecules does not change at this time scale). Since the number of binding sites is a constant, the (quasi-)steady state distribution of bound TFs to each promoter is a multinomial (see Section 3.2 of the Supplemental Information for a detailed derivation of this result). Otherwise, if gene i is epigenetically closed, then $X_{ij}(T) = 0$ for all j with probability one. Therefore, the random vector describing the number of TFs bound at the promoter region of gene i, $\mathbf{B}_i$, whose components are $\mathbf{B}_i = (X_{i1}, \ldots, X_{iN_G})$ is sampled from:

$$P(\mathbf{B}_i|\mathbf{N}) = \eta_i P_+(\mathbf{B}_i|\mathbf{N}) + (1-\eta_i)P_-(\mathbf{B}_i|\mathbf{N}), \tag{3.2.11}$$

where the quantity η_i is defined as $\eta_i = H(Y_{i3} - Y_0)$ (i.e. gene i is epigenetically open, corresponding to $\eta_i = 1$, if the corresponding level of acetylation, Y_{i3}, exceeds the threshold Y_0), $\mathbf{N} = (X_1, \ldots, X_{N_G})$ is a vector containing the monomer gene product of all genes, $P_-(\mathbf{B}_i|\mathbf{N}) = \prod_{j\in\langle i\rangle} \delta_{X_{ij},0}$, with $\delta_{X_{ij},0} = 1$ when $X_{ij} = 0$, and $P_+(\mathbf{B}_i|\mathbf{N})$ is a multinomial PDF, whose generating function is given by:

$$G(p_{i1}, \ldots, p_{iN_G}) = \left(\frac{1 + \sum_k \frac{\beta_{ik}}{\delta_{ik}} x_k^2 p_{ik}}{1 + \sum_k \frac{\beta_{ik}}{\delta_{ik}} x_k^2} \right)^{e_i} \tag{3.2.12}$$

where e_i denotes the number of binding sites at the promoter region of gene i, and the other parameters are defined in Table 3.2 (transition rates) and Table 3.4 (rescaled parameters).

Eq. (3.2.10) describes the inner dynamics of the fast components (enzymes and enzyme-substrate complexes) of the ER system for each gene i. The resulting stochastic dynamics describes how the enzymes switch between their free state and their complex state at constant rates (since the number of the different substrates is constant under the hypothesis of time scale separation). Since the number of enzymes is conserved, the (quasi-)steady distribution of the number of enzymes of each type in complex form is a binomial (see Section A.2.1.4 for a detailed derivation of this result). The corresponding generating functions are given by:

$$G_{HDM}(p_i) = \left(\frac{\kappa_{i2} + \kappa_{i3} + (\kappa_{i5} + \kappa_{i6})y_{i3} + (\kappa_{i1} + y_{i3})y_{i2}p_i}{(\kappa_{i2} + \kappa_{i3}) + (\kappa_{i1} + y_{i3})y_{i2} + (\kappa_{i5} + \kappa_{i6})y_{i3}} \right)^{e_{HDM}} \tag{3.2.13}$$

$$G_{HDAC}(p_i) = \left(\frac{\kappa_{i10} + \kappa_{i11} + (\kappa_{i13} + \kappa_{i14})y_{i2} + (\kappa_{i9} + \kappa_{i12}y_{i2})y_{i3}p_i}{(\kappa_{i10} + \kappa_{i11}) + (\kappa_{i9} + \kappa_{i12}y_{i2})y_{i3} + (\kappa_{i13} + \kappa_{i14})y_{i2}} \right)^{e_{HDAC}} \tag{3.2.14}$$

where $i = 1, \ldots, N_G$ and κ_{ij} are defined in Table 3.4. The number of free HDM and HDAC molecules is then obtained from the conservation equations $Y_{i4} = e_{HDM} - Y_{i5}$ and $Y_{i6} = e_{HDAC} - Y_{i7}$, where e_{HDM} and e_{HDAC} denote the total number of HDM and HDAC molecules, respectively.

3.2.2.0.2 Outer solution

The outer solution, corresponding to the dynamical evolution of the slow variables, is obtained by sampling the fast variables, whose values are needed to compute the reaction rates for the slow variables, from their QSSA PDFs (see Eqs. (3.2.11)-(3.2.14)):

$$x_i(\tau) = x_i(0) + \sum_{k=1}^{R_G} r_{G_{ik}} \frac{1}{S} \mathcal{P}\left(S \int_0^\tau w_k(\mathbf{x}(\sigma); \mathbf{y}_i(\sigma)) d\sigma \right) \tag{3.2.15}$$

$$y_{ij}(\tau) = y_{ij}(0) + \sum_{k=1}^{R_E} r_{E_{ijk}} \frac{1}{E} \mathcal{P}\left(E \frac{1}{\epsilon_2} \int_0^\tau v_{ik}(\mathbf{y}_i(\sigma)) d\sigma \right), \quad j = 1, 2, 3 \tag{3.2.16}$$

The QSSA PDFs of the fast variables are conditioned by the current value of the slow variables. We complete our asymptotic analysis by looking at the large S behaviour of the slow GRN variables (see Eq. (3.2.5)). We resort to a law of large numbers enunciated and proved by Kurtz which states that $S^{-1}\mathcal{P}(Su) \to u$ when $S \gg 1$ [94, 10, 5, 84]. We can apply this result straightforwardly to Eq. (3.2.15), which eventually leads to the asymptotic reduction of the full ER-GRN system:

$$\frac{\mathrm{d}x_i}{\mathrm{d}\tau} = R_i + \omega_{i1}x_{ii} - \omega_{i2}x_i - 2\sum_{j=1}^{N_G} \left(\beta_{ij}\eta_i \left(\frac{e_j}{E} - \sum_{k=1}^{N_G} x_{jk} \right) x_i^2 - \delta_{ij}x_{ij} \right), \quad \text{for } i = 1, \ldots, N_G \tag{3.2.17}$$

$$y_{ij} = y_{ij}(0) + \sum_{k=1}^{R_E} r_{E_{ijk}} \frac{1}{E} \mathcal{P}\left(E \frac{1}{\epsilon_2} \int_0^\tau v_{ik}(y_i(\sigma)) d\sigma \right), \; j = 1, 2, 3. \tag{3.2.18}$$

The resulting dynamics consists on a hybrid system where the dynamics of the TF monomers, $x_i(\tau)$, Eq. (3.2.17), is described in terms of a piece-wise deterministic Markov process [38, 20], i.e. by a system of ODEs perturbed at discrete times by two random processes, one corresponding to stochastic ER (Eq. (3.2.18)) and the other to TF dimers binding to the promoter regions. The latter are sampled from their QSSA PDFs, Eq. (3.2.11). The stochastic dynamics of the slow ER variables, Eq. (3.2.18) is in turn coupled to the random variation of the associated fast variables (ER enzymes, HDM and HDAC, and complexes). The number of complexes, Y_{i5} and Y_{i7}, are sampled from their QSSA PDFs, Eqs. (3.2.13) and (3.2.14). The corresponding numerical method used to simulate such system is described in detail in Section A.2.1.5.

3.2.3 Transitions between ER states: minimum action path approach

Noise-induced transitions are essential to understand ER dynamics and their effect on cell-fate determination [51]. Throughout the bistable regime, sufficiently large fluctuations in the stochastic ER system will induce switching between the open and silenced states. The rate at which such transitions occur can be described using reaction-rate theory [69] and large deviation theory [54], which show that the waiting time between transitions is exponentially distributed. The average switching time, τ_s, increases exponentially with system size, which in this case is given by the scale of ER substrates, Y [54, 158, 19, 21]:

$$\tau_s = C e^{Y\mathcal{S}}, \tag{3.2.19}$$

where C is a constant and $\mathcal{S}$ is the action of the stochastic switch. Eq. (3.2.19) is derived from considering the probability distribution of the so-called *fluctuation paths*, $\varphi(\tau)$, which connect the mean-field steady states in a time τ. According to large deviation theory [54, 158], we have $P(\varphi(\tau)) \sim e^{-Y\mathcal{A}_{FW}(\varphi(\tau))}$, which implies that the probability of observing paths different from the optimal, i.e. the path φ_* that minimises the action, is exponentially supressed as system size, Y, increases. This means that, for large enough system size, the behaviour of the system regarding large fluctuations is characterised by the optimal path, which is such that:

$$\mathcal{S} \equiv \mathcal{A}_{FW}(\varphi_*) = \min_{\tau, \varphi(\tau)} \mathcal{A}_{FW}(\varphi(\tau)). \tag{3.2.20}$$

An explicit form of the functional $\mathcal{A}_{FW}(\varphi(\tau))$ can be given if the dynamics is given by the corresponding chemical Langevin equation [61]:

$$\mathrm{d}y_{ij}(\tau) = f_j(\mathbf{y}_i)\mathrm{d}t + g_{ij}(\mathbf{y}_i)\mathrm{d}B_{j_t}, \tag{3.2.21}$$

where B_{j_t} denotes a Wiener process, and the mean-field drift, $f_j(\mathbf{y}_i)$, is $f_j(\mathbf{y}_i) = \sum_{k=1}^{R_E} r_{E_{ijk}} v_{ik}(\mathbf{y}_i)$, and the noise matrix, $g_{ij}(\mathbf{y}_i)$, $g_{jj} = \sqrt{\sum_{k=1}^{R_E} r_{E_{ijk}}^2 v_{ik}(\mathbf{y}_i)}$, $g_{ij} = 0$ if $i \neq j$. The rescaled variables $\mathbf{y}_i = (y_{ij} = \frac{Y_{ij}}{Y})$, $j = 1, \ldots, 7$, and the rescaled rates $v_{ik}(\mathbf{y}_i)$ are defined in Section 3.2.2. In this case, the action functional $\mathcal{A}_{FW}(\varphi(\tau))$ is the Freidlin-Wentzel (FW) functional:

$$\mathcal{A}_{FW}(\varphi) = \int_0^\tau ||\dot{\varphi}(t) - f(\varphi(t))||^2_{g(\varphi(t))} dt, \tag{3.2.22}$$

The norm $||\cdot||^2_{g(\varphi(t))} = \left\langle \cdot, D(\varphi(t))^{-1} \cdot \right\rangle$, where $D(\varphi(t)) = g(\varphi(t))\, g(\varphi(t))^T$ is the diffusion tensor. Using Eq. (3.2.22), the optimal value of the action, $\mathcal{S}$, can be found by numerical minimisation, which provides both the optimal or minimum action path (MAP) and the rate at which the ER system switch state driven by intrinsic noise. Details regarding implementation of the action-optimisation algorithm are given in Section A.2.3. A complete description of τ_s requires to estimate the pre-factor C, which is not provided by the FW theory, but can be easily estimated using stochastic simulation.

3.2.4 ER-systems ensemble generation and analysis

In Chapter 1 (Section 2.3.3) we have proposed to analyse an ensemble of ER systems in order to study the robustness of the different ER scenarios under heterogeneous conditions regarding the availability of co-factors associated with the activity of ER enzymes, which we take into account by considering variations (variability) in the kinetic constants c_{ij} (see Table 3.1 and 3.3). Such an ensemble is generated using approximate Bayesian computation (ABC) [156, 172], whereby we generate an ensemble of parameter sets $\theta_i = (c_{ij},\, i = 1, \ldots, N_G,\, j = 1, \ldots, 16)$ compatible with simulated data for the epigenetic regulation systems (see Section 2.2.3.2 for the details regarding ABC implementation).

Our approach follows closely the one presented in Chapter 2, since we will be using the data sets generated there. To summarise, we start by generating synthetic (simulated) data (denoted as "raw data" in Fig. 3.2) regarding the ER system of a genetic network of N_G genes epigenetically poised for differentiation, i.e. open

differentiation-promoting genes and silenced pluripotency-promoting genes (see the example shown in Fig. 3.2). This simulated data will play the role of the experimental data, x_0, to which we wish to fit our model. The data set consists of 10 realisations and 25 time points per realisation for each of the N_G genes. For each time point, t_i, we consider two summary statistics: the mean over realisations, $\bar{x}(t_i)$, and the associated standard deviation, $\sigma(t_i)$. We then run the ABC rejection sampler method until we reach an ensemble of 10000 parameter sets which fit the simulated data, x_0, within the prescribed tolerances for the mean and standard deviation. The heterogeneity associated with the rates c_{ij} of the parameter sets from the ensemble is interpreted in terms of variability in the availability of certain cofactors.

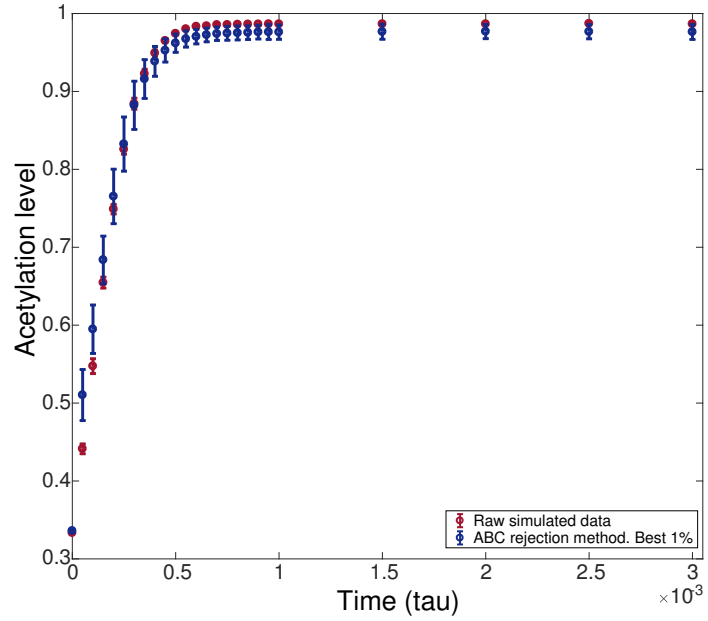
In our particular case $N_G = 2$, i.e. we have a 10000 parameter set ensemble of the ER system for the pluripotency-promoting gene and another 10000 parameter set ensemble for the gene promoting differentiation. Fig. 3.2 shows results comparing the reference (raw simulated) data to a subensemble average consisting of the 100 sets that best fit the data, for the differentiation-promoting gene, Fig. 3.2 (a) and pluripotency-promoting gene, Fig. 3.2 (b).

In our discussion, we will focus on this subensemble, i.e. the subensemble of the 100 parameter sets that best fit raw data (see Fig. 3.2). Specifically, we will consider those ER systems that for normal HME activity, i.e. when the number of HMEs is equal to its characteristic number, $e_{HDM} = e_{HDAC} = Z$, exhibit bistable behaviour. Within this subensemble, 90 sets for the differentiation-promoting ER system and 100 sets for the pluripotency-promoting ER system satisfy this condition.

3.3 Results

In order to focus our discussion, we illustrate the application of our model formulation on the case we want to study: a gene regulatory circuit with two genes, one whose product promotes differentiation and another one whose protein induces pluripotency. These two genes are further assumed to interact through mutual competitive inhibition (see Fig. 3.1(a)). Although such a system may appear to be too simplistic to describe realistic situations, there is evidence that mutual inhibition between two key transcription factors controls binary cell fate decisions in a number of situations [77, 105]. Our results are straightforward to generalise to more complex situations.

(a)



(b)

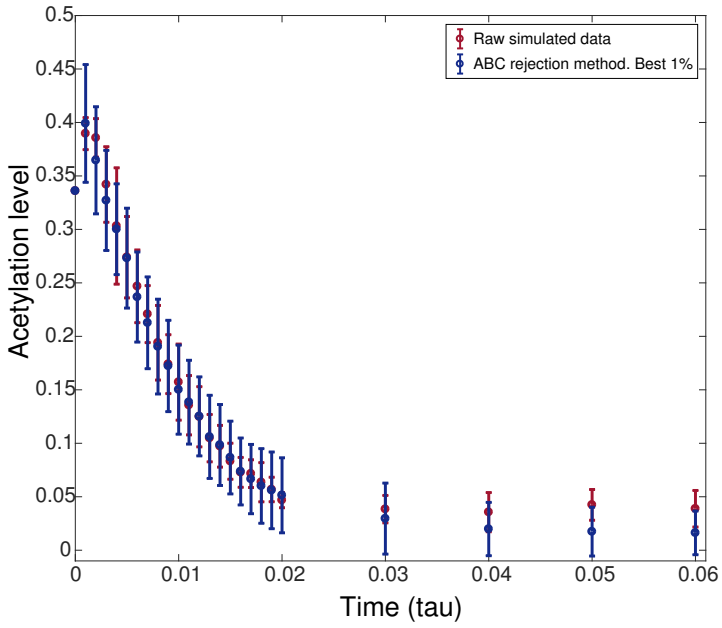


Figure 3.2: Caption on the following page.

Figure 3.2: (Previous page). Comparison between raw simulated data (red) with best fitted data resulting from the parametric sensitivity analysis of the epigenetic regulatory system (blue). (a) Evolution to the open state of the differentiation-regulating gene. (b) Evolution to the silenced state of the pluripotency-regulating gene. Raw simulated data is generated by using the SSA on the model defined by the rates shown in Table 3.3 with parameter values given in Tables A.5 and A.6, for (a) and (b), respectively. Resulting fitted data correspond to the 100 ABC parameter sets that best fit the raw data.

We proceed to analyse how ER sculpts the epigenetic landscape over the substrate of the phase space given by the model of the gene regulatory network (GRN). The latter provides the system with a variety of cell fates, corresponding to the stable steady states of the dynamical system underpinning the model of gene regulatory network [86]. The transitions between such cellular states, both deterministic and stochastic, depend upon the ability of the cell regulatory systems to elevate or lower the barriers between them. Epigenetic regulation is one of such mechanisms. Here, we examine how ER is affected by ensemble variability associated with variations in the availability of the necessary co-factors on which histone modifying enzymes (HMEs) depend to carry out their function. In particular, we will show that such variability is enough to produce a variety of behaviours, in particular differentiation-primed and stem-locked states.

3.3.1 The GRN model exhibits a complex phase space, including an undecided regulatory state

We start our analysis by studying the phase space of the dynamical system underlying our model of gene regulation, schematically illustrated in Fig. 3.1(a). Using the methodology described in detail in Section A.2.1.1, we have derived the (quasi-steady state approximation) equations for the optimal path theory of the stochastic model of the mutually inhibitory two-gene system [110]. Such equations describe the most likely relaxation trajectories towards their steady states [19, 21], under conditions of time scale separations described in detail in Section A.2.1.2.

$$\frac{\mathrm{d}q_1}{\mathrm{d}\tau} = R_1 + p_{\infty_1} p \frac{\omega_{11} \frac{\beta_{11}}{\delta_{11}} q_1^2}{1 + \frac{\beta_{11}}{\delta_{11}} q_1^2 + \frac{\beta_{12}}{\delta_{12}} q_2^2} - \omega_{12} q_1 \tag{3.3.1}$$

$$\frac{\mathrm{d}q_2}{\mathrm{d}\tau} = R_2 + p_{\infty_2} p \frac{\omega_{21} \frac{\beta_{22}}{\delta_{22}} q_2^2}{1 + \frac{\beta_{21}}{\delta_{21}} q_1^2 + \frac{\beta_{22}}{\delta_{22}} q_2^2} - \omega_{22} q_2 \tag{3.3.2}$$

where q_1 and q_2 are the variables (generalised coordinates) associated with the number of molecules of proteins, X_1 and X_2. The re-scaled variables, q_i and q_{ij}, and the re-scaled parameters, ω_{ij}, β_{ij}, and δ_{ij}, are defined in Table 3.4 (see also Section A.2.1.2).

The multiscale analysis carried out in Section A.2.1.2 shows that the parameters p, p_{∞_1} and p_{∞_2} are such that $p_{\infty_1} p = \frac{e_1}{E}$ and $p_{\infty_2} p = \frac{e_2}{E}$, where e_1 and e_2 are the number of sites in the promoter regions of our two genes exposed to and available for binding by TFs, which implies that $p_{\infty_1} p$ and $p_{\infty_2} p$ can be directly related to ER: $p_{\infty_i} p \to 0$, $i = 1, 2$, corresponds to an epigenetically silenced gene, whereas $p_{\infty_i} p \geq \mathcal{O}(1)$ is associated with an epigenetically open gene. In this section, we study the phase space of the system when both $p_{\infty_1} p$ and $p_{\infty_2} p$ are varied. This allows us to understand how the behaviour of the GRN changes when its components are subject to ER. Our results are shown in Fig. 3.3.

The system described by Eqs. (3.3.1)-(3.3.2) exhibits three types of biologically relevant steady states, namely, the *pluripotency* steady state (PSS), the *differentitation* steady state (DSS), and the *undecided* steady state (USS). Different combinations of these states can be stable or unstable depending on the parameter values (see Fig. 3.3). The PSS (DSS) corresponds to a steady state with $q_1 \ll 1$ and $q_2 = \mathcal{O}(1)$ ($q_1 = \mathcal{O}(1)$ and $q_2 \ll 1$) and the USS is associated with a state such that both $q_1 \ll 1$ and $q_2 \ll 1$.

Fig. 3.3 shows the phase space associated with the dynamical system Eqs. (3.3.1) and (3.3.2). It shows the behaviour as the parameters $p_{\infty_1} p$ and $p_{\infty_2} p$ vary. The lines shown in Fig. 3.3 correspond to the stability boundary of the different regimes. At such boundaries, saddle-node bifurcations occur, as illustrated in the example shown in Fig. A.4 (Section A.2.2). Fig. 3.3 reveals a complex phase space with seven different phases. We denote by $\mathcal{R}_P$ ($\mathcal{R}_D$) the region of the phase space where only the pluripotency (differentiation) steady state is stable. Similarly, $\mathcal{R}_U$ corresponds to those parameter values such that only the undecided steady state is stable. Furthermore, there are three bistable phases: one in which the PSS and the DSS coexist, $\mathcal{R}_{PD}$, a second one where the PSS coexists with the USS, $\mathcal{R}_{PU}$, and the third one where the DSS and USS coexist, $\mathcal{R}_{DU}$. Finally, a region exists where stable PSS, stable DSS, and stable USS, $\mathcal{R}_{PUD}$, coexist. Fig. A.7 shows examples of trajectories

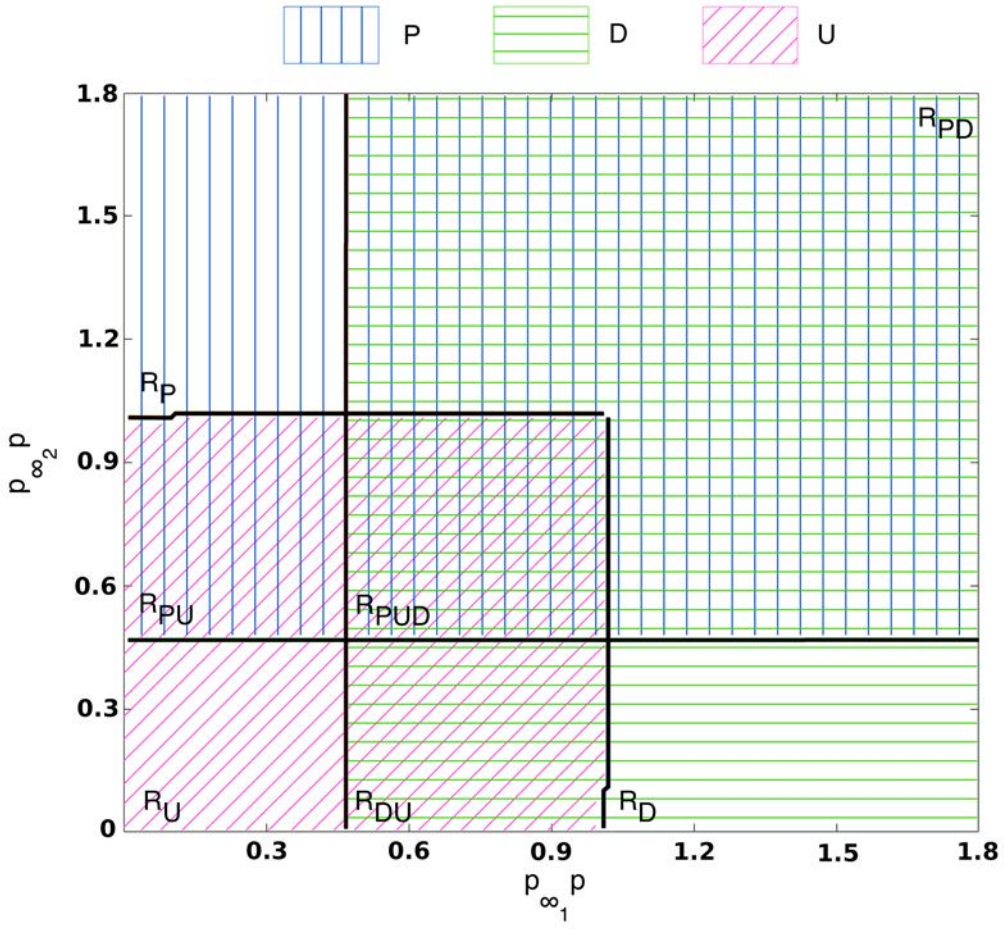


Figure 3.3: Phase diagram of the two-gene system, Eqs. (3.3.1)-(3.3.2). Vertical blue (horizontal green) hatching denotes regions where the pluripotency (differentiated) state is stable. Diagonal pink hatching denotes regions where the undecided state is stable. Regions of the phase diagram where different hatchings overlap correspond to regions of bistability or tristability. In the labels in the plot, P stands for pluripotency, D stands for differentiation and U for undecided. This phase diagram was obtained using the methodology formulated in [71]. Parameters values: $\omega_{11} = \omega_{21} = 4.0$. Other parameter values as per Table A.12.

illustrating the dynamics described by Eqs. (3.3.1)-(3.3.2) for different values of the pair $(p_{\infty_1}p, p_{\infty_2}p)$ corresponding to the different regions shown in Fig. 3.3. In particular, we show how the long term behaviour of different initial conditions differ as $(p_{\infty_1}p, p_{\infty_2}p)$ varies, so that different cell fates (co)exist related to different levels of TF accessibility.

3.3.2 Co-factor heterogeneity gives rise to both pluripotency-locked and differentiation-primed states

In the previous section, we have analysed the dynamical landscape provided by the dynamical system describing the GRN. We now proceed to study the effect of ER on the robustness of the different phases shown in Fig. 3.3 (see also Fig. A.7). We here put forward that ER is essential to the robustness of such phases and, consequently, to the stability of the associated cell fates, since transitions in bistable ER systems can induce (or facilitate) transitions between the GRN phases. Such transitions are associated with differentiation and reprogramming of cell fates. This phenomenon, so-called *epigenetic plasticity*, has been recently proposed as a major driver for disrupting cell-fate regulatory mechanisms in cancer and ageing [51]. We further focus on the role of heterogeneity within the ensemble described in Section 3.2.4 (see also Section 2.3.3).

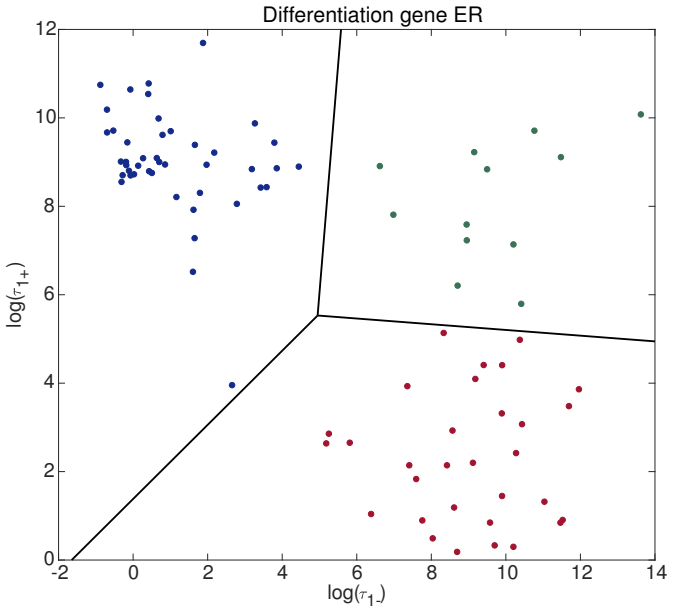
In order to characterise robustness of the different ER systems within the subensemble, we have focused on the analysis of the average transition times between the *open* and *closed* ER states. We define τ_{1_+} (τ_{1_-}) as the average transition time for a differentiation ER system- DERS- to switch from closed to open (open to closed). Similarly, the quantities τ_{2_+} and τ_{2_-} are analogously defined for the pluripotency ER systems-PERSs. The results are shown in Fig. 3.4(a) and (b), where we present scatter plots of the average transition times within the ensemble of DERSs (Fig. 3.4(a)) and PERSs (Fig. 3.4(b)). These figures show scatter plots where each point represents an ER system (i.e. a given parameter set) within our subensemble. The vertical and horizontal axes show the average switching time from closed-to-open and open-to-closed, respectively. We observe that the heterogeneity exhibited by the differentiation ER systems is greater than the one corresponding to the pluripotency ER systems. In particular, the dispersion in τ_{2_+} is much smaller than in τ_{1_+}. Regarding τ_{2_+}, most of the pluripotency ER systems are concentrated around a narrow band. By contrast, the differentiation ER systems show large degrees of heterogeneity in both τ_{1_+} and τ_{1_-}.

Heterogeneity in the differentiation ER systems exhibits an interesting pattern, whereby such systems organise themselves in three clusters obtained through k-means clustering, shown as blue, green and red dots in Fig. 3.4(a). DERSs within the *blue* cluster are charaterised by long closed-to-open waiting times and short open-to-closed waiting times. DERSs belonging to the *red* cluster are the specular image of those within the blue cluster, i.e. they have short closed-to-open waiting times and long open-to-closed waiting times. Finally, DERSs in the *green* cluster are characterised by large values of both τ_{1_+} and τ_{1_-}.

Insight into the stochastic dynamics, particularly regarding heterogeneity of the robustness of the open and silenced ER states to intrinsic noise, can be gained by analysing the corresponding optimal escape paths. Four examples of such paths, computed according to the MAP theory (see Section 3.2.3), for two DERSs (DERS1 and DERS2) and two PERSs (PERS1 and PERS2) are shown in Figs. A.8(a)-(d). A comparison between the value of the minimum action $\mathcal{S}$ (see Eqs. (3.2.20)-(3.2.22)), for the optimal escape paths corresponding to DERS1 and DERS2 (Figs. A.8(a) and (c)), and for PERS1 and PERS2 (Figs. A.8(b) and (d)) shows a tendency for DERSs to exhibit much more variability (see Table 3.5, where we report the minimum action values for the optimal paths). Whilst the action value for PERS1 is about twice the value of PERS2 in Fig. A.8(b) (closed to open), there is an over 8-fold increase when comparing the action values of DERS1 and DERS2 in Fig. A.8(a) (closed to open). Similarly, when comparing the action $\mathcal{S}$ for the open to closed optimal paths, we observe that the variability associated with the DERSs (Fig. A.8(c)) is also larger than the one in PERSs (Fig. A.8(d)). This property partly explains the difference between Fig. 3.4(a) and Fig. 3.4(b) regarding DERS and PERS heterogeneity, respectively. A similar argument can be put forward to help us explain the heterogeneity within the DERS ensemble (Fig. 3.4(a)). Blue cluster DERSs exhibit optimal closed-to-open paths with larger value of the optimal action than that found in their red cluster counterparts (see Fig. A.8(a) and Table 3.5, where DERS1 belongs to the red cluster and DERS2 to the blue cluster). This property has the consequence that the closed-to-open waiting time, τ_{1_+}, is longer for blue cluster DERSs.

To quantify the effects of bistable ER on the landscape related to the gene regulatory system (see Fig. 3.3), we proceed to estimate the probability, $\mathcal{Q}$, that the combined activity of each pair of DERS and PERS within our ensemble produces a global epigenetic regulatory state compatible with differentiation. DERS-PERS pairs

(a)



(b)

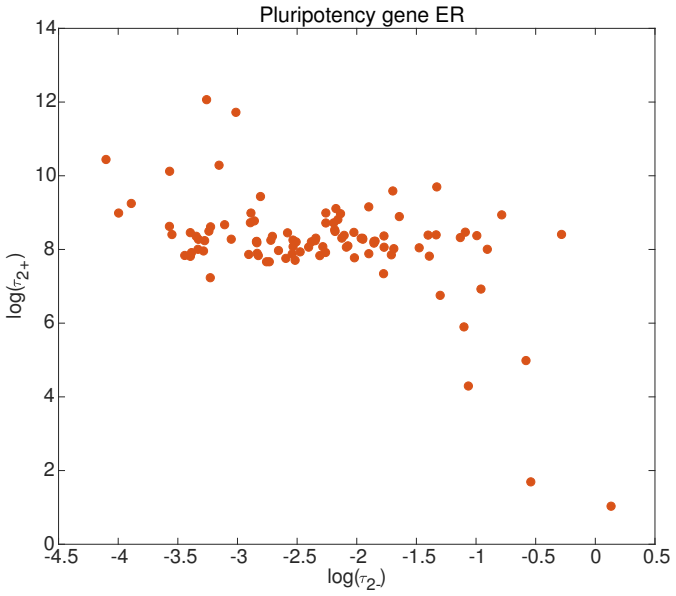


Figure 3.4: Caption on the following page.

Figure 3.4: (Previous page). Scatter plots showing heterogeneity in the behaviour of bistable differentiation ER systems (DERSs) and pluripotency ER systems (PERSs). The vertical axis corresponds to the average opening time and the horizontal axis, to the average closing time. Each dot in plot (a) represents a DERS within the ensemble (see Section 3.2.4). Different colours and black lines show the three clusters resulting from a k-means analysis discussed in Sections 3.3.2 and 3.3.3. Dots in plot (b) represent PERSs within the ensemble defined in Section 3.2.4.

Table 3.5: Minimum action values, $\mathcal{S}$, corresponding to the optimal escape paths shown in Fig. A.8 (see Section 3.2.3 and Section 3.3.2 for details). Parameter values are given in Tables A.15-A.18, for PERS1, PERS2, DERS1 and DERS2, respectively.

ER system	Open to closed	Closed to open
DERS1	0.05387	0.007012
DERS2	0.09947	0.05813
PERS1	0.01502	0.1836
PERS2	0.02043	0.07645

with high values of $\mathcal{Q}$ are associated with *differentiation-primed* states. By contrast, those DERS-PERS combinations with low $\mathcal{Q}$ are identified with *pluripotency-locked* states.

We proceed forward with this programme by recalling that escape times from a stable attractor in a stochastic multi-stable system are exponentially distributed [54, 158]. This implies that the PDFs for the escape times for both DERSs and PERSs are fully determined by the corresponding values of $\tau_{1\pm}$ and $\tau_{2\pm}$. We also assume that, for a given ER-GRN system, the DERS and the PERS evolve independently of each other.

We consider PDF of the waiting time associated with a scenario of full reprogramming of the epigenetic landscape, τ_P. Such a scenario assumes that the system is initially in a pluripotency-locked ER state where the DERS is closed and the PERS is open, which we denote as D_-P_+. For the system to make its transit into the differentiation-primed state D_+P_-, corresponding to open DERS and closed PERS, there are two possible routes: $D_-P_+ \to D_-P_- \to D_+P_-$ (route 1) and $D_-P_+ \to D_+P_+ \to D_+P_-$ (route 2). Simultaneous switch of both ER systems is

considered highly unlikely and therefore ignored. The PDF of the waiting time of the transition $D_-P_+ \to D_+P_-$, denoted by $P_{+-,-+}(\tau_P)$, is given by:

$$P_{+-,-+}(\tau_P) = \mathcal{Z}^{-1}\left(P_1(\tau_P) + P_2(\tau_P)\right), \tag{3.3.3}$$

where

$$P_1(\tau) = \tau_{1-}^{-1}\tau_{2+}^{-1}e^{-\tau/\tau_{1-}}\left(\frac{e^{-\tau/\tau_{2+}} - e^{-\tau/\tau_{2-}}}{\tau_{2-}^{-1} - \tau_{2+}^{-1}}\right)$$

$$P_2(\tau) = \tau_{1-}^{-1}\tau_{2+}^{-1}e^{-\tau/\tau_{2+}}\left(\frac{e^{-\tau/\tau_{1+}} - e^{-\tau/\tau_{1-}}}{\tau_{1-}^{-1} - \tau_{1+}^{-1}}\right),$$

and

$$\mathcal{Z}^{-1} = \frac{(\tau_{1-} + \tau_{2+})\left((\tau_{1-}^{-1} + \tau_{2-}^{-1})(\tau_{1+}^{-1} + \tau_{2+}^{-1})\right)}{\tau_{2+}^{-1} + \tau_{1+}^{-1} + \tau_{2-}^{-1} + \tau_{1-}^{-1}}.$$

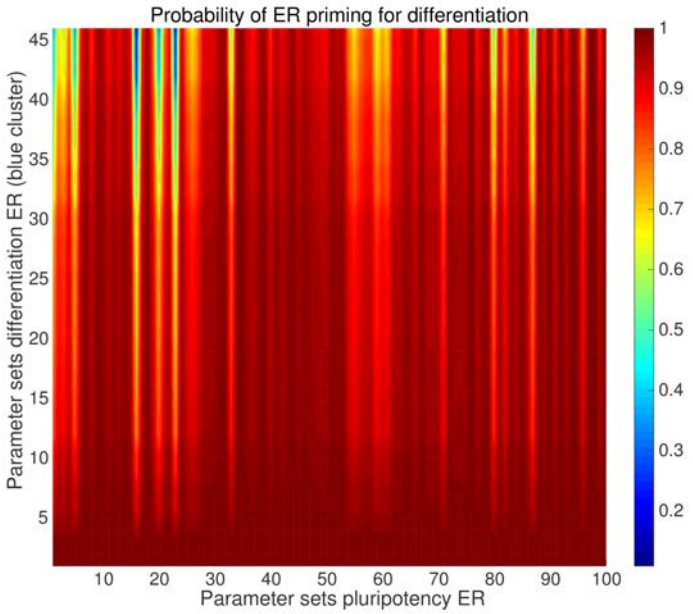
$P_1(\tau_p)$ and $P_2(\tau_p)$ are the probabilities related to each of the landscape reprogramming routes. The probability that the ER landscape has undergone reprogramming from pluripotency-locked into differentiation-primed state within the time interval $(0, \tau_P]$, $\mathcal{Q}$, is thus given by:

$$\mathcal{Q} \equiv \int_0^{\tau_P} P_{+-,-+}(\tau)\mathrm{d}\tau, \tag{3.3.4}$$

where in our case, τ_P has been taken as the mean time of τ_{1+}, that is, the mean time for the differentiation ER systems (DERSs) to switch from the closed to the open state, switch needed for a cell to differentiate. Furthermore, we have chosen τ_{1+} since it exhibits a larger range of variability than the time for the pluripotency ER systems to switch from its open to its closed state, which is also a necessary condition for differentiation to happen.

We investigate the DERSs belonging to the different clusters of Fig. 3.4(a) regarding their likelihood to produce pluripotency-locked epigenetic landscapes (results shown in Fig. 3.5). The analysis shows that DERSs within the red cluster (Fig. 3.5(c)) correspond to differentiation-primed epigenetic landscapes ($\mathcal{Q} = 1$)for all the DERS-PERS pairs. By contrast, the blue cluster (Fig. 3.5(a)) and the green cluster (Fig. 3.5(b)) contain DERSs associated with both differentiation-primed (large $\mathcal{Q}$) and pluripotency-locked (small $\mathcal{Q}$) epigenetic landscapes. As discussed in the next section, the latter are more abundant within the blue cluster.

(a)



(b)

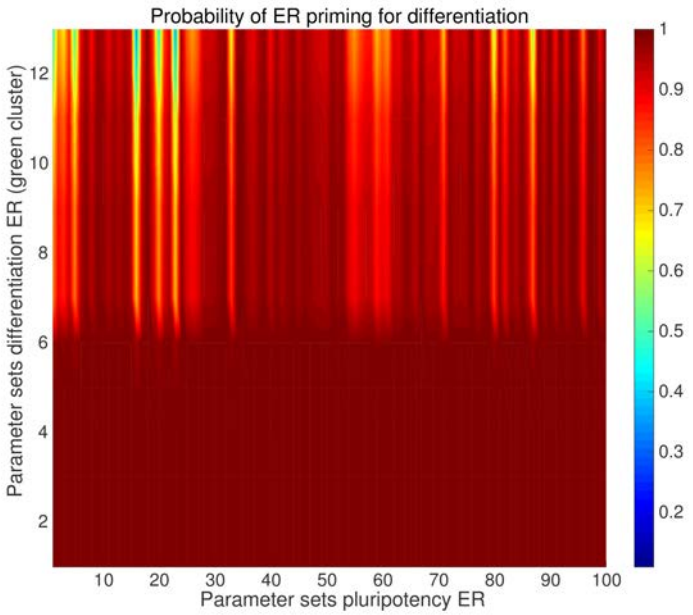


Figure 3.5: Figure continues on the following page. Caption on the following page.

(c)

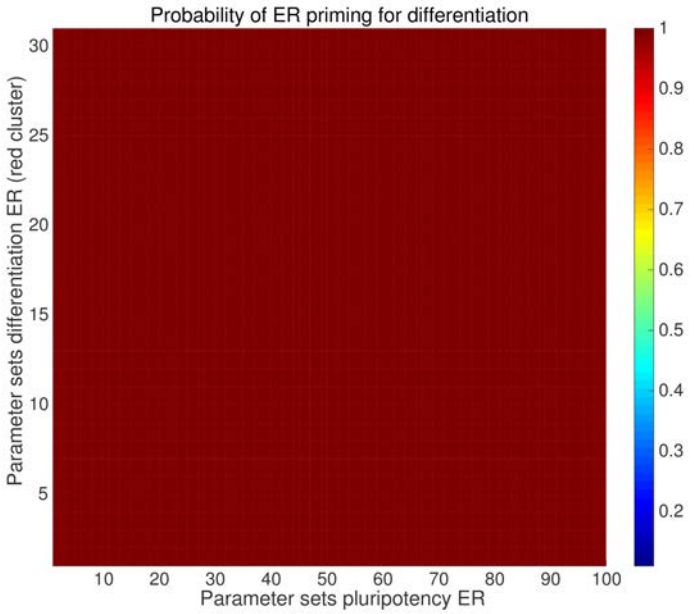


Figure 3.5: (Previous page). Differentiation probability $\mathcal{Q}$ within the ensembles of DERSs and PERSs corresponding to the three clusters of Fig. 3.4(a) (see Section 3.3.2). For all three plots, the horizontal axis runs over the whole ensemble of PERSs. The vertical axis of plots (a), (b), and (c) runs over all the DERSs within the blue cluster, the green cluster, and the red cluster, respectively.

3.3.3 Analysis of ensemble heterogeneity

We now proceed to analyse the patterns observed in our ensemble of ER systems regarding both the differences between the three clusters observed in the ensemble of DERSs (Fig. 3.4(a)) and the distinctive features that characterise pluripotency-locked DERS-PERS pairs. In order to do this, we follow the methodology put forward in Section 2.3.3, whereby ensemble statistics (cumulative distribution functions (CDFs)) of the parameters c_{ij} (see Table 3.1) corresponding to the DERSs/PERSs associated with the subensemble of systems exhibiting a particular behaviour are analysed. By comparing such CDFs to either the general population (i.e. whole ensemble) or to different subensembles, we can detect statistically significant biases, which allows us to identify key parameters (and their biases) associated with the

behaviour displayed by the focal subensemble.

3.3.3.1 Significant differences within the ensemble of DERSs

We start this analysis by studying the pattern emerging in the ensemble of DERSs, Fig. 3.4(a). As discussed in the previous section, DERSs organise themselves in three clusters, which exhibit remarkable differences regarding their capability to trigger differentiation-primed epigenetic landscapes (see Figs. 3.4(a) and 3.5). Our results are shown in Fig. 3.6, where we depict the empirical CDFs for the different kinetic parameters of the ER reactions for the differentiation gene, c_{1_j} (see Table 3.3). We proceed to look for which c_{1_j} there are statistically significant differences, by comparing the CDF of each cluster with that corresponding to the whole DERS ensemble, and also the CDFs of the clusters among them (see Fig. 3.6). Each of these two-sample comparison is carried out by means of the Kolmogorov-Smirnov (KS) test. Statistically significant differences were found in the cases we comment below. The p-values are reported in Section A.2.4.

Red cluster versus blue cluster. As discussed in the previous section, the differences between DERSs within the blue and red clusters are essential to ascertain the main features that distinguish differentiation-primed and pluripotency-locked systems. The bias detected within the red (blue) cluster in the corresponding CDFs (see Fig. 3.6) is towards bigger (smaller) values for c_{1_1} (unrecruited demethylation) and $c_{1_{15}}$ (unrecruited acetylation) and towards smaller (larger) values for $c_{1_{11}}$ (unrecruited deacetylation) and $c_{1_{16}}$ (recruited acetylation). The behaviour of c_{1_1}, $c_{1_{11}}$, and $c_{1_{15}}$ is straightforward to interpret. The trends observed in the data are consistent with DERSs within red cluster being more prone to differentiation-primed ER landscapes, as they promote removal of negative marks and addition of positive marks.

Red cluster versus green cluster. In this case, the bias detected within the red (green) cluster in the corresponding CDFs (see Fig. 3.6) is to larger (smaller) values for c_{1_3} (unrecruited demethylation) and to smaller (bigger) values for $c_{1_{16}}$ (recruited acetylation). The tendency in the data corresponding to c_{1_3} is compatible with the features of the red cluster DERSs, as it involves an increase in the removal of negative marks.

Blue cluster versus green cluster. Fig. 3.6 shows that DERSs within the green cluster have smaller values of c_{1_3} (unrecruited demethylation) and larger values of

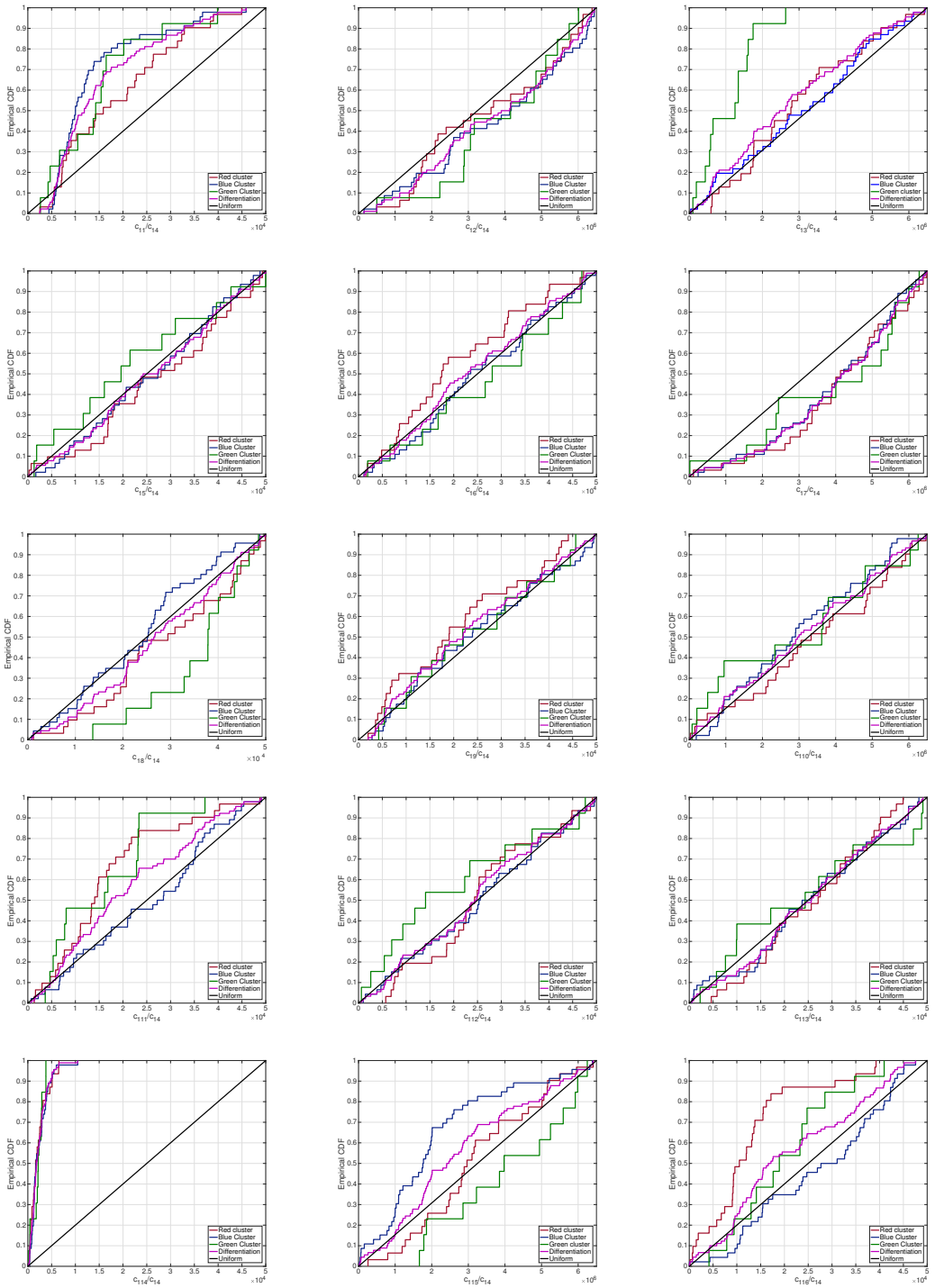


Figure 3.6: Caption on the following page.

Figure 3.6: (Previous page). Empirical CDFs for the whole ensemble of DERS parameter sets (magenta lines). This ensemble has been generated according to the methodology explained in Section 3.2.4 (see also Section 2.2.3.2). We also show the partial empirical CDFs corresponding to each of the clusters from Fig. 3.4(a) (red, green, and blue lines). We analyse a total of 90 DERS parameter sets. The red cluster includes 31 sets, the green cluster contains 13 sets, and the blue cluster has 46 sets. For reference, we also show the CDF for a uniform distribution (black line).

c_{1_8} (recruited methylation) than their blue cluster counterparts. Both of such effects stimulate addition of negative marks. However, DERSs in the green cluster also exhibit lower $c_{1_{11}}$ (unrecruited deacetylation) and bigger $c_{1_{15}}$ (unrecruited acetylation), which both encourage addition of positive marks. This can explain why the green cluster DERSs exhibit both long τ_{1-} and τ_{1+} (see Fig. 3.4(a)).

3.3.3.2 Significant differences between differentiation-primed and pluripotency-locked ER landscapes

The quantity $\mathcal{Q}$ allows us to classify each pair DERS-PERS drawn from our ensemble regarding their degree of resilience to switch into a state prone to differentiation. If $\mathcal{Q}$ is larger than a threshold value $\mathcal{T}$, the corresponding DERS-PERS pair is categorised as differentiation-primed. By contrast, when $\mathcal{Q} < \mathcal{T}$, the DERS-PERS pair is classified as pluripotency-locked.

We first proceed to compare within the whole population (without discriminating between clusters) those DERSs such that $\mathcal{Q} \geq \mathcal{T}$ (differentiation-primed ER landscapes) against those with $\mathcal{Q} < \mathcal{T}$ (pluripotency-locked ER landscapes). We take $\mathcal{T} = 0.7$. The results are shown in Fig. A.9. The CDFs of the parameters c_{1_1} (unrecruited demethylation), $c_{1_{14}}$ (recruited deacetylation), and $c_{1_{15}}$ (unrecruited acetylation) are biased towards higher values for the subensemble associated with differentiation-primed ER landscapes ($\mathcal{Q} \geq \mathcal{T}$). The requirement for $\mathcal{Q}$ to be $\mathcal{Q} \geq \mathcal{T}$ biases the CDF of $c_{1_{16}}$ (recruited acetylation) towards lower values than in the general population. The interpretation of the results regarding c_{1_1} and $c_{1_{15}}$ is clear, since they encourage the removal of negative marks and the addition of positive marks and thus promote expression of the differentiation gene. The CDFs of $c_{1_{14}}$ and $c_{1_{16}}$ corresponding to differentiation-primed ER landscapes are virtually identical to the CDFs associated with the general population (see Fig. A.9). These features are

therefore inherent in bistable behaviour (see Section 3.2.1), rather than being specific to differentiation-primed DERSs.

If we now restrict our analysis to those DERSs within the blue cluster (see Fig. 3.7), we observe that the parameters whose CDFs differ significantly when splitted into differentiation-primed and pluripotency-locked are c_{1_1} (unrecruited demethylation) and $c_{1_{14}}$ (recruited deacetylation). As in the analysis in the whole ensemble, only the result regarding c_{1_1} is relevant for the analysis of the features yielding to differentiation-primed ER landscapes.

Regarding the PERSs, the results are less compelling. The results are shown in Fig. A.10. Our analysis shows that significative differences can be found between the empirical distributions of three parameter values: c_{1_3} (unrecruited demethylation), c_{1_8} (recruited methylation), and $c_{1_{15}}$ (unrecruited acetylation). PERSs such that $\mathcal{Q} \geq \mathcal{T}$ exhibit larger values of all three parameters.

3.3.4 Ensemble-based strategies for unlocking resilient pluripotency

The results of the previous sections suggest a number of strategies to unlock resilient pluripotency states which hinder differentiation. One of our main conclusions is that such states of resilient pluripotency are mostly vinculated to DERS-PERS combinations such that the DERS belongs to either the blue or the green cluster. In view of this, a possible strategy in order to encourage differentitation-primed ER landscape consists on changing a selected combination of parameter values according to a rationale provided by the analysis carried out in the previous two sections. Our results are shown in Fig. 3.8.

One possible strategy consists on first transforming a blue cluster DERS into a green cluster one, and then completing the *reprogramming* of the DERS by transforming the resulting set into a red cluster DERS. A candidate strategy involves first changing a parameter whose CDF is significantly different when the blue cluster is compared with the green cluster. The second step is then to change a parameter that exhibits significant difference between the green and red cluster. Taking the results of the previous section into consideration, we consider the reduction of $c_{1_{11}}$ (unrecruited deacetylation) and the increase of c_{1_3} (unrecruited demethylation). The result of this reprogramming strategy is shown in Fig. 3.8(a), where we show that a blue cluster DERS (big blue circle) is first transformed into a green cluster one

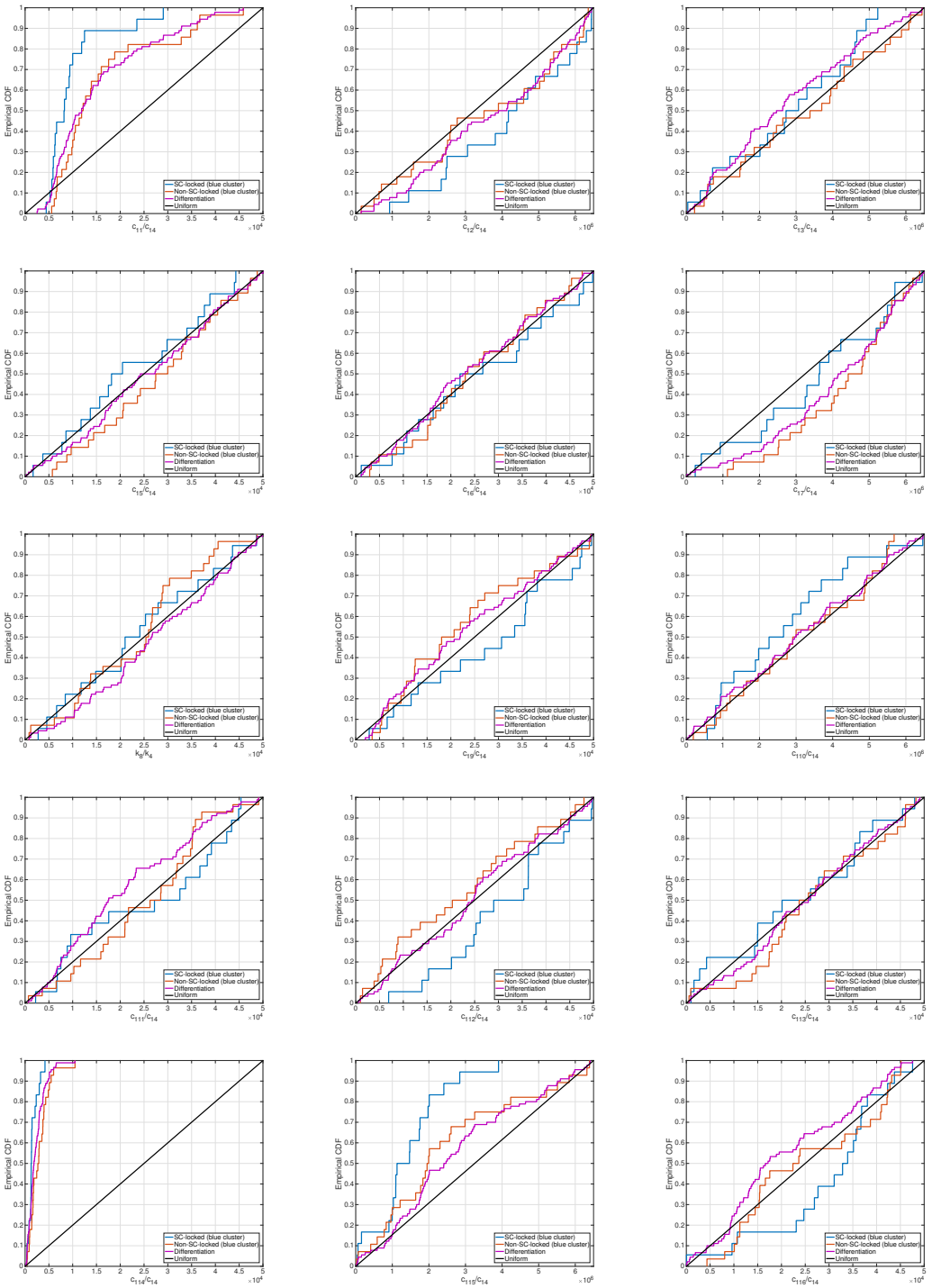


Figure 3.7: Caption on the following page.

Figure 3.7: (Previous page). Empirical CDFs for the DERS parameter sets within the blue cluster. This ensemble has been generated according to the methodology explained in Section 3.2.4 (see also Section 2.2.3.2). The DERSs within the blue cluster have been divided into two subsets: those such that $\mathcal{Q} < \mathcal{T}$ (SC-locked, blue lines) and those such that $\mathcal{Q} \geq \mathcal{T}$ (non-SC-locked, orange lines), with $T = 0.7$. For comparison, we plot the CDFs of the whole DERS ensemble (magenta lines), and, for guidance the CDF corresponding to a uniformly distributed random variable (black lines).

(green square in Fig. 3.8(a)), and then, finally, into a red cluster DERS (red square in Fig. 3.8(a)). The initial blue cluster DERS has been chosen as the set with the largest value of $c_{1_{11}}$, which has been shown to be a significant difference when comparing the blue cluster to the red one, and the blue cluster to the green one, leading to the idea that this property is linked to the blue cluster (idea which is reinforced because $c_{1_{11}}$ is not significant when comparing the red and the green cluster).

The efficiency of such a strategy to unlock resilient pluripotency and to encourage differentiation is shown in Fig. 3.8(b), where we present statistics of the differentiation time, τ_D, for the original blue cluster DERS and for the corresponding reprogrammed one (two step reprogramming, *red cluster-like*). These simulations have been done for the full ER-GRN, Eqs. (3.2.17)-(3.2.18), using the hybrid multiscale simulation algorithm described in Section 3.2.2 (and fully developed in Section A.2.1.5). The resulting differentiation times for the ER-GRN with reprogrammed ER landscape are orders of magnitude smaller than those with original ER-GRN within the blue cluster DERS.

An alternative strategy, that involves changing the value of one parameter only, consists on increasing the value of c_{1_3} (unrecruited demethylation). Such a strategy is not obvious, since c_{1_3} is not one of the parameters whose empirical CDF has significant differences when DERS in the red cluster are directly compared with those in the blue cluster. However, since the CDF of c_{1_3} is significantly different when both the blue cluster and the red cluster are compared to the green cluster, it is conceivable that increasing c_{1_3} without further intervention could reprogram blue cluster DERSs. The result of this reprogramming strategy is shown in Fig. 3.8(a) (red diamond). Simulation results shown in Fig. 3.8(b) (two step reprogramming) confirm the viability of this approach. In fact, based on the statistics of the differentiation time, both strategies are virtually indistinguishable.

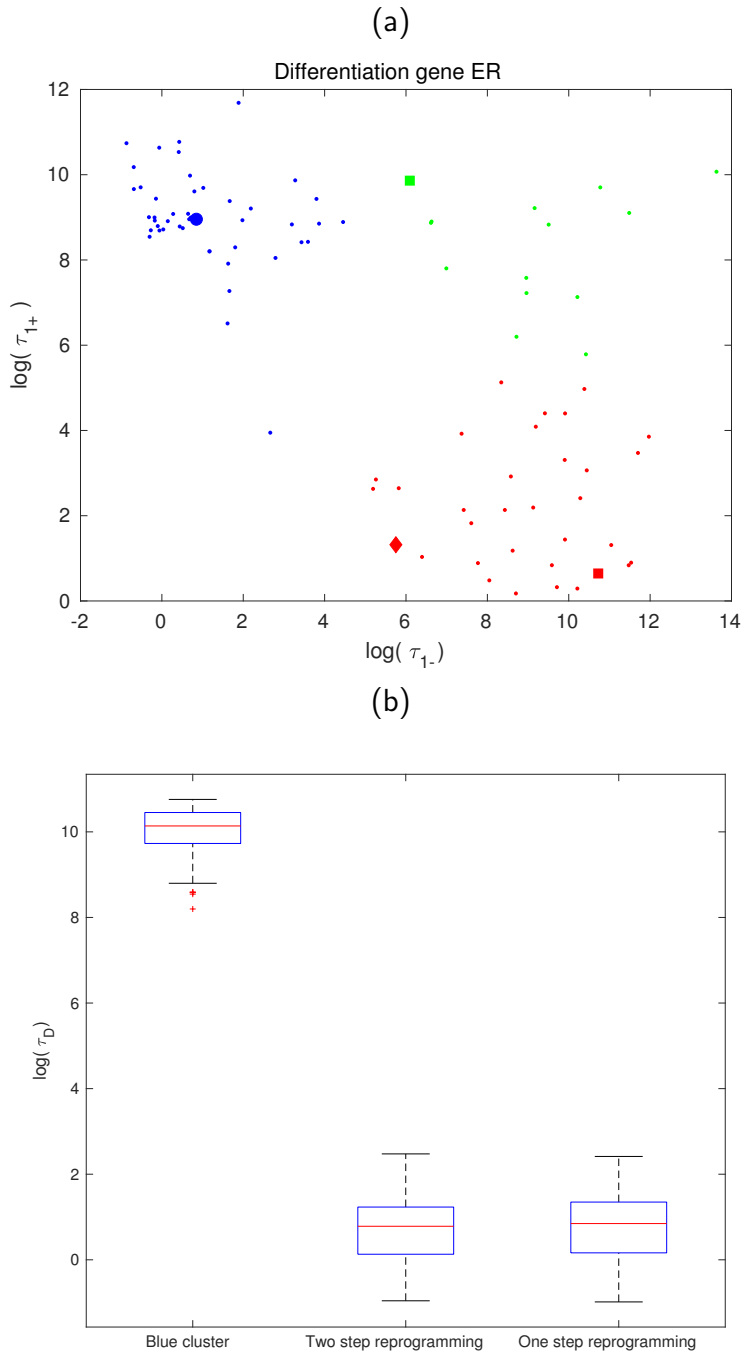


Figure 3.8: Caption on the following page.

Figure 3.8: (Previous page). Plots showing the effect of the different reprogramming strategies of blue cluster DERSs, as evaluated in terms of the statistics of the differentiation time (τ_D). (a) Two step reprogramming is illustrated by the green square (first step), which finally becomes the red square (second step). One step reprogramming is depicted as the red diamond (see Section 3.3.4 for details). (b) Comparison of τ_D for the original DERS (big blue circle in (a)) and the ones resulting from the reprogramming strategies. We consider a base-line scenario where the number of HMEs is exactly equal to average, i.e. $e_{HDM} = e_{HDAC} = Z$. We then compare the simulation results obtained for different scenarios regarding the different strategies to the base-line scenario. Parameter values: $Z = 5$ and $Y = 15$. Other parameter values given in Table A.19.

3.3.5 Loss of HDAC activity hinders differentiation in our ER-GRN model

Besides variability associated with cofactor heterogeneity, our model allows us to address the issue of variability regarding HME activity. HME activity is affected by both normal physiological processes, such as ageing, and pathologies such as cancer. For example, impaired activity of HDM and HDAC has been observed in relation to cancer and ageing. Here, we analyse the impact of HDM and HDAC loss of activity on the dynamics of differentiation. In particular, we simulate differentiation in our ER-GRN model to obtain statistics of the differentiation time to assess the effect of loss of HME activity. The simulations shown in this section have all been carried out using the hybrid multiscale simulation algorithm described in Section A.2.1.5.

In order to clarify the effect of loss of HME activity on the ER model, we first consider the phase diagram of its mean-field limit in different situations (see Section 2.3.1 for details). The results are shown in Fig. A.11. Figs. A.11(a) and (c) show the phase diagram for two DERSs, one belonging to the red cluster (DERS1) and another, to the blue cluster (DERS2). Regarding the features of the phase diagram, the main distinction between blue cluster DERSs and red cluster DERSs, as illustrated in the examples shown in Figs. A.11(a) and (c), is that the surface occupied by the bistable region (shaded blue region) is much larger in blue cluster DERSs, because of the displacement of its lower boundary. By comparison, the bistability region of the PERSs is narrower than that of the DERSs (see Figs. A.11(b) and

(d)). In particular the boundary that separates the bistable phase from the closed phase (area at the left of the blue shaded region) is displaced towards smaller HDM activity in the DERS phase diagrams.

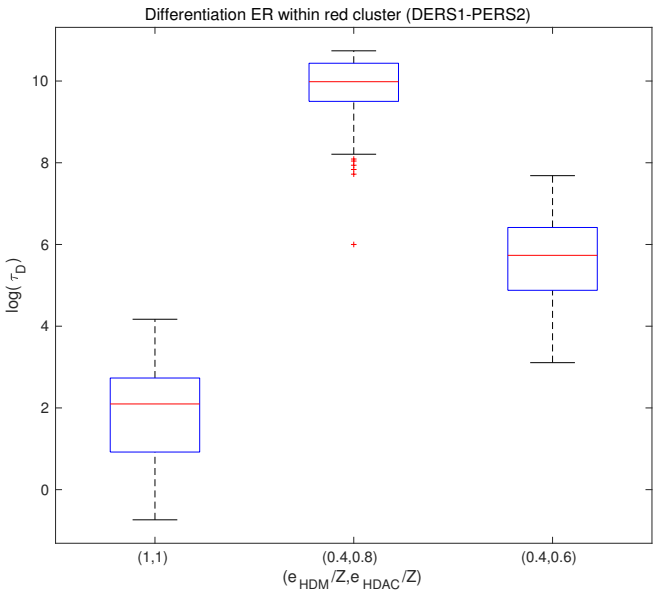
This property suggests that a possible strategy to promote differentiation would be to decrease HDM activity, as this would drive the PERS into its closed phase whilst allowing the DERS to remain within its bistability region. In order to assess this and other scenarios, we consider a base-line scenario where the number of HMEs is exactly equal to average, i.e. $e_{HDM} = e_{HDAC} = Z$. We then compare different scenarios regarding the abundance of HDM and HDAC to the base-line scenario.

Contrary to what could be expected, simulation results show that the strategy of reducing HDM activity alone beyond the PERS closing boundary further hinders differentiation. As can be seen in Fig. 3.9 (c), a decrease in HDM activity actually leads to longer differentiation time (see also [110]). Similarly, Figs. 3.9 (a) and (b), which show statistics of the differentiation time, reveal that a decrease in both HDM and HDAC activity also leads to an increment in differentiation times, that is, this strategy fails to decrease the differentiation time below the base-line scenario. In both cases, such hindrance of differentiation is the product of the increase in the opening times (τ_{1+}) of the DERS. This effect occurs because, as HDM and HDAC activity is reduced, the DERS is driven towards the boundary between the closed state and the bistability region. Close to such a region, the DERS closed state becomes more stable and thus the corresponding τ_{1+} increases. By contrast, further reduction of HDAC activity moves the DERSs system closer to their boundary between the bistability region and the open state, resulting in a reduction of the differentiation time. However, since the differentiation times remain above those corresponding to the base line HDM and HDAC activity scenario (see Fig. 3.9 (a) and (b)), we conclude that loss of both HDM and HDAC activity contributes towards hindering differentiation.

3.4 Conclusion

In this Chapter we have presented a model of epigenetic plasticity which has helped us to uncover some of the details and mechanisms underlying epigenetic regulation of phenotypic robustness, in particular regarding the robustness of pluripotent states. We have further uncovered how epigenetic heterogeneity regulates the deci-

(a)



(b)

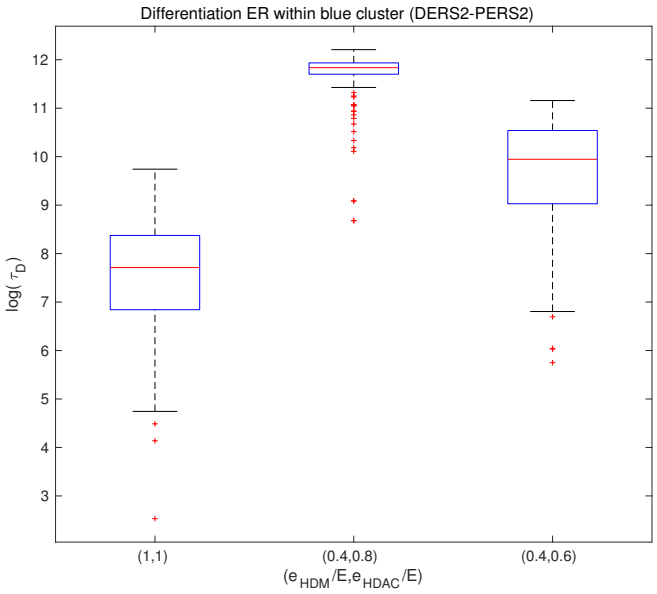


Figure 3.9: Plot continues on the following page. Caption on the following page.

(c)

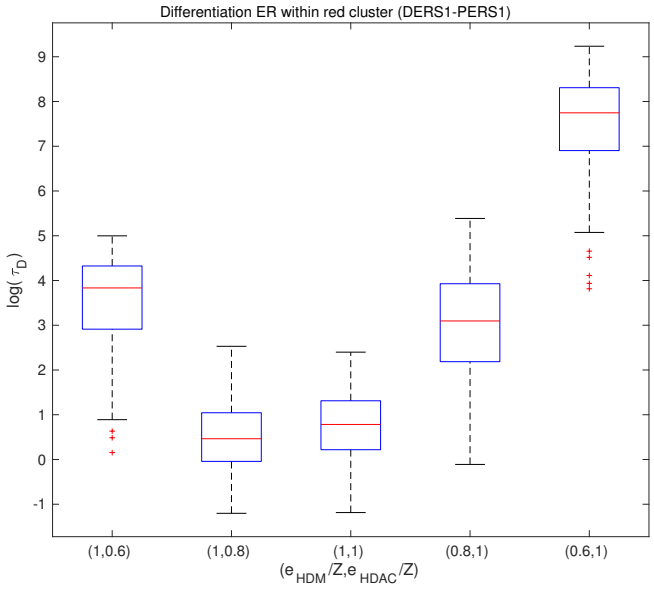


Figure 3.9: (Previous page). Plots showing the effect of the variation of HDM and HDAC on the statistics of the differentiation time (τ_D). We consider a base-line scenario where the number of HMEs is exactly equal to average, i.e. $e_{HDM} = e_{HDAC} = Z$. We then compare the simulation results obtained for different scenarios regarding the abundance of HDM and HDAC to the base-line scenario, i.e. by changing the values of (e_{HDM}, e_{HDAC}). Parameter values: $Z = 5$ and $Y = 15$. Other parameter values given in Tables A.15, A.16, **??** and **??**, Section A.2.5.

sion mechanisms and kinetics driving phenotypic robustness in a stem-lock model of pathological pluripotency. Our deconstruction of epigenetic plasticity and phenotypic malleability provides crucial insights into how pathological states of permanently acquired pluripotency can be therapeutically unlocked by exploiting epigenetic heterogeneity.

We have added an ER layer to previous approaches in which cell phenotypes were associated with the attractors of complex gene regulatory systems and their robustness, with the resilience of such attractors tuned by the presence of intrinsic noise, environmental fluctuations, and other disturbances [79, 155, 181, 58, 170, 127, 97].

Our approach is based on two main pillars: namely, a framework for the generation of the ensemble of ER systems, and a multiscale asymptotic analysis-based method for model reduction of the stochastic ER-GRN model (see Section 3.2.2). The ensemble generation method allows the definition of epi-phenotypes based on epigenetic-regulatory modes compatible with a given state of the whole ER-GRN system [53]. We initially chose an epi-phenotype in which the ER of differentiation gene(s) (DERS) is open/active whereas the ER of pluripotency gene(s) (PERS) is closed/silent. We then used Approximate Bayesian Computation (ABC) to generate an ensemble of DERS and PERS compatible with the above-mentioned phenotype (see Section 3.2.4, Fig. 3.2 and Chapter 2). With such an ensemble generated, we then proceeded to evaluate its hidden, intrinsic heterogeneity in terms of the physical properties of the ER-GRN systems. This approach is closely related to the notion of neutral networks formulated to analyse systems with genotype-phenotype maps [164, 165, 166]. By making a number of assumptions regarding separation of characteristic scales, we re-scaled both the variables and the parameters of the ER-GRN system, which allowed us to discriminate the underlying separation of time scales and consequently construct an asymptotic expansion, leading to a stochastic QSSA of the system. This approximation reduces a rather complex stochastic system (Eqs. (3.2.1)-(3.2.3)) to a hybrid, piece-wise deterministic Markov system (see Section 3.2.2). Furthermore, our model reduction procedure gives rise to an efficient and scalable, hybrid numerical method to simulate the ER-GRN system (see Section A.2.1.5). Although the model reduction was formulated for a GRN with an arbitrary number of mutually inhibiting genes, such a procedure is applicable to broader situations.

When analysing the behaviour of the mean-field limit of the GRN, it became apparent that, even in the simplest case considered involving a gene regulatory circuit of only two genes, the system exhibited a complex space that included several multi-stable phases. We observed a regime of tri-stability where the expected stem-lock (pluripotent) and differentiated steady-states coexist with a third state, the indecision state, in which the expression level of both genes is very low. From a developmental perspective, the latter state could serve the purpose of priming cells for differentiation, as the expression level of the pluripotency gene has decreased in a manner that could release repression upon the differentiation gene. Of note, the transitions between the different phases can be triggered by changes directly related to epigenetic regulation (i.e., cofactors of chromatin-modifying enzymes), which thereby act as *bona fide* molecular bridges connecting epigenetic and phenotypic plasticity by translating changes in ER states into variations of GRN states.

Having analysed the phase diagram of the GRN system and established its connection with ER (see Section 3.3.1), we have assessed the role of epigenetic heterogeneity in generating stem-lock pluripotent states. Such states can be viewed as examples of the so-called overly restricted epigenetic states, which present accentuated epigenetic barriers that block cell state transitions and which are biologically unable to disengage self-renewal pathways [51]. The opposite situation of so-called overly permissive epigenetic states is accompanied by lowered epigenetic barriers that allow the promiscuous sampling of alternative cell states [51]. Yamanaka originally appreciated the link between epigenetic heterogeneity and plasticity when aiming to explain the extremely low efficiency of somatic cell reprogramming at the population level [176]. We now know that an epigenetic predisposition to reprogramming fates exists in somatic cells and, therefore, the potential to acquire stem cell-like traits might in part reflect a pre-existing heterogeneity in cell states [129]. Furthermore, by perturbing the epigenetic state of somatic populations via inhibition of some epigenetic enzymes (e.g., the histone methyltransferase Ezh2, which catalyses repressive H3K27 methylation [159]), such heterogeneity can be harnessed to fine-tune the cellular response to reprogramming-to-pluripotency factors. Indeed, our findings support a scenario in which a sub-ensemble of ER systems with higher reprogramming potential pre-exists within the ensemble of ER systems compatible with a terminally differentiated cell state, and that such a sub-ensemble could be harnessed by targeting chromatin-modifying enzymes such as HDMs and HDACs.

A careful evaluation of our ensemble of ER systems (i.e., combinations of DERSs and PERSs) concerning stem-locked systems associated with differentiation-repressive ER states (see Sections 3.3.2, 3.3.3, and 3.3.4) has concluded that DERS heterogeneity had a stronger influence on such pluripotency-locked systems when compared with that of PERSs. Accordingly, we found that the ensemble of DERSs can be divided into three different clusters, with each one exhibiting distinct properties regarding stem locking. The so-called red cluster appeared to generate differentiation-permissive ER systems irrespective of their PERSs counterparts. By contrast, the so-called blue and green clusters contained DERSs yielding pluripotency-locked ER systems irrespective of their PERSs companions. In light of these findings, we conducted a detailed comparative analysis to uncover the underlying, statistically significant differences between DERSs within the differentiation-permissive sub-ensemble and those associated with the differentiation-repressive epigenetic states (see Section 3.3.2). This approach allowed us to detect which kinetic ER parameters were key to determine whether a DERS within a given system produces either permissive or re-

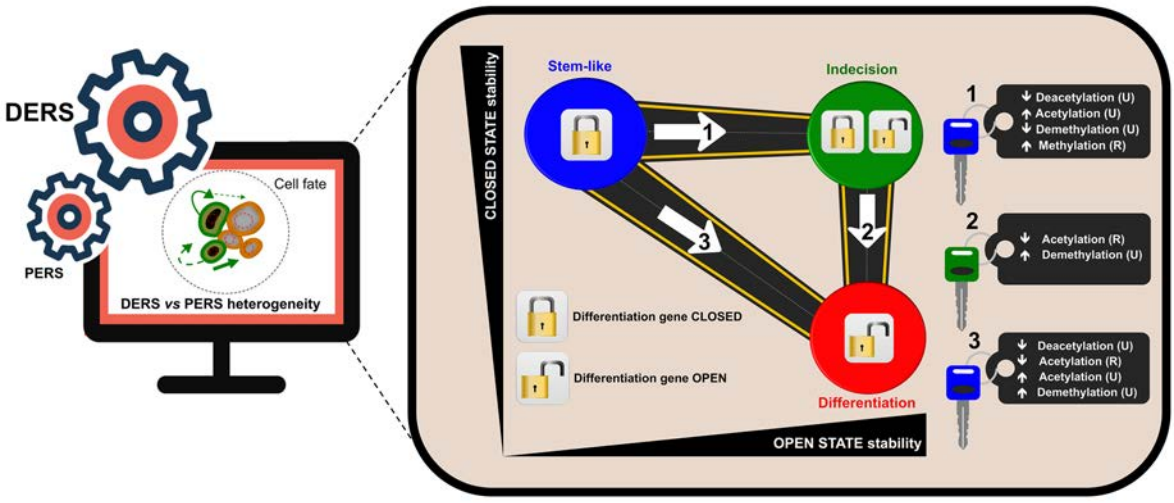


Figure 3.10: Strategies to unlock pluripotent stem-like states in ageing and cancer. Epigenetic regulation heterogeneity of differentiation genes (DERS), but not that of pluripotency genes (PERS), was predominantly in charge of the entry and exit decisions of the pluripotent stem-like states (blue). The application of a hybrid numerical method validated the likelihood of epigenetic heterogeneity-based strategies capable of unlocking and directing the transit from differentiation-refractory to differentiation-primed (red) epistates via kinetic changes in epigenetic factors. (Note: The epigenetic parameters regulating the entry into robust epi-states throughout the entire ER-GRN system revealed a regime of tri-stability in which pluripotent stem-like (blue) and differentiated (red) steady-states coexisted with a third indecisive (green) state). (R: Recruited; U: Unrecruited).

pressive ER differentiation systems (see Section 3.3.3). Remarkably, the elucidation of the identity of such critical regulators (see Fig. 3.10) would allow the formulation of strategies aimed to unlock differentiation-repressive epigenetic states by solely changing the values of such parameters (i.e., epigenetic cofactors). The feasibility of such strategies was verified by direct simulation of the ER-GRN system using our hybrid simulation method.

Our mathematical deconstruction of epigenetic plasticity suggests, for the first time, that epigenetic heterogeneity may underlie the predisposition of cell populations to pathological reprogramming processes that cause a permanent, locked

stem-like state disabled for reparative differentiation and prone to malignant transformation. Just as the potential of single somatic cells to generate pluripotent lineages reflects a pre-existing epigenetic heterogeneity permissive for the enhancement of reprogramming fates [110, 53, 129], we show that ER heterogeneity could generate a subpopulation in which robustness of the pluripotent phenotype is inherently boosted. Moreover, the uncovering of the epigenetic mechanisms underpinning such stem-locked states might help in the formulation of strategies capable, for instance, of unlocking the chronic epigenetic plasticity of senescence-damaged tissues while stimulating differentiation of such stem cell-like states to successfully achieve tissue rejuvenation (see Fig. 3.11). As we enter a new era of therapeutic approaches to target ageing per se (e.g., senolytic agents), our current mathematical modelling and computation simulation might pave the way to incorporate new systemic strategies based on the local availability of epigenetic cofactors capable of fine-tuning the senescence-inflammatory regulation of reparative reprogramming in ageing and cancer.

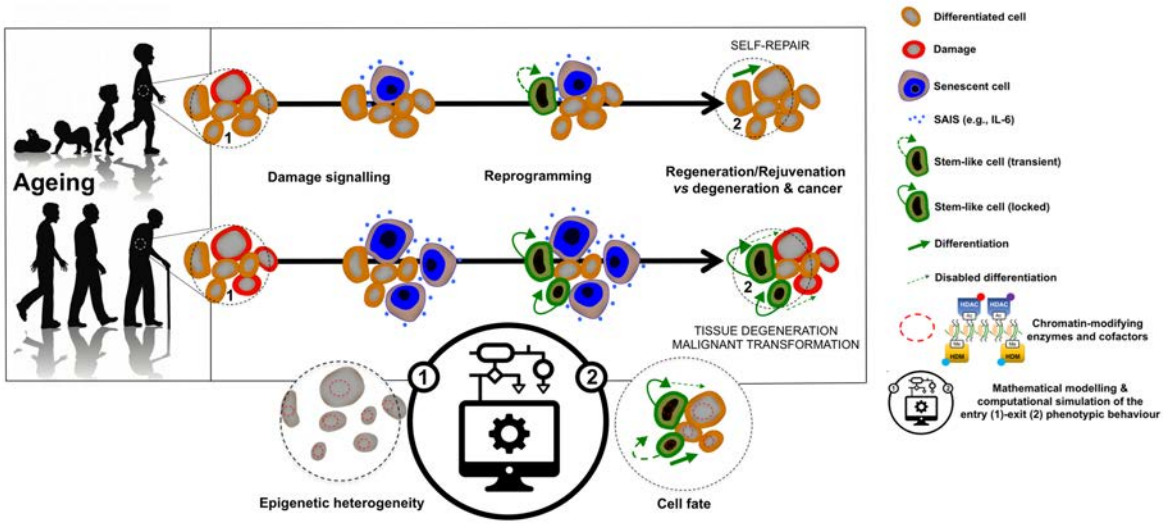


Figure 3.11: Physiological and pathological cell fate reprogramming: A mathematical approach. Reprogramming-like phenomena in response to damage signalling may constitute a reparative route through which human tissues respond to injury, stress, and disease via induction of a transient acquisition of epigenetic plasticity and phenotype malleability. However, tissue regeneration/rejuvenation should involve not only the transient epigenetic reprogramming of differentiated cells, but also the committed re-acquisition of the original or alternative committed cell fate. Chronic or unrestrained epigenetic plasticity would drive ageing/cancer phenotypes by impairing the repair or the replacement of damaged cells; such uncontrolled phenomena of *in vivo* reprogramming might also generate cancer-like cellular states. Accordingly, we now know that chronic senescence-associated inflammatory signalling might lock cells in highly plastic epigenetic states disabled for reparative differentiation and prone to malignant transformation. We herein introduce a first-in-class stochastic, multiscale reduction method of combined epigenetic regulation (ER)-gene regulatory network (GRN) to mathematically model and computationally simulate how ER heterogeneity regulates the entry-exit mechanisms and kinetics of physiological and pathological cell fate reprogramming. (SAIS: Senescence-associated inflammatory signalling)

Chapter 4

Ageing and epigenetic dysregulation effects in cell reprogramming

4.1 Summary

In this Chapter, we present an extension of the model formulated in Chapter 2. The current model does not consider any longer an infinite abundance in the media of the acetyl and methyl groups necessary for the acetylation and methylation modifications, respectively. Instead, we consider that these marks have a characteristic scale, Z. We show that the epigenetic regulatory system exhibits different behaviours depending on how Z is related to the other existing characteristic scales. Furthermore, the model presented below, allows us to account for the activity of not just the enzymes removing the epigenetic marks (HDMs and HDACs), but also the activity of those enzymes performing the epigenetic histone modifications, histone methyltransferases (HMTs) and histone acetyltransferases (HATs). The incorporation of these features leads to a more realistic model able to reproduce the plastic and resilient scenarios described in Chapter 2, as well as clarifying the connections between the appearance of the plastic cell behaviour and both ageing and epigenetic dysregulation. In particular, when introducing ageing-related factors in our model, such as upregulation of the activity of HATs for the pluripotency gene [80], we identify a healthy, physiological cell reprogramming, where cell dedifferentiation is reversible. When this effect occurs concomitantly with other epigenetic alterations, such as downregulation of the HDAC activity [122, 174], we find a pathological irreversible form of cell reprogramming as the epigenetic regulatory system enters a

pluripotency-locked state. These results allow us to formulate epigenetic strategies for rescueing the epigenetic regulatory system from the pathological cell reprogramming or avoiding it.

4.2 Model formulation and analysis

In this Section, we present the details regarding the extended model of epigenetic regulation. The current model relaxes the assumption that ER proceeds within an infinite abundance of acetylation and methylation marks (which was assumed in Chapter 2) by assuming that marks are finitely abundant. Under these conditions, epigenetic regulatory systems of the differentiation-promoting gene and the pluripotency-promoting gene compete for the marks. This feature allows to have a model where both genes are coupled, rather than having two replicas of the same stochastic model of ER (as studied in Chapter 2).

4.2.1 Epigenetic regulation model

As in the previous chapters, we formulate a model of ER acting upon a GRN of 2 mutually inhibiting genes, where one gene promotes differentiation and one gene promotes pluripotency. The ER model assumes that unmodified nucleosomes (U) can be modified by acquiring negative and positive marks. These marks can be added (removed) by recruited and unrecruited mechanisms under the presence of the corresponding histone (de)modifying-enzyme (see Section 2.2.1 for a complete description of this phenomena). Nucleosomes are considered to be either unmodified (U), acetylated (A) or methylated (M), and the dynamics of the model is described by the transitions rates between these three states. Direct transitions between A and M nucleosomes are considered highly unlikely, so the transitions are modelled according to the linear sequence

$$M \rightleftharpoons U \rightleftharpoons A,$$

(see Section 2.2.1 for further details).

Since we are studying a GRN with one gene promoting differentiation and one gene promoting pluripotency, we will assume that we have nucleosomes specific for each gene. Furthermore, we will also assume that the histone-modifying enzymes (HMTs, HATs, HDMs and HDACs) are gene-specific. The rationale for this assumption comes from the fact that experimental results point to the importance of local

changes in the chromatin patterns [33, 122, 48, 138, 104]. Specifically, different patterns of methylation and acetylation have been observed in pluripotency-promoting and differentiation-promoting genes. Inglés et al. [80] have found that centenarians overexpress pluripotency-associated genes, and Heyn et al. [153] have shown differences in the DNA methylome when comparing centenarians and newborn individuals. In particular, Heyn et al. find that tissue-specific genes are hypomethylated in centenarians when compared to octogenarians and newborns, i.e. there is a decrease in the methylation marks of differentiation-associated genes with ageing. Both findings remark the importance of being able to study addition and removal of methylation and acetylation marks separately for each gene and thus, the need for gene-specific enzymes. So we will denote with a subscript D or a subscript P, standing for differentiation and pluripotency, respectively, to which gene the nucleosome or enzyme belongs to.

In our modelling scenario, the dynamics is expressed in terms of the transition rates as depicted in Table 4.2 with variables defined in Table 4.1. More specifically, the reactions involved in our model are:

1. HDM-mediated demethylation:

$$M_D + HDM_D \leftrightarrows C_{M_D} \rightarrow U_D + HDM_D + Met$$
$$M_P + HDM_P \leftrightarrows C_{M_P} \rightarrow U_P + HDM_P + Met$$

   HDM-mediated demethylation can be both unrecruited, where the rates associated with each reaction are constant (see Table 4.2, reactions 1 to 3 (differentiation gene), reactions 25 to 27 (pluripotency gene)), and recruited, where all the associated rates are taken to be proportional to the number of A-nucleosomes (see Table 4.2, reactions 4 to 6 (differentiation gene), reactions 28 to 30 (pluripotency gene)).

2. HDAC-mediated deacetylation:

$$A_D + HDAC_D \leftrightarrows C_{A_D} \rightarrow U_D + HDAC_D + Acet$$
$$A_P + HDAC_P \leftrightarrows C_{A_P} \rightarrow U_P + HDAC_P + Acet$$

   HDAC-mediated deacetylation can be both unrecruited (see Table 4.2, reactions 7 to 9 (differentiation gene), reactions 37 to 39 (pluripotency gene)), or

recruited, where all the associated rates are proportional to the number of M-nucleosomes (see Table 4.2, reactions 10 to 12 (differentiation gene), reactions 40 to 42 (pluripotency gene)).

3. HMT-mediated methylation:

$$U_D + Met + HMT_D \leftrightarrows \tilde{C}_{M_D} \rightarrow M_D + HMT_D$$
$$U_P + Met + HMT_P \leftrightarrows \tilde{C}_{M_P} \rightarrow M_P + HMT_P$$

HMT-mediated methylation can also be unrecruited (Table 4.2, reactions 13 to 15 (differentiation gene), reactions 31 to 33 (pluripotency gene)) or recruited, in which case the associated rate is proportional to the number of methylated nucleosomes (Table 4.2, reactions 16 to 18 (differentiation gene), reactions 34 to 36 (pluripotency gene)).

4. HAT-mediated acetylation:

$$U_D + Acet + HAT_D \leftrightarrows \tilde{C}_{A_D} \rightarrow A_D + HAT_D$$
$$U_P + Acet + HAT_P \leftrightarrows \tilde{C}_{A_P} \rightarrow A_P + HAT_P$$

As with the previous processes, HAT-mediated acetylation can also be unrecruited (Table 4.2, reactions 19 to 21 (differentiation gene), reactions 43 to 45 (pluripotency gene)) or recruited, in which case the transition rate is proportional to the number of acetylated nucleosomes (Table 4.2, reactions 22 to 24 (differentiation gene), reactions 46 to 48 (pluripotency gene)).

Similarly to what we assumed in 2.2.1, we consider the scenario where both histone hypomethylation (low level of negative (M) marks) and hyperacetylation (high level of positive (A) marks) allow genes to be expressed when the associated transcription factors are present [100]. On the contrary, we relate histone hypermethylation and hypoacetylation to silent states where genes are not expressed even in the presence of the appropriate transcription factors. In this chapter, we focus on the conditions for the resilient and plastic scenarios to arise (as described in Section 2.3), as well as their robustness, particularly in connection with the abundance or activity of the histone modifying enzymes (HMEs) and the effects of ageing. As argued previously, changes in the abundance of HMEs are well justified due to the variability in the pool of cofactors employed by HMEs. Not only this, but also, the levels of activity of the HMEs change during ageing [48, 125, 122]

Variable	Description
X_1	Number of methylation marks (Met)
X_2	Number of acetylation marks ($Acet$)
X_4	Number of unmodified nucleosomes differentiation gene (U_D-nucleosome)
X_5	Number of methylated nucleosomes differentiation gene (M_D-nucleosome)
X_6	Number of acetylated nucleosomes differentiation gene (A_D-nucleosome)
X_7	Number of HDM enzyme molecules differentiation gene (HDM_D)
X_8	Number of methylated nucleosome diff. gene -HDM enzyme complexes (C_{M_D})
X_9	Number of HDAC enzyme molecules differentiation gene ($HDAC_D$)
X_{10}	Number of acetylated nucleosome diff. gene -HDAC enzyme complexes (C_{A_D})
X_{11}	Number of HMT enzyme molecules differentiation gene (HMT_D)
X_{12}	Number of methylated nucleosome diff. gene -HMT enzyme complexes ($\tilde{C}_{M_D}$)
X_{13}	Number of HAT enzyme molecules differentiation gene (HAT_D)
X_{14}	Number of acetylated nucleosome diff. gene -HAT enzyme complexes ($\tilde{C}_{A_D}$)
X_{15}	Number of methylated nucleosome plurip. gene -HDM enzyme complexes (C_{M_P})
X_{16}	Number of acetylated nucleosome plurip. gene -HDAC enzyme complexes (C_{A_P})
X_{17}	Number of methylated nucleosome plurip. gene -HMT enzyme complexes ($\tilde{C}_{M_P}$)
X_{18}	Number of acetylated nucleosome plurip. gene -HAT enzyme complexes ($\tilde{C}_{A_P}$)
X_{19}	Number of unmodified nucleosomes pluripotency gene (U_P-nucleosome)
X_{20}	Number of methylated nucleosomes pluripotency gene (M_P-nucleosome)
X_{21}	Number of acetylated nucleosomes pluripotency gene (A_P-nucleosome)
X_{22}	Number of HDM enzyme molecules pluripotency gene (HDM_P)
X_{23}	Number of HDAC enzyme molecules pluripotency gene ($HDAC_P$)
X_{24}	Number of HMT enzyme molecules pluripotency gene (HMT_P)
X_{25}	Number of HAT enzyme molecules pluripotency gene (HAT_P)

Table 4.1: Random variables

Transition rate	Event
$W_1(x) = k_1 X_5 X_7$	Formation of M-nucleosome diff.-HDM enzyme complex (unrecruited)
$W_2(x) = k_2 X_8$	M-nucleosome diff.-HDM enzyme complex splits (unrecruited)
$W_3(x) = k_3 X_8$	Demethylation and HDM enzyme release (unrecruited)
$W_4(x) = k_4 X_5 X_6 X_7$	Formation of M-nucleosome diff.-HDM enzyme complex (recruited)
$W_5(x) = k_5 X_6 X_8$	M-nucleosome diff.-HDM enzyme complex splits (recruited)
$W_6(x) = k_6 X_6 X_8$	Demethylation and HDM enzyme release (recruited)
$W_7(x) = k_7 X_6 X_9$	Formation of A-nucleosome diff.-HDAC enzyme complex (unrecruited)
$W_8(x) = k_8 X_{10}$	A-nucleosome diff.-HDAC enzyme complex splits (unrecruited)
$W_9(x) = k_9 X_{10}$	Deacetylation and HDAC enzyme release (unrecruited)
$W_{10}(x) = k_{10} X_6 X_5 X_9$	Formation of A-nucleosome diff.-HDAC enzyme complex (recruited)
$W_{11}(x) = k_{11} X_{10} X_5$	A-nucleosome diff.-HDAC enzyme complex splits (recruited)
$W_{12}(x) = k_{12} X_{10} X_5$	Deacetylation and HDAC enzyme release (recruited)
$W_{13}(x) = k_{13} X_4 X_1 X_{11}$	Formation of U-nucleosome diff.-HMT enzyme complex (unrecruited)
$W_{14}(x) = k_{14} X_{12}$	U-nucleosome diff.-HMT enzyme complex splits (unrecruited)
$W_{15}(x) = k_{15} X_{12}$	Methylation and HMT enzyme release (unrecruited)
$W_{16}(x) = k_{16} X_4 X_1 X_{11} X_5$	Formation of U-nucleosome diff.-HMT enzyme complex (recruited)
$W_{17}(x) = k_{17} X_{12} X_5$	U-nucleosome diff.-HMT enzyme complex splits (recruited)
$W_{18}(x) = k_{18} X_{12} X_5$	Methylation and HMT enzyme release (recruited)
$W_{19}(x) = k_{19} X_4 X_2 X_{13}$	Formation of U-nucleosome diff.-HAT enzyme complex (unrecruited)
$W_{20}(x) = k_{20} X_{14}$	U-nucleosome diff.-HAT enzyme complex splits (unrecruited)
$W_{21}(x) = k_{21} X_{14}$	Acetylation and HAT enzyme release (unrecruited)
$W_{22}(x) = k_{22} X_4 X_2 X_{13} X_6$	Formation of U-nucleosome diff.-HAT enzyme complex (recruited)
$W_{23}(x) = k_{23} X_{14} X_6$	U-nucleosome diff.-HAT enzyme complex splits (recruited)
$W_{24}(x) = k_{24} X_{14} X_6$	Acetylation and HAT enzyme release (recruited)

Table 4.2: Table continues in the following page

Table 4.2: Random processes and their transition rates. X_i are as in Table 4.1.

Transition rate	Event
$W_{25}(x) = k_{25}X_{20}X_{22}$	Formation of M-nucleosome plurip.-HDM enzyme complex (unrecruited)
$W_{26}(x) = k_{26}X_{15}$	M-nucleosome plurip. -HDM enzyme complex splits (unrecruited)
$W_{27}(x) = k_{27}X_{15}$	Demethylation and HDM enzyme release (unrecruited)
$W_{28}(x) = k_{28}X_{20}X_{21}X_{22}$	Formation of M-nucleosome plurip. -HDM enzyme complex (recruited)
$W_{29}(x) = k_{29}X_{21}X_{15}$	M-nucleosome plurip. -HDM enzyme complex splits (recruited)
$W_{30}(x) = k_{30}X_{21}X_{15}$	Demethylation and HDM enzyme release (recruited)
$W_{31}(x) = k_{31}X_{19}X_1X_{24}$	Formation of U-nucleosome plurip. -HMT enzyme complex (unrecruited)
$W_{32}(x) = k_{32}X_{17}$	U-nucleosome plurip. -HMT enzyme complex splits (unrecruited)
$W_{33}(x) = k_{33}X_{17}$	Methylation and HMT enzyme release (unrecruited)
$W_{34}(x) = k_{34}X_{19}X_1X_{24}X_{20}$	Formation of U-nucleosome plurip.-HMT enzyme complex (recruited)
$W_{35}(x) = k_{35}X_{17}X_{20}$	U-nucleosome plurip. -HMT enzyme complex splits (recruited)
$W_{36}(x) = k_{36}X_{17}X_{20}$	Methylation and HMT enzyme release (recruited)
$W_{37}(x) = k_{37}X_{21}X_{23}$	Formation of A-nucleosome plurip. -HDAC enzyme complex (unrecruited)
$W_{38}(x) = k_{38}X_{16}$	A-nucleosome plurip. -HDAC enzyme complex splits (unrecruited)
$W_{39}(x) = k_{39}X_{16}$	Deacetylation and HDAC enzyme release (unrecruited)
$W_{40}(x) = k_{40}X_{21}X_{20}X_{23}$	Formation of A-nucleosome plurip. -HDAC enzyme complex (recruited)
$W_{41}(x) = k_{41}X_{16}X_{20}$	A-nucleosome plurip. -HDAC enzyme complex splits (recruited)
$W_{42}(x) = k_{42}X_{16}X_{20}$	Deacetylation and HDAC enzyme release (recruited)
$W_{43}(x) = k_{43}X_{19}X_2X_{25}$	Formation of U-nucleosome plurip. -HAT enzyme complex (unrecruited)
$W_{44}(x) = k_{44}X_{18}$	U-nucleosome plurip. -HAT enzyme complex splits (unrecruited)
$W_{45}(x) = k_{45}X_{18}$	Acetylation and HAT enzyme release (unrecruited)
$W_{46}(x) = k_{46}X_{19}X_2X_{25}X_{21}$	Formation of U-nucleosome plurip. -HAT enzyme complex (recruited)
$W_{47}(x) = k_{47}X_{18}X_{21}$	U-nucleosome plurip. -HAT enzyme complex splits (recruited)
$W_{48}(x) = k_{48}X_{18}X_{21}$	Acetylation and HAT enzyme release (recruited)

The transition rates (Table 4.2) associated with each of these reactions are modelled using the law of mass action [59]. This model is an extension of the one formulated in Chapter 2 (see Section 2.2.1), which, in turn, is based on previous models by Menéndez et al. [110]. This model extends the model previously presented in two ways. On one side, it accounts explicitly for the activity of four types of HMEs: the above-modelled HDMs and HDACs, but also, the presence of HMTs and HATs. Furthermore, the modelling approach adopted here treats both removal and addition of M and A marks by representing them using the Michaelis-Menten model of enzymatic catalysis. This allows us to analyse how changes in the HMEs abundance can directly affect the behaviour of the ER system, which was not possible in previous models [110]. In particular, we aim to study the effects of altered activity of HMEs and their connection with ageing-related cell reprogramming. Ageing is also a new feature incorporated into this modelling framework. The precise form in which ageing is considered will be fully described later.

4.2.2 Mean-field equations and QSSA

As we have done previously (see Chapter 2), in order to gain some insight into the behaviour of the stochastic ER model, we analyse its mean-field limit regarding time scale separation and the quasi-steady state approximation (QSSA). For a full account of the technicalities we refer to [1, 40].

The mean-field equations, which describe the time evolution of the ensemble average of the variables X_i (see Table 4.1), related to the stochastic system with rates given in Table 4.2 are:

$$\frac{dQ_i}{dt} = \sum_{j=1}^{48} r_{j,i} W_j(\mathbf{Q}), \tag{4.2.1}$$

where $\mathbf{Q}$ is a vector whose entries, Q_i, are defined as $Q_i \equiv \langle X_i \rangle$ (i.e., its average value). In order to proceed further, we assume that the variables describing the system are divided into three groups according to their characteristic scales. The different characteristic scales correspond to the characteristic number of epigenetic marks, Z, the characteristic number of sites, S, and the characteristic scale for the HMEs and complexes, E. More specifically, we consider the situation where the subset of chemical species X_i, with $i = 1, 2$, corresponding to the marks, scales as $X_i = Zx_i$, where $x_i = O(1)$, the subset of chemical species X_i, with $i = 4, 5, 6, 19, 20, 21$, corresponding to the substrates, scales as $X_i = Sx_i$, where $x_i = O(1)$ and the remaining

species are such that X_i, with $i = 7, \ldots, 18, 22, 23, 24, 25$, corresponding to the enzymes and complexes, scale as $X_i = Ex_i$, where $x_i = O(1)$. We further assume that S and E must be such that $\epsilon = \frac{E}{S} \ll 1$ (Briggs-Haldane approximation), which is a standard assumption in enzyme kinetics (see [88]). Of key importance is the relationship between the characteristic scale Z and the characteristic scale S. Depending on this, the system will have a media with abundance or non-abundance of epigenetic marks. Later, we will deal with this discussion. It is noteworthy to mention that when $\frac{S}{Z} \to 0$ (i.e. overabundance of marks), the current model reduces to the model of Chapter 2, where marks where assumed to be infinite.

In order to proceed further, we define the following scale transformation for the transition rates in Table 4.2: $W_j(\mathbf{Q}) = k_{16}S^2EZ\omega_j(\mathbf{q})$, where for the rescaling we have used the transition rate of highest order. We further rescale the time variable so that a dimensionless variable, τ, is defined as $\tau = k_{16}SEZt$. It is now straightforward to verify that, upon rescaling, the mean-field equations become:

$$\frac{Z}{S}\frac{dq_i}{d\tau} = \sum_{j=1}^{48} r_{j,i}\omega_j(\mathbf{q}),\ i = 1, 2, \tag{4.2.2}$$

$$\frac{dq_i}{d\tau} = \sum_{j=1}^{48} r_{j,i}\omega_j(\mathbf{q}),\ i = 4, 5, 6, 19, 20, 21, \tag{4.2.3}$$

$$\epsilon\frac{dq_i}{d\tau} = \sum_{j=1}^{48} r_{j,i}\omega_j(\mathbf{q}),\ i = 7, \ldots, 18, 22, \ldots, 25, \tag{4.2.4}$$

with $\epsilon = E/S$. Eqs. (4.2.2)-(4.2.4) show that if $\epsilon = E/S \ll 1$ holds, then, the mean-field equations (Eqs. (4.2.2)-(4.2.4)) naturally display multiple scales structure, which we will exploit to simplify our analysis by means of a QSSA [137]. Furthermore, depending on the ratio $\varepsilon_4 := \frac{S}{Z}$, further simplification will also be possible. We will mainly distinguish three relevant cases: $\varepsilon_4 \ll 1$, $\varepsilon_4 = O(1)$ and $\varepsilon_4 \gg 1$. Notice that we denote this ratio with a subscript 4 for our convenience, since in this way it is easy to see that it is the ratio about the abundance of epigenetic marks, which is one of the features introduced in this Chapter 4.

Regarding Eqs. (4.2.2)-(4.2.4), the QSS approximation consists on assuming that $\epsilon\frac{dq_i}{d\tau} \simeq 0$ which leads to a differential-algebraic system of equations which provides us with the QSSA. In our particular case, the QSSA is given by:

$$\begin{aligned}
\frac{Z}{S}\frac{dq_1}{d\tau} &= (\kappa_3+\kappa_6q_6)q_8-(\kappa_{13}+q_5)q_1q_{11}q_4+(\kappa_{14}+\kappa_{17}q_5)q_{12}+ \\
&+ (\kappa_{27}+\kappa_{30}q_{21})q_{15}-(\kappa_{31}+\kappa_{34}q_{20})q_1q_{24}q_{19}+(\kappa_{32}+\kappa_{35}q_{20})q_{17} \qquad (4.2.5)\\
\frac{Z}{S}\frac{dq_2}{d\tau} &= (\kappa_9+\kappa_{12}q_5)q_{10}-(\kappa_{19}+\kappa_{22}q_6)q_2q_{13}q_4+(\kappa_{20}+\kappa_{23}q_6)q_{14}+ \\
&+ (\kappa_{39}+\kappa_{42}q_{20})q_{16}-(\kappa_{43}+\kappa_{46}q_{21})q_2q_{25}q_{19}+(\kappa_{44}+\kappa_{47}q_{21})q_{18} \qquad (4.2.6)\\
\frac{dq_4}{d\tau} &= (\kappa_3+\kappa_6q_6)q_8-(\kappa_{13}+q_5)q_1q_{11}q_4+(\kappa_{14}+\kappa_{17}q_5)q_{12}+(\kappa_9+\kappa_{12}q_5)q_{10}- \\
&- (\kappa_{19}+\kappa_{22}q_6)q_2q_{13}q_4+(\kappa_{20}+\kappa_{23}q_6)q_{14} \qquad (4.2.7)\\
\frac{dq_5}{d\tau} &= -(\kappa_1+\kappa_4q_6)q_5q_7+(\kappa_2+\kappa_5q_6)q_8+(\kappa_{15}+\kappa_{18}q_5)q_{12} \qquad (4.2.8)\\
\frac{dq_6}{d\tau} &= -(\kappa_7+\kappa_{10}q_5)q_6q_9+(\kappa_8+\kappa_{11}q_5)q_{10}+(\kappa_{21}+\kappa_{24}q_6)q_{14} \qquad (4.2.9)\\
\frac{dq_{19}}{d\tau} &= (\kappa_{27}+\kappa_{30}q_{21})q_{15}-(\kappa_{31}+\kappa_{34}q_{20})q_1q_{24}q_{19}+(\kappa_{32}+\kappa_{35}q_{20})q_{17}+ \\
&+ (\kappa_{39}+\kappa_{42}q_{20})q_{16}-(\kappa_{43}+\kappa_{46}q_{21})q_2q_{25}q_{19}+(\kappa_{44}+\kappa_{47}q_{21})q_{18} \qquad (4.2.10)\\
\frac{dq_{20}}{d\tau} &= -(\kappa_{25}+\kappa_{28}q_{21})q_{20}q_{22}+(\kappa_{26}+\kappa_{29}q_{21})q_{15}+(\kappa_{33}+\kappa_{36}q_{20})q_{17} \qquad (4.2.11)\\
\frac{dq_{21}}{d\tau} &= -(\kappa_{37}+\kappa_{40}q_{20})q_{21}q_{23}+(\kappa_{38}+\kappa_{41}q_{20})q_{16}+(\kappa_{45}+\kappa_{48}q_{21})q_{18} \qquad (4.2.12)\\
q_7 &= \frac{p_{HDM_D}(\kappa_2+\kappa_3+(\kappa_5+\kappa_6)q_6)}{(\kappa_1+\kappa_4q_6)q_5+\kappa_2+\kappa_3+(\kappa_5+\kappa_6)q_6} \qquad (4.2.13)\\
q_8 &= \frac{p_{HDM_D}(\kappa_1+\kappa_4q_6)q_5}{(\kappa_1+\kappa_4q_6)q_5+\kappa_2+\kappa_3+(\kappa_5+\kappa_6)q_6} \qquad (4.2.14)\\
q_9 &= \frac{p_{HDAC_D}(\kappa_8+\kappa_9+(\kappa_{11}+\kappa_{12})q_5)}{(\kappa_7+\kappa_{10}q_5)q_6+\kappa_8+\kappa_9+(\kappa_{11}+\kappa_{12})q_5} \qquad (4.2.15)\\
q_{10} &= \frac{p_{HDAC_D}(\kappa_7+\kappa_{10}q_5)q_6}{(\kappa_7+\kappa_{10}q_5)q_6+\kappa_8+\kappa_9+(\kappa_{11}+\kappa_{12})q_5} \qquad (4.2.16)\\
q_{11} &= \frac{p_{HMT_D}(\kappa_{14}+\kappa_{15}+(\kappa_{17}+\kappa_{18})q_5)}{(\kappa_{13}+q_5)q_4q_1+\kappa_{14}+\kappa_{15}+(\kappa_{17}+\kappa_{18})q_5} \qquad (4.2.17)\\
q_{12} &= \frac{p_{HMT_D}(\kappa_{13}+q_5)q_4q_1}{(\kappa_{13}+q_5)q_4q_1+\kappa_{14}+\kappa_{15}+(\kappa_{17}+\kappa_{18})q_5} \qquad (4.2.18)\\
q_{13} &= \frac{p_{HAT_D}(\kappa_{20}+\kappa_{21}+(\kappa_{23}+\kappa_{24})q_6)}{(\kappa_{19}+\kappa_{22}q_6)q_4q_2+\kappa_{20}+\kappa_{21}+(\kappa_{23}+\kappa_{24})q_6} \qquad (4.2.19)
\end{aligned}$$

$$q_{14} = \frac{p_{HATD}(\kappa_{19} + \kappa_{22}q_6)q_4q_2}{(\kappa_{19} + \kappa_{22}q_6)q_4q_2 + \kappa_{20} + \kappa_{21} + (\kappa_{23} + \kappa_{24})q_6} \tag{4.2.20}$$

$$q_{15} = \frac{p_{HDMP}(\kappa_{25} + \kappa_{28}q_{21})q_{20}}{(\kappa_{25} + \kappa_{28}q_{21})q_{20} + \kappa_{26} + \kappa_{27} + (\kappa_{29} + \kappa_{30})q_{21}} \tag{4.2.21}$$

$$q_{16} = \frac{p_{HDACP}(\kappa_{37} + \kappa_{40}q_{20})q_{21}}{(\kappa_{37} + \kappa_{40}q_{20})q_{21} + \kappa_{38} + \kappa_{39} + (\kappa_{41} + \kappa_{42})q_{20}} \tag{4.2.22}$$

$$q_{17} = \frac{p_{HMTP}(\kappa_{31} + \kappa_{34}q_{20})q_{19}q_1}{(\kappa_{31} + \kappa_{34}q_{20})q_{19}q_1 + \kappa_{32} + \kappa_{33} + (\kappa_{35} + \kappa_{36})q_{20}} \tag{4.2.23}$$

$$q_{18} = \frac{p_{HATP}(\kappa_{43} + \kappa_{46}q_{21})q_{19}q_2}{(\kappa_{43} + \kappa_{46}q_{21})q_{19}q_2 + \kappa_{44} + \kappa_{45} + (\kappa_{47} + \kappa_{48})q_{21}} \tag{4.2.24}$$

$$q_{22} = \frac{p_{HDMP}(\kappa_{26} + \kappa_{27} + (\kappa_{29} + \kappa_{30})q_{21})}{(\kappa_{25} + \kappa_{28}q_{21})q_{20} + \kappa_{26} + \kappa_{27} + (\kappa_{29} + \kappa_{30})q_{21}} \tag{4.2.25}$$

$$q_{23} = \frac{p_{HDACP}(\kappa_{38} + \kappa_{39} + (\kappa_{41} + \kappa_{42})q_{20})}{(\kappa_{37} + \kappa_{40}q_{20})q_{21} + \kappa_{38} + \kappa_{39} + (\kappa_{41} + \kappa_{42})q_{20}} \tag{4.2.26}$$

$$q_{24} = \frac{p_{HMTP}(\kappa_{32} + \kappa_{33} + (\kappa_{35} + \kappa_{36})q_{20})}{(\kappa_{31} + \kappa_{34}q_{20})q_{19}q_1 + \kappa_{32} + \kappa_{33} + (\kappa_{35} + \kappa_{36})q_{20}} \tag{4.2.27}$$

$$q_{25} = \frac{p_{HATP}(\kappa_{44} + \kappa_{45} + (\kappa_{47} + \kappa_{48})q_{21})}{(\kappa_{43} + \kappa_{46}q_{21})q_{19}q_2 + \kappa_{44} + \kappa_{45} + (\kappa_{47} + \kappa_{48})q_{21}} \tag{4.2.28}$$

where the re-scaled parameters κ_j are defined in Table 4.3, and where we have used the following conservation laws,

$$q_7 + q_8 = p_{HDMD}, \text{ where } p_{HDMD} = \frac{HDM_{DT}}{E},$$

$$q_9 + q_{10} = p_{HDACD}, \text{ where } p_{HDACD} = \frac{HDAC_{DT}}{E},$$

$$q_{11} + q_{12} = p_{HMTD}, \text{ where } p_{HMTD} = \frac{HMT_{DT}}{E},$$

$$q_{13} + q_{14} = p_{HATD}, \text{ where } p_{HATD} = \frac{HAT_{DT}}{E},$$

$$q_{15} + q_{22} = p_{HDMP}, \text{ where } p_{HDMP} = \frac{HDM_{PT}}{E},$$

$$q_{16} + q_{23} = p_{HDACP}, \text{ where } p_{HDACP} = \frac{HDAC_{PT}}{E},$$

$$q_{17} + q_{24} = p_{HMTP}, \text{ where } p_{HMTP} = \frac{HMT_{PT}}{E},$$

$$q_{18} + q_{25} = p_{HATP}, \text{ where } p_{HATP} = \frac{HAT_{PT}}{E},$$

with the subscript T denoting total, e.g. HDM_{DT} denotes the total number of HDM molecules for the differentiation gene (that is, those in its free form, q_7, and those bound forming a complex, q_8). These conservation laws account for the fact that the total number of enzyme molecules, i.e. the enzyme molecules in their free form and those forming a complex, must be constant. Hence, the quantities $p_{HME_{gene}}$ are defined as the total number of molecules of the HME affecting a particular gene over its characteristic scale, E. Hence, E denotes the average of abundance of the HMEs which, for simplicity, has assumed to be the same for all HMEs for both genes. This result opens interesting avenues to investigate, since ageing appears to alter the activity of some of these HMEs [175, 125, 115, 33, 48, 138, 104]. Therefore, our model allows us in a natural manner to explore the effects of these anomalies on the stability of epigenetic regulatory states because the values of $p_{HME_{gene}}$ are undetermined, so they can be used as parameters.

In order to reduce further the dimension of our system, Eqs. (4.2.5)-(4.2.28), we can also employ other conservation laws corresponding to the number of nucleosomes at the promoter region of the differentiation and the pluripotency gene, N_D and N_P, respectively. After rescaling, these conservation laws result into,

$$q_4 + q_5 + q_6 + \frac{E}{S}(q_8 + q_{10} + q_{12} + q_{14}) = \frac{N_D}{S}, \tag{4.2.29}$$

$$q_{19} + q_{20} + q_{21} + \frac{E}{S}(q_{15} + q_{16} + q_{17} + q_{18}) = \frac{N_P}{S}, \tag{4.2.30}$$

where using the assumption that $\frac{E}{S} \approx 0$, Eqs. (4.2.29)-(4.2.30) can be approximated by

$$q_4 + q_5 + q_6 \approx \frac{N_D}{S},$$

$$q_{19} + q_{20} + q_{21} \approx \frac{N_P}{S}.$$

For simplicity, and without loss of generality, we will assume that $\frac{N_D}{S} \approx 1$ and $\frac{N_P}{S} \approx 1$. Therefore, we obtain that

$$q_4 \approx 1 - q_5 - q_6, \tag{4.2.31}$$

and

$$q_{19} \approx 1 - q_{20} - q_{21}. \tag{4.2.32}$$

Table 4.3: Re-scaled variables and dimensionless parameters

Rescaled variables	Dimensionless parameters
$q_1 = X_1/Z$	$\kappa_1 = k_1/(k_{16}SZ)$, $\kappa_2 = k_2/(k_{16}S^2Z)$, $\kappa_3 = k_3/(k_{16}S^2Z)$
$q_2 = X_2/Z$	$\kappa_4 = k_4/(k_{16}Z)$, $\kappa_5 = k_5/(k_{16}SZ)$
$q_4 = X_4/S$	$\kappa_6 = k_6/(k_{16}SZ)$, $\kappa_7 = k_7/(k_{16}SZ)$
$q_5 = X_5/S$	$\kappa_8 = k_8/(k_{16}S^2Z)$, $\kappa_9 = k_9/(k_{16}S^2Z)$
$q_6 = X_6/S$	$\kappa_{10} = k_{10}/(k_{16}Z)$, $\kappa_{11} = k_{11}/(k_{16}SZ)$
$q_7 = X_7/E$	$\kappa_{12} = k_{12}/(k_{16}SZ)$, $\kappa_{13} = k_{13}/(k_{16}S)$
$q_8 = X_8/E$	$\kappa_{14} = k_{14}/(k_{16}S^2Z)$, $\kappa_{15} = k_{15}/(k_{16}S^2Z)$
$q_9 = X_9/E$	$\kappa_{17} = k_{17}/(k_{16}SZ)$, $\kappa_{18} = k_{18}/(k_{16}SZ)$
$q_{10} = X_{10}/E$	$\kappa_{19} = k_{19}/(k_{16}S)$, $\kappa_{20} = k_{20}/(k_{16}S^2Z)$
$q_{11} = X_{11}/E$	$\kappa_{21} = k_{21}/(k_{16}S^2Z)$, $\kappa_{22} = k_{22}/(k_{16})$
$q_{12} = X_{12}/E$	$\kappa_{23} = k_{23}/(k_{16}SZ)$, $\kappa_{24} = k_{24}/(k_{16}SZ)$
$q_{13} = X_{13}/E$	$\kappa_{25} = k_{25}/(k_{16}SZ)$, $\kappa_{26} = k_{26}/(k_{16}S^2Z)$
$q_{14} = X_{14}/E$	$\kappa_{27} = k_{27}/(k_{16}S^2Z)$, $\kappa_{28} = k_{28}/(k_{16}Z)$
$q_{15} = X_{15}/E$	$\kappa_{29} = k_{29}/(k_{16}SZ)$, $\kappa_{30} = k_{30}/(k_{16}SZ)$
$q_{16} = X_{16}/E$	$\kappa_{31} = k_{31}/(k_{16}S)$, $\kappa_{32} = k_{32}/(k_{16}S^2Z)$
$q_{17} = X_{17}/E$	$\kappa_{33} = k_{33}/(k_{16}S^2Z)$, $\kappa_{34} = k_{34}/(k_{16})$
$q_{18} = X_{18}/E$	$\kappa_{35} = k_{35}/(k_{16}SZ)$, $\kappa_{36} = k_{36}/(k_{16}SZ)$
$q_{19} = X_{19}/S$	$\kappa_{37} = k_{37}/(k_{16}SZ)$, $\kappa_{38} = k_{38}/(k_{16}S^2Z)$
$q_{20} = X_{20}/S$	$\kappa_{39} = k_{39}/(k_{16}S^2Z)$, $\kappa_{40} = k_{40}/(k_{16}Z)$
$q_{21} = X_{21}/S$	$\kappa_{41} = k_{41}/(k_{16}SZ)$, $\kappa_{42} = k_{42}/(k_{16}SZ)$
$q_{22} = X_{22}/E$	$\kappa_{43} = k_{43}/(k_{16}S)$, $\kappa_{44} = k_{44}/(k_{16}S^2Z)$
$q_{23} = X_{23}/E$	$\kappa_{45} = k_{45}/(k_{16}S^2Z)$, $\kappa_{46} = k_{46}/(k_{16})$
$q_{24} = X_{24}/E$	$\kappa_{47} = k_{47}/(k_{16}SZ)$, $\kappa_{48} = k_{48}/(k_{16}SZ)$
$q_{25} = X_{25}/E$	$\tau = k_{16}SEZt$, $\epsilon = E/S$, $\varepsilon_4 = S/Z$

A quick look through the system, Eqs. (4.2.5)-(4.2.28), allows us to see that depending on the ratio $\frac{S}{Z}$ different behaviours will arise, which will imply other possible reductions of the size of the algebraic-differential system. Let us focus now on the analysis of the system regarding the ratio $\varepsilon_4 = \frac{S}{Z}$. We will consider three possible scenarios:

1. $Z \gg S$ ($\varepsilon_4 \ll 1$). When taking $Z \gg S$, we are assuming that the characteristic scale for the epigenetic marks, Z, is much larger than the characteristic scale for the sites, S. This assumption corresponds to the scenario considered in Section 2.2.1, where the marks have not been modelled explicitly because they have never been considered a limiting factor. Therefore, when taking $\varepsilon_4 = 0$ we expect to obtain a model very similar to the one previously studied in Chapter 2. This scenario is essential, since this will allow us to find parameter values for this model, so as to have the same base-line behaviour (bistability in a determined region of the phase space, monostability at the normal enzymatic level) as with the previous model. Hence, the limit of setting $\varepsilon_4 = 0$ will be the starting point for our discussion.

2. $Z \approx S$, i.e. $\varepsilon_4 = O(1)$. Taking as reference the behaviour of the system for $\varepsilon_4 = 0$, we continue the solutions found by increasing ε_4. Modifying ε_4 in this way corresponds to moving from a media where marks are infinitely abundant ($\varepsilon_4 = 0$) to a media where marks are as abundant as sites ($\varepsilon_4 = 1$), covering the regime of a media where marks are still abundant, but finite ($0 < \varepsilon_4 < 1$).

3. $Z \ll S$ ($\varepsilon_4 \gg 1$). In this case, epigenetic marks are much less abundant than nucleosomes, which implies that the sites cannot be all modified. This corresponds to a regime where the majority of nucleosomes will be unmodified, that is, with neither methylation nor acetylation marks. Since we are interested in studying the open(closed) state of the genes, which are characterised by high(low) levels of acetylation and low(high) levels of methylation, we need that marks are, at least, of the same order as sites. Otherwise, we would not be able to distinguish open and closed states. This limit thus provides no relevant information for the problem we are studying.

We now consider in more detail the implications of regimes 1 and 2 on the behaviour of Eqs. (4.2.5)-(4.2.28). It is noteworthy that these different scalings for the ratio $\frac{S}{Z}$ have a clear biological interpretation in terms of ageing. Ageing is known

to alter the methylation and acetylation patterns. Therefore, by changing the abundance of epigenetic marks, we could induce some of the patterns observed in ageing. For instance, in a media where marks are scarce, hypoacetylated or hypomethylated scenarios will be forced to be adopted. According to our model notation, reducing the abundance of the marks in the media corresponds to moving from $\varepsilon_4 = 0$ to $0 < \varepsilon_4 \leq 1$.

4.2.2.1 Case 1: $\varepsilon_4 = 0$ ($Z \gg S$)

Taking $\varepsilon_4 = 0$ means that marks are infinitely abundant in the media (previously studied model, where the marks where not explicitly modelled, see Section 2.2.1). In particular, when setting $\frac{S}{Z} = 0$ in Eqs. (4.2.5)-(4.2.6) gives us

$$\frac{dq_1}{d\tau} \approx 0 \Rightarrow q_1(\tau) = K_1,$$
$$\frac{dq_2}{d\tau} \approx 0 \Rightarrow q_2(\tau) = K_2,$$

with K_1 and K_2 constant values. Furthermore, when using the conservation law for the methylation marks, we have that

$$q_1 + \frac{S}{Z}(q_5 + q_{20}) + \frac{E}{Z}(q_8 + q_{12} + q_{15} + q_{17}) = \frac{Met_T}{Z},$$

where using the assumption that both $S/Z = 0$ and $E/Z = 0$, leads to

$$q_1 \approx \frac{Met_T}{Z}. \tag{4.2.33}$$

For simplicity, we will denote the ratio $p_{Met} =: \frac{Met_T}{Z}$. Hence, we can conclude that when $Z \gg S$ holds, then, $q_1 \approx p_{Met}$ (a constant value). An analogous discussion for q_2 will lead us to the same result, so we write $q_2 \approx p_{Acet}$.

In this case, the system of Eqs. (4.2.5)-(4.2.28) is reduced to:

$$\frac{dq_5}{d\tau} = -(\kappa_1 + \kappa_4 q_6)q_5 q_7 + (\kappa_2 + \kappa_5 q_6)q_8 + (\kappa_{15} + \kappa_{18} q_5)q_{12} \tag{4.2.34}$$
$$\frac{dq_6}{d\tau} = -(\kappa_7 + \kappa_{10} q_5)q_6 q_9 + (\kappa_8 + \kappa_{11} q_5)q_{10} + \\ + (\kappa_{21} + \kappa_{24} q_6)q_{14} \tag{4.2.35}$$

$$\begin{aligned}\frac{dq_{20}}{d\tau} &= -(\kappa_{25}+\kappa_{28}q_{21})q_{20}q_{22}+(\kappa_{26}+\kappa_{29}q_{21})q_{15}+\\ &+ (\kappa_{33}+\kappa_{36}q_{20})q_{17} \qquad (4.2.36)\\ \frac{dq_{21}}{d\tau} &= -(\kappa_{37}+\kappa_{40}q_{20})q_{21}q_{23}+(\kappa_{38}+\kappa_{41}q_{20})q_{16}+\\ &+ (\kappa_{45}+\kappa_{48}q_{21})q_{18} \qquad (4.2.37)\end{aligned}$$

together with the algebraic equations, Eq. (4.2.13)-(4.2.28), Eqs. (4.2.31)-(4.2.32) and $q_1 = p_{Met}$, $q_2 = p_{Acet}$.

Interestingly, in this limit, as epigenetic marks are abundant, HMEs do not need to compete for them, which implies that the system described by Eqs. (4.2.34)-(4.2.37) decouples into two independent subsystems. One of these subsystems describes the behaviour of the differentiation gene, Eqs. (4.2.34)-(4.2.35), and the other subsystem corresponds to the dynamics of the pluripotency gene, Eqs. (4.2.36)-(4.2.37).

4.2.2.2 Case 2: $\varepsilon_4 = O(1)$ ($Z \approx S$)

Consider now the case where $S/Z = O(1)$, i.e. $S = \varepsilon_4 Z$, with $\varepsilon_4 \in (0, 1]$. We wish now to study the system for $0 < \varepsilon_4 \leq 1$. Recall that we are still assuming that $E \ll S$ ($\epsilon \approx 0$).

The main difference between the case $\varepsilon_4 = 0$ and $0 < \varepsilon_4 \leq 1$ is that we can no longer assume that the marks are infinitely abundant in the media. If we go back to the system described by Eqs. (4.2.5)-(4.2.28), we can see that the first two equations, Eqs. (4.2.5)-(4.2.6), are the ones corresponding to the time evolution of the epigenetic marks. In the previous case, we have been able to use the multiple-scale assumption and eliminate these two equations, by assuming that marks were constant over time. Now, we have that marks are of the same order as sites. At first, this fact could seem to imply that the system to study will be 6-dimensional (rather than 4-dimensional). However, using the conservation laws for the total number of marks, we will be able to introduce the nonconstant behaviour of q_1 and q_2 without needing to increase the dimension of the problem. Specifically,

$$q_1 + \frac{S}{Z}(q_5 + q_{20}) + \frac{E}{S}(q_8 + q_{15} + q_{12} + q_{17}) = Met_T,$$

where using that $S/Z = \varepsilon_4$ and that $E/S \approx 0$, we obtain the following expression

$$\varepsilon_4(q_5 + q_{20}) + q_1 \approx p_{Met},$$

from where it is straightforward to obtain an expression for q_1,

$$q_1 \approx p_{Met} - \varepsilon_4(q_5 + q_{20}), \tag{4.2.38}$$

where $p_{Met} = \frac{Met_T}{Z}$. Using the analogous conservation law for the acetylation marks, we can derive the expression for q_2,

$$q_2 \approx p_{Acet} - \varepsilon_4(q_6 + q_{21}), \tag{4.2.39}$$

where $p_{Acet} = \frac{Acet_T}{Z}$.

Within the limit $\varepsilon_4 = O(1)$, the behaviour of q_i, $i = 7, \ldots, 18, 22, \ldots, 25$ remains unchanged because $\epsilon = \frac{E}{S} \ll 1$ still holds. Therefore, as in the previous case, we just need to study the time evolution for q_5, q_6, q_{20} and q_{21}, by means of the system described by Eqs. (4.2.34)-(4.2.37), which is supplemented with the algebraic expressions given by Eqs. (4.2.13)-(4.2.28), Eqs. (4.2.31)-(4.2.32) and Eqs. (4.2.38)-(4.2.39).

When assuming $\varepsilon_4 = O(1)$, the algebraic expressions for q_{11}, q_{12}, q_{13}, q_{14}, q_{17}, q_{18}, q_{24} and q_{25}, will have an extra non-linear term, since instead of having $q_1 = p_{Met}$ or $q_2 = p_{Acet}$, they will have $q_1 = p_{Met} - \varepsilon_4(q_5 + q_{20})$ or $q_2 = p_{Acet} - \varepsilon_4(q_6 + q_{21})$. This difference has a huge impact on the behaviour of the system. Now, since q_1 and q_2 are not constant and they involve one variable of each gene, they add a coupling on to the 4-dimensional system. This implies that now the system behaves as a unit, where the gene promoting differentiation and the gene promoting pluripotency are competing for the marks, i.e. they are not independent one from the other. Therefore, any change occurring to the epigenetic status of one gene, will have effects on the other one.

Although we have not considered this case, we want to notice that within the limit $\varepsilon_4 = O(1)$, we could assume the abundance of acetylation and methylation marks to be different. More precisely, we could have two parameter values, ε_{4a} and ε_{4m}, denoting the abundance of the acetyl and methyl groups, respectively, with both parameters $O(1)$ but not necessarily ε_{4a}=ε_{4m}. This will allow to explore cases where acetyl groups are more abundant than methyl groups, or vice versa. For simplicity, we have not considered this case.

4.2.3 Parameter values and viability conditions

Similarly to what we have done in Chapter 2, we consider a network of two genes, one gene promoting differentiation and the other one promoting pluripotency.

Within such a context, the epi-phenotype of a normal somatic, differentiated cell demands that those genes promoting pluripotent behaviour and/or proliferation should be silent, whereas genes promoting differentiation and quiescent behaviour should be active.

Although having presented a more complex model (see Section 4.2) able to couple the activity of both genes through the epigenetic marks, we initially use the case where the activity of both genes is decoupled, i.e. we take $\varepsilon_4 = 0$ to generate the parameter values. We proceed in this way because this case mimics the case used in Chapter 2, where we considered that the epigenetic regulatory model was composed of two replicas of the stochastic ER model, as described in Section 2.2.1, each one related to a gene of the GRN. In the case presented in here, the ER system is characterised by just one set of kinetic rates, κ_j (see Table 4.3), so that the behaviour of the system can be tuned to the demands of the viability conditions (to be specified below) by changing these parameters. The variability in the rates k_j, which are related to the activity of the HMEs that carry out the epigenetic modifications, can be interpreted in terms of the existing heterogeneity in the availability of cofactors, such as NAD+ or acetyl-CoA, which are needed for the HMEs to perform their function, and which change their abundance with ageing. It is important to remark that the kinetic rate constants, κ_j, as given in Table 4.3, are dimensionless, so there is an undetermined scale in our system, k_{16}, which can be used to fit real data (if available).

For the remainder of this Chapter, and according to the notation introduced in Table 4.1 and Table 4.3, an *open* epigenetic state for the differentiation gene will refer to a steady state of the system where $q_4 \simeq q_5 \simeq 0$ and $q_6 \simeq 1$ (predominance of positive marks). Similarly, for a pluripotency gene, the open state will be characterised by $q_{19} \simeq q_{20} \simeq 0$ and $q_{21} \simeq 1$. A *closed* or *silent* epigenetic state for the differentiation gene is associated with $q_4 \simeq 0$, $q_5 \simeq 1$ and $q_6 \simeq 0$ at equilibrium (predominance of negative marks), whilst the corresponding closed state for the pluripotency promoting gene is characterised by $q_{19} \simeq 0$, $q_{20} \simeq 1$ and $q_{21} \simeq 0$.

In the Appendix (Section A.3.1) we give the tables of parameter values used in our ER system so as to generate the plastic and resilient epiphenotypes, as described in Chapter 2. These parameter sets have been chosen following the same viability criteria as in Chapter 2 (see Section 2.2.3.1), namely, that for a media with abundant marks ($\varepsilon_4 = 0$) at normal enzymatic level (for all the HMEs this model includes), the differentiation(pluripotency) gene has just one stable steady state, corresponding to its open(closed) state. Furthermore, for these parameter sets it is also required

that for $p_{HDM_D} < 1$, $p_{HDAC_D} < 1$ ($p_{HDM_P} < 1$, $p_{HDAC_P} < 1$) the differentiation (pluripotency) gene has a bistability region, where the open and the closed states of the gene coexist.

4.3 Results

In this Section, we study the effects that ageing and epigenetic dysregulation may have in the resilient and plastic scenarios (as defined in Section 2.3 in Chapter 2) regarding their robustness with respect to changes in the abundance of epigenetic marks and HME activity, and their consequences in cell reprogramming.

4.3.1 The plastic and resilient scenarios are recovered under the assumption of abundant marks

The first analysis we perform is to study whether in this model, the previous two scenarios identified in Section 2.3, namely, the plastic and the resilient scenario, are both still feasible. In order to do this, we set $\varepsilon_4 = 0$, which implies that the model has no coupling, i.e. there is no competition for the marks because the epigenetic marks are abundant in the media (as assumed in Chapter 2). When $\varepsilon_4 = 0$, the 4-dimensional system of ODEs presented (see Section 4.2) decouples into two independent subsystems, each of 2-ODEs, one for the ER system of the gene promoting differentiation and another one for the ER system of the pluripotency-promoting gene. This is exactly the situation modelled in Chapter 2, when we had two replicas of the stochastic model of ER. This similarity between both systems is the one we want to take advantage of so as to identify the two relevant scenarios previously considered.

Fig. 4.1 and Fig. 4.2 show, respectively, the plastic and the resilient scenario. As can be seen, the main difference between Figure 4.1 and 4.2 is the overlapping/nonoverlapping bistability regions. In Fig. 4.1 it can be observed that the bistability region of the differentiation gene, described by the area between the red and the blue dashed lines, and the bistability region of the pluripotency gene, depicted by red and blue solid lines, have a region where they overlap. In this section of the phase diagram, with the zone of interest zoomed at in Fig 4.1(b), it can be seen that the open state boundary for the pluripotency gene (red solid line) appears at the left (i.e. for lower values of p_{HDM}) of the closed state boundary for the differentiation gene (blue dashed line), indicating that the bistability regions coexist in this area.

In this region both genes are bistable, which indicates, that the system could move to the closed-differentiation, open-pluripotency state, i.e. the cell could reprogram, giving rise to the so-called plastic scenario. This diagram has been obtained with the parameter values from Table A.22 in Section A.3.1.

In contrast, Fig. 4.2 illustrates a resilient epiphenotype, where the bistability regions do not overlap. In this scenario, cell reprogramming is not possible because both genes cannot be bistable at once, which is why this scenario is labelled as the resilient scenario. In the situation depicted in Fig. 4.2, whenever the differentiation gene can be closed (region at the left of the blue dashed line), the pluripotency gene is also closed, impeding thus, reprogramming to occur. This refractory scenario has been generated with the parameter values given in Table A.21, Section A.3.1.

4.3.2 Cell reprogramming is still feasible in a media with non-abundant marks

In the previous section, we have shown that with the model presented in this Chapter, the two relevant scenarios (resilient and plastic) presented in Chapter 2 can still be obtained. The current model allows us to investigate the effect of altering the abundance of the epigenetic marks, since this factor is modelled in terms of ε_4.

We are interested in studying whether the plastic scenario stays feasible in a media of non-abundant epigenetic marks (i.e. $0 < \varepsilon_4 \leq 1$) or if on the contrary, it ceases to exist. The latter would imply that the plastic scenario is just an artifact of our model assumptions but it would not be biologically plausible, since epigenetic marks can be really large, but not infinite as assumed when taking $\varepsilon_4 = 0$.

In order to assess the robustness of the plastic scenario regarding variation on the abundance of marks, we consider the parameter values from the plastic scenario, Table A.22, and we increase ε_4. The resulting phase-diagram for $\varepsilon_4 = 10^{-5}$ is shown in Fig. 4.3. As can be observed (see Fig. 4.3(b) for better observation), in a media of non-abundant marks, the plastic scenario is still conserved. This finding reveals that cell reprogramming is feasible in a media of non-abundant marks and, in fact, it implies that cell reprogramming is attainable in a realistic assumption, i.e. when marks are finite.

Although it may seem that $\varepsilon_4 = 10^{-5}$ is a very small parameter, we want to

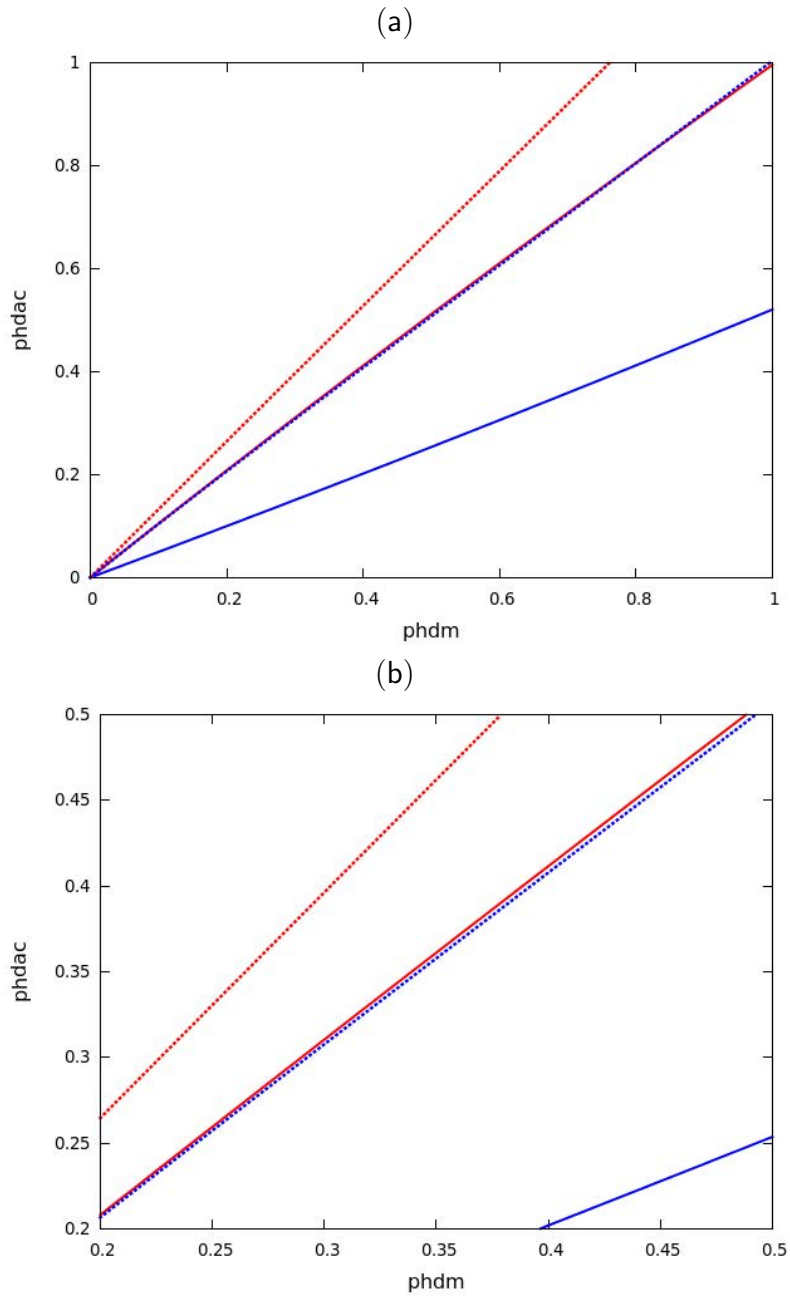


Figure 4.1: Caption on the following page.

Figure 4.1: (Previous page). Plots showing the existence of the plastic scenario, as described in Section 2.3, when the marks are abundant in the media, which is obtained by setting $\varepsilon_4 = 0$. In both plots, solid (dashed) lines correspond to the bistability limits of the pluripotency(differentiation)-promoting gene. Red(blue) color denotes the boundary where the open(closed) state ceases to be stable. This means that the region at the left of the solid (dashed) red line, the only possible stable steady state is the closed one for the pluripotency(differentiation) gene, the region between the solid (dashed) red and solid (dashed) blue line is where the pluripotency(differentiation) gene is bistable, and in the region at the right of the blue solid (dashed) line the only possible stable steady state is the open state for the pluripotency (differentiation) gene. (a) Phase space of a plastic scenario corresponding to the ER model with parameter set given by Table A.22 in Section A.3.1. (b) Zoom of the phase space shown in (a), where the overlapping between both bistabilities regions is clearer (region between the red solid line and the blue dashed line).

emphasise the important step taken when moving from $\varepsilon_4 = 0$ to $\varepsilon_4 = 10^{-5}$, since it means changing the scenario from an infinite number of marks, to a media where they have a finite presence. In fact, once the existence of the plastic scenario has been confirmed for a media with non-infinite marks, we have wondered on the effect of a further reduction of the abundance of the epigenetic marks, i.e. an increase on the ε_4 parameter. Taking the initial plastic scenario (described by Table A.22 in Section A.3.1) and setting $\varepsilon_4 = 0.01$ and $\varepsilon_4 = 0.1$, gives rise to the scenarios depicted in Fig 4.4 and Fig. 4.5, respectively.

Fig. 4.4 and Fig. 4.5 reveal several interesting results. The first thing to notice when observing Fig. 4.4 and 4.5 is that none of them exhibits plastic behaviour (i.e. the overlap between both bistability regions vanishes). In both cases, thus, cell reprogramming is not achievable. Furthermore, when comparing Fig. 4.4 (a) to Fig. 4.5, where the only difference is an increase in ε_4, it can be seen that the bistability region for the differentiation gene is shifted to the left, whilst the bistability region for the pluripotency gene is shifted to the right. This combined movement of both bistability regions have as consequence that the region in between these bistability regions, corresponding to open differentiation gene and closed pluripotency gene, i.e. normal somatic, differentiated cell, increases its area. This can be interpreted as an increased robustness of the differentiated epi-phenotype, since now, in order to enter a bistability region, the variation in the HMEs activity needs to be larger than previously, when both bistability regions were really close. Notice that in all the

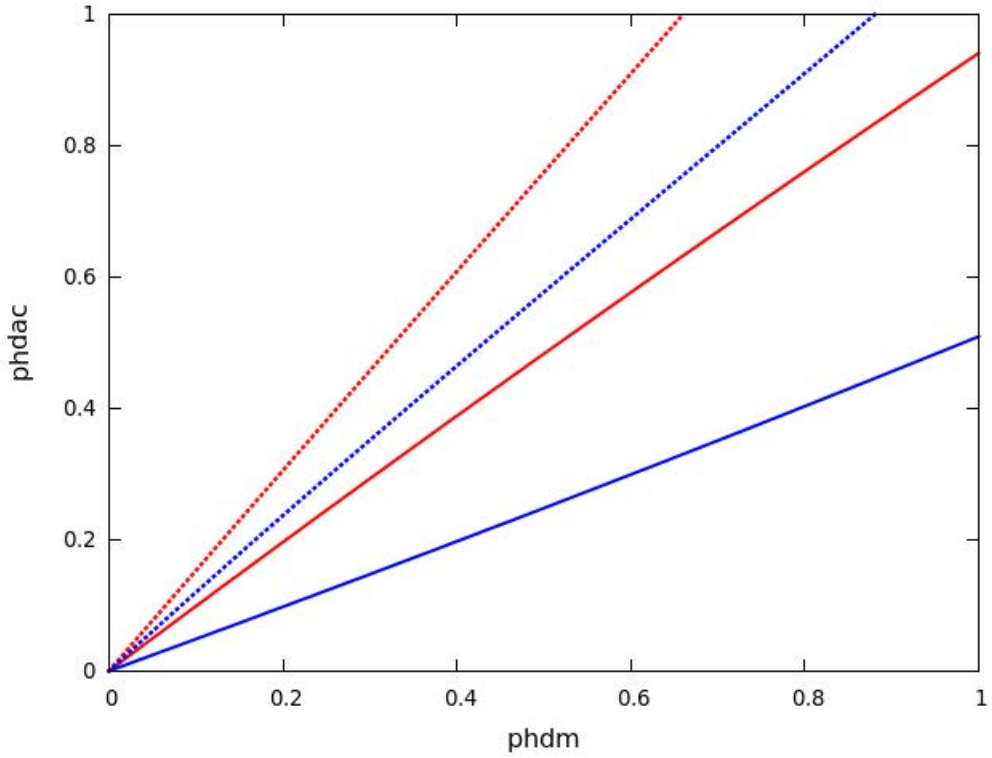


Figure 4.2: Plot showing the existence of the resilient scenario, as described in Section 2.3, when the marks are abundant in the media, which is obtained by setting $\varepsilon_4 = 0$. Solid (dashed) lines correspond to the bistability limits of the pluripotency(differentiation)-promoting gene. Red(blue) color denotes the boundary where the open(closed) state ceases to be stable. This means that the region at the left of the solid (dashed) red line, the only possible stable steady state is the closed one for the pluripotency(differentiation) gene, the region between the solid (dashed) red and solid (dashed) blue line is where the pluripotency(differentiation) gene is bistable, and in the region at the right of the blue solid (dashed) line the only possible stable steady state is the open state for the pluripotency (differentiation) gene. Parameter set for this resilient scenario given by Table A.21 in Section A.3.1.

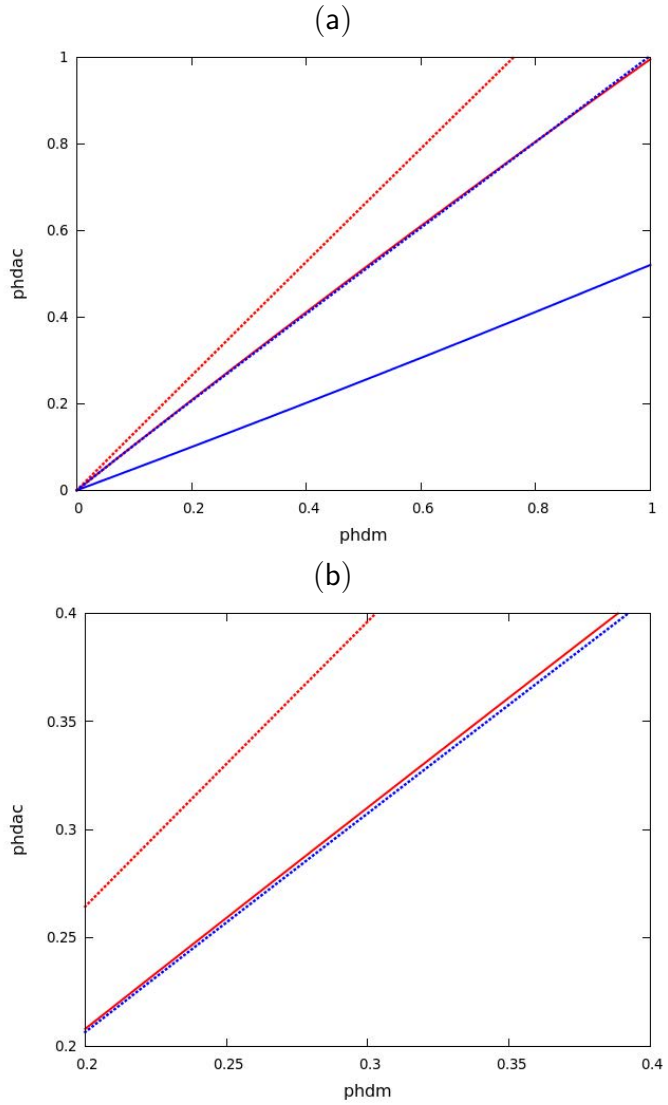


Figure 4.3: Plots showing the existence of the plastic scenario, as described in Section 2.3 even when the marks are not abundant in the media, $\varepsilon_4 = 10^{-5}$. In both plots, solid (dashed) lines correspond to the bistability limits of the pluripotency(differentiation)-promoting gene. Red(blue) color denotes the boundary where the open(closed) state ceases to be stable. In the region at the left(right) of the red(blue) line the only possible stable steady state is the closed(open), and the region between the red and the blue line exhibits bistable behaviour. (a) Phase space of a plastic scenario corresponding to the ER model with parameters given by Table A.22 in Section A.3.1 (b) Zoom of the phase space shown in (a), where the overlapping between both bistabilities regions is clearer (region between the red solid line and the blue dashed line).

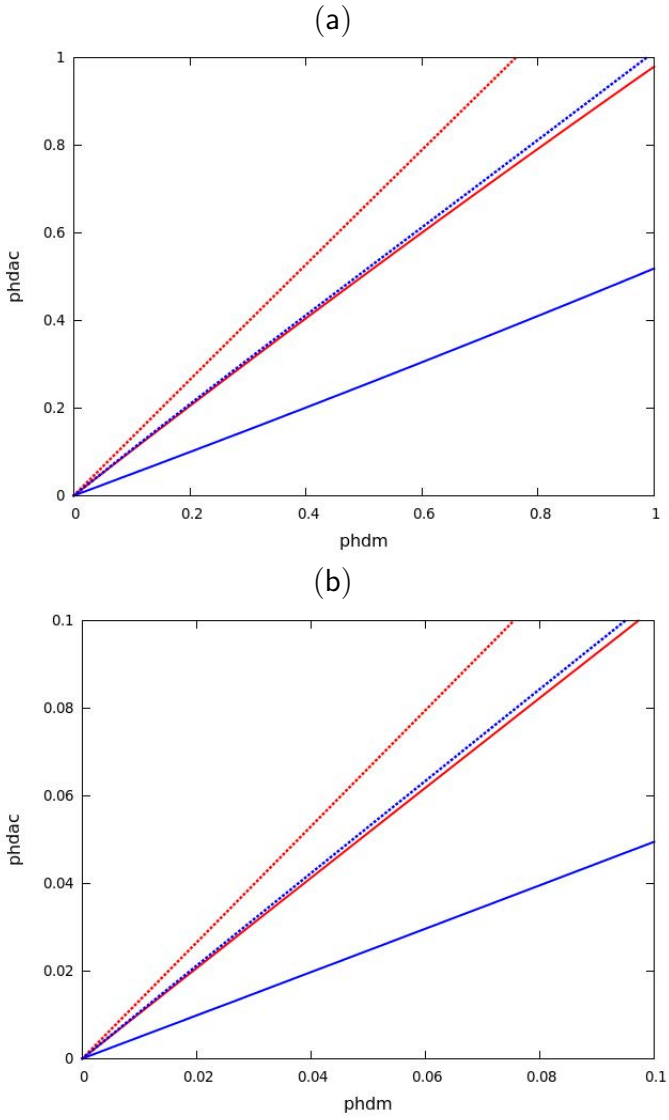


Figure 4.4: Plots showing the loss of the plastic scenario ($\varepsilon_4 = 10^{-2}$). In both plots, solid (dashed) lines correspond to the bistability limits of the pluripotency(differentiation)-promoting gene. Red(blue) color denotes the boundary where the open(closed) state ceases to be stable. At the region at the left(right) of the red(blue) line the only possible stable steady state is the closed(open), and the region between the red and the blue line exhibits bistable behaviour. (a) Phase space of a resilient scenario, as described in Section 2.3, corresponding to the ER model with parameters given by Table A.22 in Section A.3.1. (b) Zoom of the phase space shown in (a), where the non-overlapping between both bistability regions is clearer.

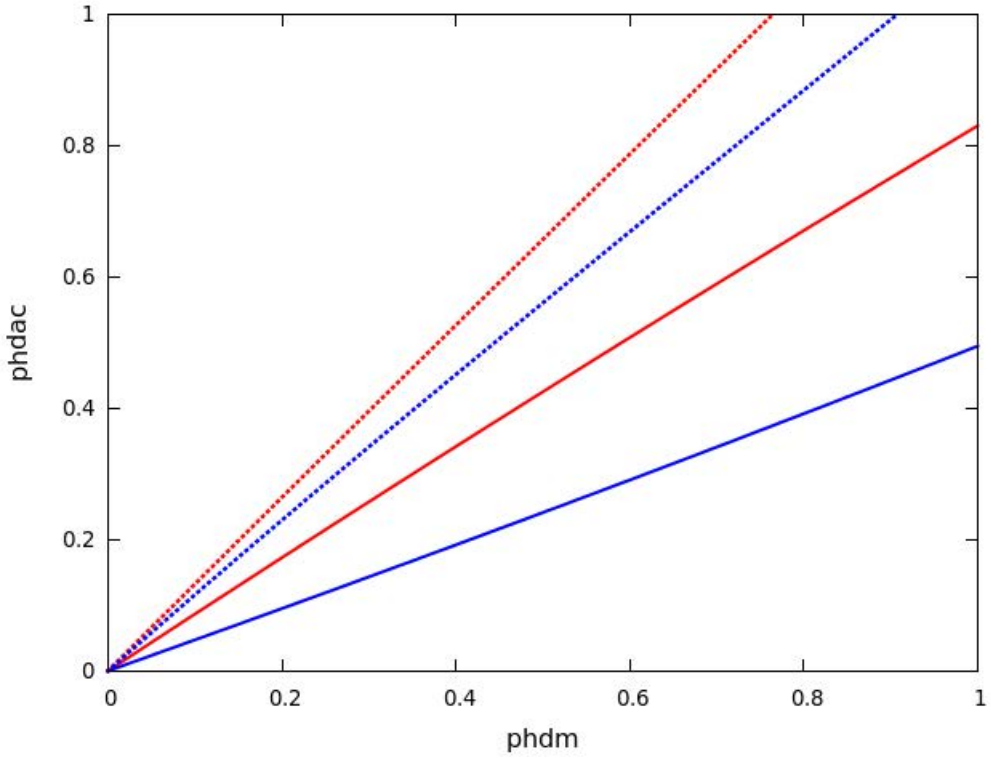


Figure 4.5: Plot showing that the differentiated epiphenotype increases its robustness as ε_4 increases. Plot for $\varepsilon_4 = 10^{-1}$. It can be observed that the bistability regions do not overlap and the distance separating them, distance between the dashed blue line and the solid red line, increases when compared to Fig. 4.4. Solid (dashed) lines correspond to the bistability limits of the pluripotency(differentiation)-promoting gene. Red(blue) color denotes the boundary where the open(closed) state ceases to be stable. At the region at the left(right) of the red(blue) line the only possible stable steady state is the closed(open), and the region between the red and the blue line exhibits bistable behaviour. ER system given by Table A.22 in Section A.3.1.

above figures, the x axis has represented the HDM activity, assuming to be the same for both genes, whilst the y-axis has represented the activity of HDAC enzymes, again assuming to be the same for both genes, i.e. setting $p_{HDM_D} = p_{HDM_P}$ (and denoting it as $phdm$) and setting $p_{HDAC_D} = p_{HDAC_P}$ (and denoting it as $phdac$). Let us recall that $phdm = 1$ or $phdac = 1$ represents normal activity (abundance) of these enzymes.

In Section 4.3.1 we have proved the existence of the plastic scenario. In this Section we have shown that the plastic scenario is still feasible in a media with non-infinite epigenetic marks, although we have also concluded that a further increase on the ε_4 value renders the system without the plastic behaviour, increasing the robustness of the differentiated state. However, as discussed in Section 1.2.2.1, ageing tissues exhibit cell reprogramming [115]. Our results suggest that in order to find a plausible explanation for the correlation between ageing and cell reprogramming, the effect of other factors needs to be incorporated, because by solely decreasing the abundance of epigenetic marks, the plastic behaviour is not obtained. It is well-known that ageing tissues are characterised by epigenetic dysregulations, such a decrease on the activity of the HDACs (sirtuin family) which is assumed to shorten lifespan, or a decrease of HDMs activity found in several cancers (IDH mutation) [17, 12, 30, 128, 41, 122] . In the next section, we will analyse the precise epigenetic alterations related to ageing, so as to see if we can recover the existing cell plasticity characteristic of ageing.

4.3.3 Analysing ageing effects: appearance of cancer-related pathological plasticity and how to revert it

According to the results by Inglés et al. [80], centenarian individuals exhibit overexpression of pluripotency-related genes, when compared to younger individuals. Therefore, in order to study the effects of ageing in the model, we will study the effect of increasing the number of HAT enzymes for the pluripotency gene, modelled through the parameter p_{HAT_P}, since this implies that the pluripotency gene will have extra help in order to acetylate, which is an epigenetic mark for gene expression (open gene). In order to verify whether increasing the number of pluripotency gene HATs would favour reprogramming, as per the experimental results in ageing tissues [151, 179], we start with the parameter set given by Table A.22, such that for $\varepsilon_4 = 0$ plastic behaviour was obtained, but plasticity was lost when increasing ε_4 (see Fig. 4.4 and Fig. 4.5). In particular, we assume that the number of HAT enzymes for the pluripotency gene becomes greater by a 50%, i.e. $p_{HAT_P} = 1.5$. The abundance of all

the other HMEs remains unaltered. The results for this situation, with $\varepsilon_4 = 0.1$, are illustrated in Fig. 4.6. It is worth remarking that for $\varepsilon_4 \neq 0$, there is competition for the marks, so changing the activity of some epigenetic enzymes adding or removing marks to/from the pluripotency gene, such as p_{HAT_P}, has also consequences on the differentiation gene. If the activity of one enzyme adding (removing) marks increases on one gene, it implies that there are less(more) epigenetic marks available for the other gene to use.

Fig. 4.6 represents the bifurcation diagram obtained when setting $p_{HAT_P} = 1.5$ with respect to the abundance of HDAC enzymes, which for simplicity, has been assumed to be the same for the differentiation and the pluripotency genes, i.e. $p_{HDAC_D} = p_{HDAC_P}$, which therefore are denoted by the general notation p_{HDAC}. Analysing Fig. 4.6, we can observe that at normal enzymatic level for the HDAC enzymes, $p_{HDAC} = 1$, the system exhibits tristable behaviour (stable steady states shown in red). This behaviour can be appreciated in more detail by observing Fig. 4.6 (a), where we can see that q_6, variable corresponding to the (re-scaled) acetylated sites at the promoter region of the differentiation gene, has two steady states with high values (associated with open differentiation gene) and one steady state with lower values, corresponding to closed differentiation gene (Fig. 4.6 (b) provides a better representation of the two steady states with high q_6). When looking at Fig. 4.6 (c), which represents the steady state values of q_{21}, corresponding to the (re-scaled) acetylated sites at the promoter region of the pluripotency gene, we can observe that at normal enzymatic level, $p_{HDAC} = 1$, there are two stable solutions with high values of q_{21} (corresponding to open pluripotency gene), and one steady state solution with lower values of q_{21} which is related to closed pluripotency gene. Again, the solutions for high values of q_{21} are better depicted in Fig. 4.6 (d). Therefore, we can see that at normal enzymatic level, there exists different possibilities regarding the state the cell can adopt. Specifically, these diagrams (see Fig. 4.6) shows the existence of five possible steady state solutions, of which, three are stable: one corresponds to both genes open, another one corresponds to the differentiation gene open and the pluripotency gene closed and the third one, which is the one leading to cell reprogramming, has differentiation gene closed and pluripotent gene open. It is important to remark that since $\varepsilon_4 = 0.1$, the behaviour of one gene is not decoupled from the behaviour of the other gene and this is why, we cannot combine any possible stable steady state of the differentiation gene to any possible stable steady state of the pluripotency gene, because they are steady states of the 4-dimensional system, not of the 2-dimensional subsystems (as analysed in Chapter 3). This explains why both differentiation and pluripotency genes closed is not a stable solution in this case.

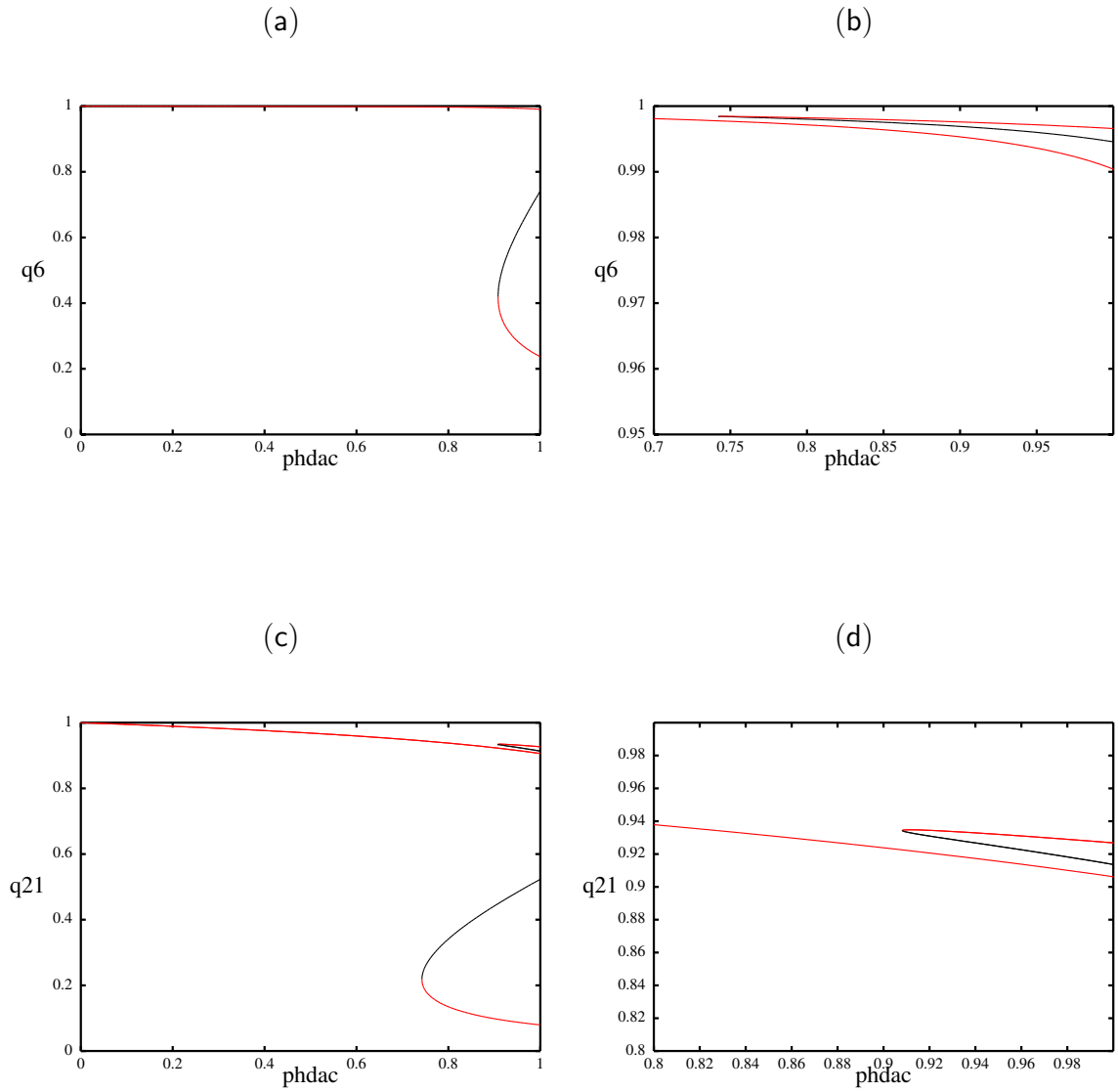


Figure 4.6: One-parameter bifurcation diagrams showing the value of the variable q_6 (a,b) and q_{21} (c,d) at steady states versus the parameter p_{HDAC}, assuming $p_{HDAC_P} = p_{HDAC_D}$, for $\varepsilon_4 = 0.1$ and $p_{HAT_P} = 1.5$. Red (black) lines correspond to stable(unstable) steady states. Parameter values given by Table A.22 in Section A.3.1. Plots (b) and (d) are zooms of the top right region of plots (a) and (c), respectively, which are provided for a better illustration of the existing bistability in that area.

The situation depicted in Fig. 4.6, which is related to the previous resilient scenario depicted in Fig. 4.5 where cells were in the differentiated state for normal enzymatic levels, shows that when altering the activity of HAT_P, setting $p_{HAT_P} = 1.5$, and letting the HDAC enzymes at its normal activity ($p_{HDAC} = 1$) cell reprogramming is feasible. In other words, the up-regulation of the enzymatic activity of the HAT for the pluripotency gene seems to drive the system towards a situation where cells could renew, since they could be reprogrammed towards a stem cell-like state. This, which we relate to a 'healthy' ageing, could explain why people overexpressing the pluripotency-associated genes can become centenarians, since these individuals have the possibility to rejuvenate their tissues, by generating stem cell-like cells, and differentiating them again.

Interestingly, the 'healthy' ageing situation described in the previous paragraph, not only holds for normal activity of HMEs. As depicted in Fig. 4.6, if the activity of HDAC enzymes decreases slightly, the cell could still reprogram towards the stem-cell like state, i.e. differentiation-gene closed, pluripotency-gene open. However, it is well-documented that HDACs decrease their activity with ageing [125, 122, 48]. As can be seen in Fig. 4.6, if this reduction exceeds a threshold, then, the system enters into a regime where the bistable behaviour for the differentiation-gene disappears and if the reduction is more severe, then, the pluripotency-gene gets locked in the open state. Therefore, a decrease of the HDAC activity, which is known to happen with ageing, will drive the 'healthy' ageing ('healthy' plasticity) to a malignant or pathological plasticity, where cells will lose their ability to differentiate again. In other words, the cell will enter a region where the only possible fate it can take is a less differentiated state, where pluripotency genes are expressed, i.e. cell will get locked in an undifferentiated state, which may cause the accumulation of undifferentiated cells in the tissue, a feature observed in some cancers. This situation, where there is a combined effect of an increase of HAT activity for the pluripotency genes and a global decrease in activity of HDAC enzymes, is the one assumed to occur in normal ageing, which is the reason we label this situation as the physiological ageing.

The physiological ageing situation depicted in Fig. 4.6 holds for larger values of ε_4, i.e. for scenarios where the epigenetic marks are less abundant. Fig. 4.7 depicts the results for $\varepsilon_4 = 0.5$, where it can be observed that the results are qualitatively analogous, with the difference on the value of p_{HDAC} for which the pathological plasticity appears. In this case, the level of HDAC reduction must be greater than in the previous case when $\varepsilon_4 = 0.1$.

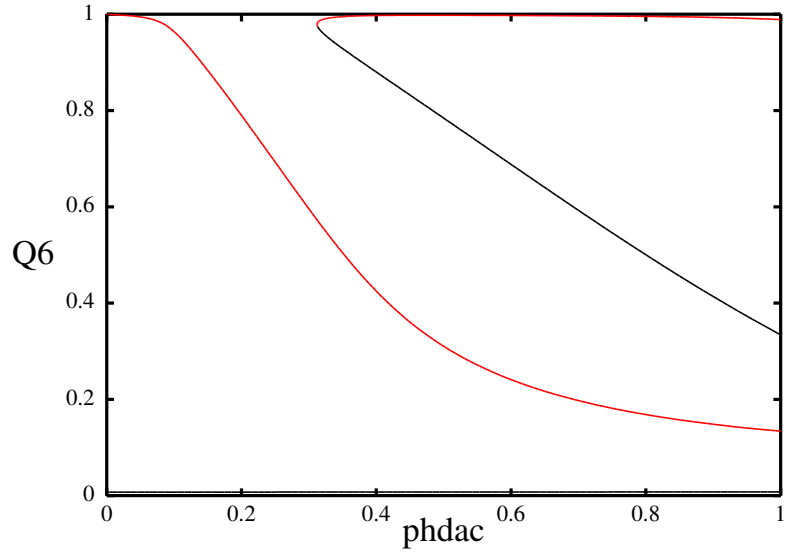
Therefore, upregulation of the pluripotency-associated genes can render cells in a situation favourable for renewing tissues and thus, it may be one of the reasons why this overexpression is found in centenarian people [80]. However, when this feature is accompanied by a decrease of the HDAC activity, the transient plasticity acquired may be permanent, which will put the cells in a situation favourable for cancer development. An immediate consequence of this fact is that if we could find ways to boost the activity of HDAC enzymes to recover its normal average, i.e. to avoid the decline associated with ageing, then we would be able to obtain results in two interesting ways, since we would move from the pathological to the 'healthy' ageing scenario. On one hand, we would avoid situations with higher oncogenic likelihood, and on the other hand, by being able to renew our tissues we may find the way to revert the ageing process. Interestingly, it has been reported in the literature that the increase of the activity of sirtuin 1 (see Section 1.2.2.4.1), which has HDAC activity, has been proved to be successful in extending lifespan, which is in agreement with our results. Our findings agree with the fact that when accounting for epigenetic dysregulation associated with ageing, pathological cell reprogramming is a realistic possibility, highlighting the key role played by HMEs as drivers of impaired ageing.

With the epigenetic strategy suggested, increasing HDAC activity, we have exemplified possible epigenetic alterations that we argue that could give rise to a 'healthy' plastic phenotype in the presence of factors associated with ageing. Since this model takes into account the activity of 8 enzymes, the number of strategies that we could design is large. The strategy we have presented was consisting on changing the activity of just one enzyme, but of course, combined strategies that alter the activity of more than one enzyme are also possible. How successful these strategies are is something needed to quantify as it will provide useful therapeutic information. Of particular interest will be testing which strategies drive the system towards a more favourable situation for reprogramming, in the sense that the appearance of the 'healthy' ageing phenotype appears without the possibility for stem-cell locking reprogramming.

4.4 Discussion

In this Chapter, we have presented a model of the epigenetic regulation of a two gene regulatory system, where one gene promotes differentiation and one gene promotes pluripotency. This model includes the one presented in Chapter 2, where each

(a)



(b)

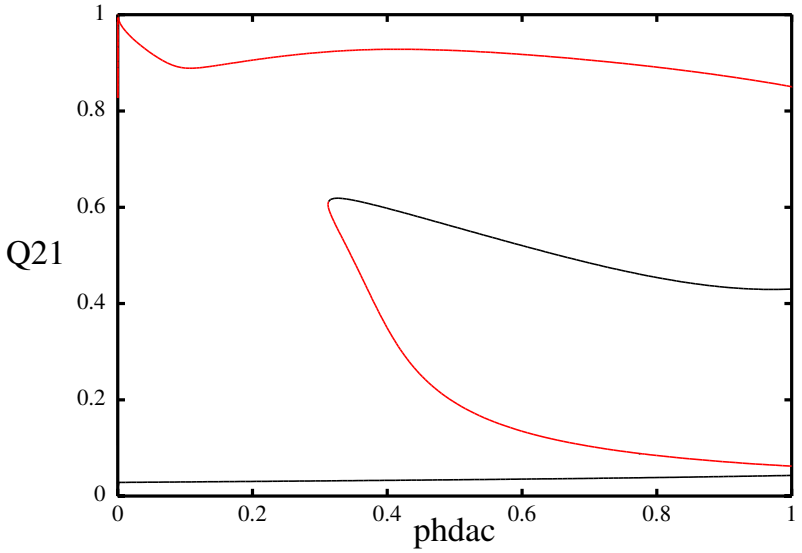


Figure 4.7: One-parameter bifurcation diagrams showing the value of the variable q_6 (a) and q_{21} (b) at steady states versus the parameter p_{HDAC}, assuming $p_{HDAC_P} = p_{HDAC_D}$, for $\varepsilon_4 = 0.5$ and $p_{HAT_P} = 1.5$. Red (black) lines correspond to stable(unstable) steady states. Parameter values given by Table A.22 in Section A.3.1

gene had its own ER model. The model presented here extends the previous one by adding competition for the epigenetic marks needed for the epigenetic alterations at the histones of the genes. The addition of competition implies that both genes are not independent and thus, the model goes one step closer towards bridging the modelling framework and the real biological problem, where systems, rather than disconnected, form a high complex unit.

We have extended the previous model by both incorporating the activity of specific gene enzymes and by adding explicit modelling of HATs and HMTs activity. By exploiting the multiple-scales arising naturally in the model, we have incorporated a new parameter able to represent the abundance of the epigenetic marks. Such abundance is known to be altered within ageing systems and our model allows us to study this alteration by changing the value of the corresponding parameter. Therefore, the extended model allows us to study the interplay between ageing and epigenetic dysregulation. As discussed in Chapter 1, epigenetic dysregulation associated with ageing has recently emerged as one of the key elements driving tumorigenesis (see Section 1.2.2.3). In particular, epigenetic heterogeneity has been linked to higher reprogramming rate (see Section 1.2.2.1), to the metabolism rewire characteristic of cancer cells (see Section 1.2.2.4.1)... Therefore, this Chapter sheds some light on the connection between the combined effects of ageing and epigenetic dysregulation to cancer.

We have placed our efforts on analysing the existence and robustness of plastic behaviour associated with the epigenetic regulatory system, as defined in Section 2.3, but also on its transient or permanent acquisition, as discussed in Chapter 3. Cell reprogramming is feasible for ER systems exhibiting the plastic epi-phenotype. Determination of the conditions which favour the appearance of this scenario allows us to understand when cells can revert to a stem cell-like state, e.g. when CSCs may appear. As discussed, one of our main results is that the plastic epiphenotype appears in realistic situations, i.e. when genes compete for a finite number of epigenetic enzymes. However, when marks become scarce, the possibility of cell reprogramming disappears and the differentiated epi-phenotype, characterised by open(closed) differentiation(pluripotency) gene, increases its robustness. Since cell reprogramming is known to occur withing ageing tissues, we have further studied ageing effects.

When taking into consideration not only the abundance of the epigenetic marks, but also ageing-related epigenetic alterations regarding variations in activity of HMEs, in particular HATs [80] and HDACs [122, 174], we can distinguish two different plas-

tic phenotypes. On one hand, there is the 'healthy' plastic phenotype, characterised by bistable behaviour of both genes, which allow, first, to revert to a stem-cell like state and later, to return to the differentiated state, since both states are stable. These state could be related to healthy reprogramming that helps the regeneration (or rejuvenation) of tissues and, consequently, revert some of the problems associated with the ageing process. On the other hand, we have characterised a 'pathological' (or physiological) plastic state, by which the acquisition of stem-cell features is irreversible and cells get locked in a undifferentiated state. Noticeably, this irreversible plasticity appears when taking epigenetic dysregulation effects, such as those observed in the sirtuin family, in cooperation with ageing.

Our modelling results show that epigenetic enzymatic alterations associated with ageing may be responsible for locking cells in a cancer-prone state, impeding their differentiation. The accumulation of undifferentiated cells has been related to cancer and according to our model, this is just a consequence of aged cells with dysregulation on the activity of HMEs. Consequently, epigenetic alterations may underlie initiating events that set the system in a cancer-prone situation. Interestingly, successful strategies so as to revert this situation and set the cell in a 'healthy' plastic state have been formulated, consisting on the restoration to base-line levels of the HDAC activity.

Our model suggests that when trying to design drugs to revert ageing-effects, epigenetic activity cannot be ignored. In fact our modelling will recommed to those studying senolytic drugs, that these drugs should be accompanied by epigenetic strategies, such the one suggested based on restoring the normal epigenetic activity of HDACs. Hence our mathematical model may be helpful for a better understanding of the existing connection between ageing and epigenetic enzymatic dysregulation and, consequently, to distinguish the 'healthy' ageing progression versus a disease-associated ageing.

Chapter 5

Conclusions & Future work

5.1 Conclusions

As Nessa Carey explains in her book [27], if we go back to the past, it can be said that during the 19th century, Darwin and Mendel defined the era of evolution and genetics. Watson and Crick marked the 20th century by coining it as the DNA era, helping to understand how genetics and evolution interact. Now, it is expected that epigenetics will trigger a similar paradigm shift in the 21st century, converting it into the epigenetic era [27]. In fact, epigenetic therapies will probably have a major clinical impact sooner than later, but in order to be successfully effective, understanding how epigenetic alterations and, more in general, epigenetic plasticity drives diseases such as cancer is a must. For this reason, trying to test and validate some models for cancer epigenetics seems a key step towards this goal [51]. This is precisely what this thesis has intended to, i.e. formulating and analysing mathematical models in order to shed some light on the role played by epigenetics in cell plasticity.

In this thesis, we have aimed at unravelling the effects that epigenetic regulation and, in particular, heterogeneity in the epigenetic regulatory systems, have in both cell fate decisions and transitions. The former has been discussed in Chapter 3, where we have shown that depending on the epigenetic regulatory system, some cells are resilient to differentiation, i.e. they want to remain in a stem cell-like state. The latter, cell fate transitions, has been discussed mainly in Chapter 2 and Chapter 4, where we have focused our discussion on identifying key properties of the epigenetic regulatory systems in order to allow a cell to dedifferentiate, i.e. to change its cell fate to a pluripotent one. This issue has been studied in more depth in Chapter 4, where we have also analysed when this cell fate transitions are reversible, which

leads to the so-called 'healthy' plastic behaviour, and when these transitions are permanent, scenario which is related to malignant plasticity due to its connections with tumorigenesis. All these results let to a general conclusion, namely, that epigenetic regulation plays a key role in determining cell fate and it is on its heterogeneity, where the answer to many questions is found.

In Chapter 2 we have formulated a stochastic model of epigenetic regulation which has allowed us to distinguish two important scenarios, the so-called resilient and plastic scenarios. The main difference between them is the existence of a regime (defined in terms of the activity of HMEs) where cell reprogramming is feasible. The plastic scenario is characterised by having a region on the HDM-HDAC phase space where both genes, the differentiation and the pluripotency, are bistable. Within this region, where both genes coexist with bistable behaviour, our model allows the ER system to switch from a differentiation-locked state to a pluripotency-permissive one, and vice versa. In other words, within the plastic scenario ER sets the system in a position where it can reprogram from the differentiated state to the stem-cell one, and differentiate back to the differentiated cell fate, if needed.

Having established that depending on the ER systems our model can exhibit plastic or resilient epi-phenotypes, we have explored the epigenetic properties that are associated with the appearance of each scenario. Interestingly, the analysis of our model suggests the possibility of an ageing-progressive evolution from the resilient to the plastic state, since the plastic scenario is characterised by reduced acetylation levels, which have been linked to ageing [33, 133, 48, 125]. This property of the plastic scenario implies that cells exhibiting initially a resilient epi-phenotype may switch to a plastic one as ageing progresses. This hypothesis is reinforced with the results from Mosteiro et al. [115], where senescent cells, which are more likely to be found within ageing tissues, induce reprogramming by releasing signalling cues into the tissue that significantly facilitate in vivo $OSKM$-induced reprogramming. Translating these results into our modelling framework, cells with properties related to ageing, such as those having lower acetylation levels, exhibit plastic epi-phenotypes corresponding to lower epigenetic barriers associated with robustness of differentiated phenotypes.

The elite model for reprogramming postulates that some cells are more prone to reprogramming than others due to existing inherent epigenetic heterogeneity. The results from Chapter 2 are in agreement with this notion, since we have identified a scenario where cell reprogramming is possible and another one where it is not, and the only difference between them is the existing inherent heterogeneity. Similarly, Pour

et al. [129] show that some cells are initially more predisposed to reprogramming, but that this 'privilege' can be acquired by tuning cell epigenetics. In particular, they show that heterogeneity in ER can be harnessed through modification of HMEs activity to increase reprogramming. The importance of cell epigenetics has also been tested by different strategies in Chapter 2, where we have shown that resilient phenotypes become plastic phenotypes, and vice versa, when altering some epigenetic features.

The results from Mosteiro et al. [115] regarding reprogramming in vivo, have shown that senescent tissues have larger reprogramming rate. Once reprogrammed, cells might get locked in the stem cell-like state, i.e. they cannot switch to a differentiated fate [108]. This issue has been studied in Chapter 3. In this Chapter, we have formulated an ER-GRN model, where the ER model formulated in Chapter 2 has been coupled to a 2-GRN model, with one gene promoting differentiation and one gene promoting pluripotency. The study of this combined model has allowed to identify three possible (stable) cell fates: the differentiated state, the pluripotent state and the so-called undecided state, where both genes are expressed at really low levels. The biological interpretation of the undecided state is of interest, since it has been eperimentally observed that cells undergoing differentiation (or reprogramming), they do it in a two step manner: first, by switching off the pluripotency (differentiation) genes, which will lead them to this undecided state, and then, by activating the expression of the gene needed to acquired their final cell fate, differentiation (pluripotency) gene for those cells undergoing differentiation (reprogramming) [116]. Interestingly, depending on the epigenetic regulatory status (on or off) of the ER systems forming the GRN, these three possible cell fates can coexist, which will allow stochastic transitions between them. However, this combined ER-GRN model endows the system with the possibility for cell fate transitions, which consists on changing the ER state of one of the genes forming the GRN network. In other words, this system affords the possibility of ER transitions between the active and inactive form, which has implications on the state the ER-GRN system can adopt and, thus, into the cell fate.

In Chapter 3, in order to characterise the properties of ER systems more likely to produce pluripotency-locked epi-states, we have focused the discussion in analysing the existing ER heterogeneity in terms of which ER states are less likely to allow cell differentiation. Our analysis of the ER ensemble heterogeneity indicates that ER of the pluripotency gene plays a minor role in determining the ability to differentiate. By contrast, ER systems for differentiation genes are the main players to determine

differentiation or locking in the stem cell-like phenotype. Our analysis has revealed the existence of three different types of behvaiour for the ER systems of the differentiation gene: those which want to differentiate, those which are undecided since transitions between the open and closed state are both long, and those which want to remain in its closed state, i.e. they are resilient to differentiation. These last set of ER systems are those associated with stem-cell locking, since the ER system for the differentiation gene wants to remain closed, which impedes the expression of the differentiation gene. Of particular interest is being able to unlock these systems, since accumulation of undifferentiated cells, i.e. cells in a stem-cell like state, has been linked to cancer.

In order to address this question, in Chapter 3 we have been sucessful in designing stategies that can change the associated behaviour of a particular ER system from the differentiation-resilient type (stem-cell locking) to the differentiation-prone behaviour. Noticeably, the change of behaviour has been obtained by just altering the activity of epigenetic kinetic enzymatic constants. This fact reinforces that concentration of epigenetic metabolites and co-factors may be of extreme importance, since alteration in the activity of HMEs, such as HDMs and HDACs, can completely change the behaviour an ER system exhibits, which when analysed in a more general framework, shapes the ER-GRN model by altering the final state the system may take.

From a more theoretical point of view, in Chapter 3 we have derived a stochastic model reduction for the ER-GRN model, which is based on the existence of multiple scales. This model reduction allowed us to formulate an efficient hybrid simulation method which has been used to test the strategies predicted by our ensemble approach. By using the hybrid scheme, we have been able to compute the time it takes the ER-GRN system to differentiate and we have concluded, that the strategies designed are successful in the sense that they greatly reduce the time needed for cell differentiation. The implications of these strategies may open new avenues to design therapeutic strategies able to change pathological cell reprogramming, where cells get locked in the stem-cell like state, to 'healthy' cell reprogramming, where cells can recover their differentiated phenotype. These suggested kinetic routes, thus, may have important consequences in the treatment of ageing associated diseases, such as cancer, where cells get locked in undifferentiated states, which drives malignant progression.

Another important result is that, although the stochastic model reduction for the

ER-GRN model has been tested in a particular ER-GRN system, it has been formulated in general terms, and thus it is applicable to more general ER-GRN systems. Similarly, the numerical scheme could also be extended to different ER-GRN systems from the one we have dealt with. Therefore, the thesis provides a general framework of wide applicability when studying complex systems with separation of time scales, as those appearing when analysing coupled ER-GRN models.

In Chapter 4 we have extended the model presented in Chapter 2, by relaxing some of the simplifying assumptions involved in it. These new features are the consideration of HAT and HMT activity and the assumption of a finite abundance of marks, which implies that differentiation and pluripotency ER systems need to compete for them. Their introduction yields to interesting results. The first thing to be observed with this model is that the scenarios described in Chapter 2, i.e. plastic and resilient, are robust, since under finite abundance of epigenetic marks, they are still feasible.

By introducing competition for the epigenetic marks, the ER system rather than being formed by two replicas of the same model, it is formed by a single ER model where both ER systems are coupled. This clearly is a more realistic model, since in a multicellular organism, the epigenetic marks source is the same for all the genes, which implies that they need to compete for them.

These additional elements allow us to take into account ageing effects in a more direct way than it was possible in the previous model. The incorporation of ageing effects leads to a situation where cells are in the plastic scenario, as previously identified in Chapter 2, in the sense that the differentiated state may switch to a state where pluripotency gene is expressed. However, this new model allows us to distinguish two relevant cases within the plastic scenario. When ageing effects are taken into account, the plastic state the cell adopts is identified with a physiological ageing or physiological plasticity, since the undecided state is the only possible state the cell can adopt, i.e. the plasticity acquired is permanent rather than transient. However, this analysis also allows to identify the 'healthy' plastic state, where the cell can revert to a stem-cell like state but also revert to the differentiated state, i.e. the acquisition of the stem-cell state is transient and not permanent. This ability is the one that could allow tissue renewal and thus, provide a way to a healthy ageing. Interestingly, this pattern has been associated with centenarian individuals, which suggest that the formula for the eternal youth could be given in these terms.

Consequently, Chapter 4 suggests that when designing drugs so as to treat ageing-related diseases, epigenetic regulation should be considered as a major player. In particular, epigenetic dysregulation has been observed in ageing and cancer. If epigenetic activity can be maintained at normal levels, or at least, avoid severe dysregulation, epigenetic barriers will be maintained, which implies that both healthy reprogramming and normal somatic phenotypes are robust, and which reduces the likelihood of cancer and ageing-associated diseases.

5.2 Future work

The model and results of this thesis can be further extended in several ways. The first and more obvious way is by coupling the more complex ER model presented in Chapter 4 to the combined ER-GRN model presented in Chapter 3.

Regarding the model presented in Chapter 4 other situations can also be studied. In particular, in the model presented, the abundance of the methyl and acetyl groups is modelled by means of a single parameter. However, with ageing, acetylation and methylation marks do not seem to exhibit the same pattern. For instance, acetylation marks are assumed to have an increment in mid ages, and then there is a general decline, situation appearing in ageing individuals. The methylation marks do not follow the same pattern and in fact, it is not clear which behaviour they exhibit. Therefore, assuming that their relative abundance is the same, does not seem a realistic assumption.

Another feature that could be analysed within the model presented in Chapter 4 is the relative abundance of nucleosomes. It is well-documented that nucleosome abundance decreases with ageing. In all our modelling framework (both in Chapter 4 and in Chapter 2) we have assumed that the number of nucelosomes over its characteristic scale was fixed to one, i.e. the number of nucleosomes has been assumed constant. However, if we add a parameter to denote this ratio, by setting this parameter to a value lower than 1, the relative abundance will decrease and hence, a new ageing feature could be introduced into the model so as to study its effects in terms of cell fate transitions.

Of course, one of the major directions for future work will go into the experimental area. In this thesis, there is a lack of biological experiments which can confirm or re-

fute our results. It would be really interesting to perform some biological experiments which could complement the results presented here. In this sense, the formulation of a more realistic model, as the one presented in Chapter 4, should make easier to match the experiments and the parameters tuned there, with the modelling parameters presented here. In this sense, testing the ability to successfully move cells from the stem-cell locked state to the differentiated-prone state according to the strategies formulated in this thesis, would be something really desirable.

In this sense, we have made some contacts with some researchers working with synthetic gene regulatory networks at UPF. Based on a synthetic GRN presented in [72], we would like to test our 2-GRN model, with self-activation and competitive binding. Precisely, we would like to show experimentally the existence of the tristability region, i.e. the one with the differentiated, the pluripotent and the undecided one. As discussed previously, this tristability situation has an important biological interpretation in cell differentiation, and this is why we are interested in showing its existence, arising from this simple 2-GRN. In order to find this region experimentally, we have worked on trying to amplify the area of the phase space it occupies, and we have seen that this area is highly sensible to the degradation rates of the genes. Therefore, something which could also be analysed further is the effect of different degradation rates.

Another possible model extension would consist on considering the addition of a feedback from the GRN model into the ER model. In order to do this, HMEs, corresponding to the ER model, would be recruited by the transcription factors, which are modelled in the GRN model. Therefore, this new feature will add feedback in the opposite direction to the one considered in here, which has been from the ER system to the GRN.

Finally, the ER models presented in this thesis have assumed the existence of two types of marks, the positive and the negative marks, which activate or inactivate gene expression. However, a more realistic situation is obtained when assuming the existence of two types of substrates, the activating or the silencing substrate. Therefore, depending on which substrate the epigenetic marks are added, they may have one effect or the opposite one. This situation is more realistic because acetylation marks are well-accepted as marks for gene activity, but methylation marks can have both effects, activating or inhibiting, depending on which residue they are added. Therefore, this new formulation will allow to study both possibilities for the methylation marks, something which is out of our scope, since in here, methylation marks

have considered to be silencing marks.

Appendices

Appendix A

Supplementary materials

A.1 Supplementary materials Chapter 2

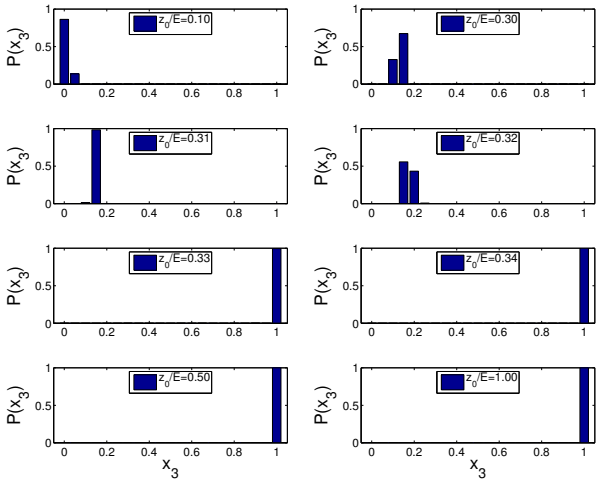
In this Section, we provide supplemental information for the results presented in Chapter 2. In particular, we present numerical results confirming the results obtained by the bifurcation analysis in Section 2.3.1, the p-values for the significative differences found by applying the Kolmogorov-Smirnov test (see Section 2.3.3), and the parameter values which give rise to the plastic and resilient scenarios, as described in Section 2.3.1.

A.1.1 Numerical results: stochastic simulation algorithm

We present simulation results verifying the bifurcation analysis of the equations discussed in Section 2.3.1. In particular, we explicitly show the existence of the hysteresis cycles predicted by our bifurcation analysis. For concreteness, since the behaviour of both differentiation- and pluripotency-promoting genes is qualitatively similar, we specifically focus on simulations of the differentiation-promoting gene. Differences with the pluripotency case only concern the quantitave value of the critical (bifurcation) points, not the behaviour of the system.

Fig. A.1 shows simulation results where we have set the number of HDAC enzymes, v_0, to $v_0 = E$ (see Table A.5). Note that according to the description given in Section 2.2.3.1, this is equivalent to fixing the HDAC concentration to $e_{HDAC} = 1$, since $e_{HDAC} = \frac{v_0}{E}$. We then vary z_0 which, according to what has been described, it is the same as varying HDM concentration, since we define $e_{HDM} = \frac{z_0}{E}$. Fig. A.1 shows results regarding the empirical distribution of x_3, $P(x_3)$, where $x_3 \equiv X_3(t_{inf})/S$ with

(a)



(b)

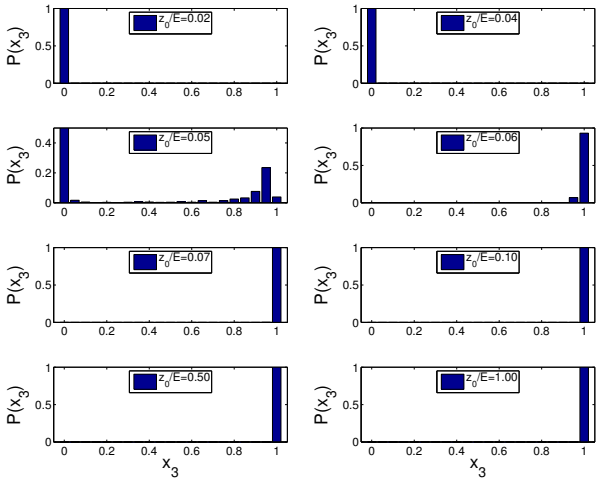


Figure A.1: Simulation results corresponding to Fig. 2.5(c). Each histogram shown in this figure is the result of 1000 realisations of the stochastic process prescribed by the rates shown in Table 2.2. Parameter values are obtained from Tables A.5 and 2.3

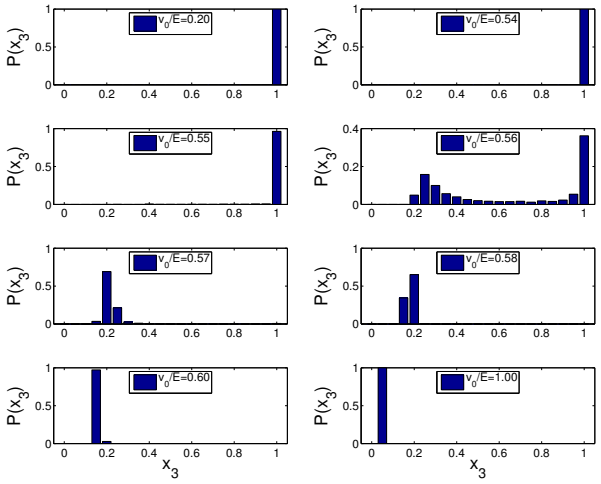
t_{inf} the duration of the simulation which is taken long enough so that the system settles onto its quasi-steady state starting from prescribed initial conditions $X_i(t = 0)$. In order to ascertain whether the system exhibits the hysteresis cycle predicted by our bifurcation analysis as z_0 changes (see Fig. 2.5(c)), we first set an initial condition with $X_1(t = 0) = 0$, $X_2(t = 0) = 0.9S$, $X_3(t = 0) = 0.1S$. This allows us to explore the behaviour of the system along the lower stable branch (corresponding to closed chromatin) of the diagram Fig. 2.5(c). Similarly, by setting initials conditions to $X_1(t = 0) = 0$, $X_2(t = 0) = 0.01S$, $X_3(t = 0) = 0.99S$, in Fig. A.1(b) we trace the behaviour of the system along the upper stable branch (associated with open chromatin).

Fig. A.1(a) shows that, for the prescribed initial condition, the system exhibits a unimodal distribution around the closed chromatin state for small values of z_0. As z_0 increases and approaches the critical value where the closed chromatin state ceases to exist, fluctuations increase (as shown by the bimodal behaviour of $P(x_3)$) thus heralding the onset of a phase transition. Beyond this point, $P(x_3)$ exhibits unimodal behaviour around the open chromatin state. These results, including the value of the critical point (which in the simulations is formally characterised by a divergence of the variance of x_3), are in agreement with those obtained from the bifurcation analysis (see Fig. 2.5(c)).

In Fig. A.1(b) we trace the other half of the hysteresis cycle. By setting initial conditions to $X_1(t = 0) = 0$, $X_2(t = 0) = 0.01S$, $X_3(t = 0) = 0.99S$, $P(x_3)$ exhibits unimodal behaviour around the open chromatin state for larger values of z_0. As z_0 is reduced and approaches the critical value for which the open chromatin state ceases to exist, fluctuations are again observed to increase, i.e. $P(x_3)$ becomes bimodal around the critical point. Beyond this point, $P(x_3)$ recovers unimodal behaviour but peaked around the closed chromatin state. Results, including the value of the critical point, are again in excellent agreement with the bifurcation analysis (see Fig. 2.5(c)).

Fig. A.2 shows the same results regarding the verification of the bifurcation analysis shown in Fig. 2.5(e). Here, we have set the number of HDM enzymes, z_0, to $z_0 = 0.2E$ (see Table A.5). We then simulate the behaviour of the system as the number of HDAC molecules, v_0, is changed. We observe the same excellent agreement between simulations and bifurcation analysis as we do between the numerical and analytical results shown in Figs. A.1 and 2.5(c), respectively.

(a)



(b)

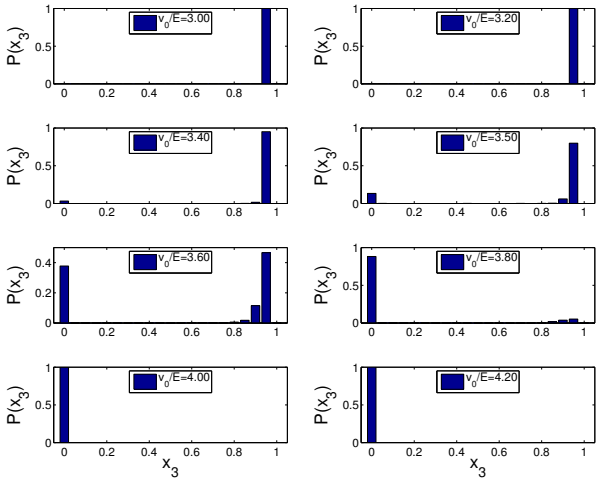


Figure A.2: Simulation results associated with Fig. 2.5(e). Each histogram shown in this figure is the result of 1000 realisations of the stochastic process prescribed by the rates shown in Table 2.2. Parameter values are obtained from Tables A.5 and 2.3.

A.1.2 Kolmogorov-Smirnov test analysis

A.1.2.1 General description

The Kolmogorov-Smirnov (KS) test [35, 29] is a non-parametric test which allows to evaluate the equality of continuous probability distributions. This well-known method can be used both to compare an empirically obtained sample with a reference probability distribution, or to compare two empiricial samples. In other words, the KS test allows us to tell whether two distributions are the same within the level of confidence we desire. Since this is a well-known technique, we will not go into its details and we will use this section to report our results. Interested readers are referred to the specialised literature for details [35, 29]. Throughout the thesis, we have imposed a level of confidence of 95 %.

In order to analyse the statistical significance of our results of Section 2.3.3 regarding the shapes of the distributions of the kinetic parameters, k_j, associated with the different scenarios we consider (refractory vs plastic), we resort to the KS test.

A.1.2.2 Comparing the viable subset with the uniform distribution

The first test we are interested in carrying out consists on checking which kinetic constants, k_j, exhibit a non-uniform distribution within the viable subset. These parameters are the ones deemed to play a substantial role in the associated behaviour (i.e. viable (base-line) conditions of the differentiation/pluripotency ER system) [50]. The null hypothesis for the test is therefore whether the empirical cumulative distribution function (CDF) of each k_j is equal to the uniform. Since we are using a confidence interval of 95%, whenever the p-value is larger than 0.05 the null hypothesis cannot be rejected, i.e. the parameter is deemed to be uniformly distributed. As we mention in Section 2.3.3, the null hypothesis is rejected only for k_1, k_3, k_6, k_7, k_{12}, k_{14}, and k_{16} (differentiation-promoting gene) and for k_3, k_8, k_{12}, k_{14}, k_{15} and k_{16} (pluripotency-promoting gene). The reported p-values are given in Table A.1 and Table A.2, respectively.

A.1.2.3 Comparing the plastic sets with the viable subset

We continue our analysis by testing the CDFs of the kinetic constants when we consider those parameter sets that exhibit plastic behaviour. We analyse which parameters have different distributions when compared to their distributions within

Parameter	p-value
k_1	$1.273016 \cdot 10^{-13}$
k_3	0.015373
k_6	0.002362
k_7	0.005169
k_{12}	$1.969483 \cdot 10^{-8}$
k_{14}	$2.467713 \cdot 10^{-51}$
k_{16}	$4.555466 \cdot 10^{-11}$

Table A.1: Reported p-values for the parameters for which the hypothesis that they are distributed uniformly was rejected (diff.-promoting gene). Confidence interval of 95%. The p-values for the rest of the kinetic constants are all larger than 0.05.

Parameter	p-value
k_3	0.048676
k_8	$1.025119 \cdot 10^{-5}$
k_{12}	$1.489366 \cdot 10^{-12}$
k_{14}	$8.060631 \cdot 10^{-4}$
k_{15}	0.026272
k_{16}	$3.781311 \cdot 10^{-4}$

Table A.2: Reported p-values for the parameters for which the hypothesis that they are distributed uniformly was rejected (plurip.-promoting gene). Confidence interval of 95%. The p-values for the rest of the kinetic constants are all larger than 0.05.

the viable set. These parameters are the ones deemed essential for the associated behaviour (i.e. plastic behaviour). The null hypothesis for the test is therefore whether the empirical CDF of each k_j for the plastic sets is equal to their empirical distributions within the whole viable set. Since we are using a confidence interval of 95%, whenever the p-value is larger than 0.05 the null hypothesis cannot be rejected, i.e. the parameter is deemed to be uniformly distributed. As reported in Section 2.3.3, the null hypothesis is rejected only for k_1, k_9 and k_{14} (differentiation-promoting gene) and for k_2 and k_6 (pluripotency-promoting gene). The reported p-values are given in Table A.3 and Table A.4, respectively.

Parameter	p-value
k_1	0.023493
k_9	0.037880
k_{14}	$1.492134 \cdot 10^{-4}$

Table A.3: Reported p-values for the parameters for which the hypothesis that they have the same distribution as that within the viable subset was rejected, diff.-promoting gene. Confidence interval of 95%. The p-values for the rest of the kinetic constants are all larger than 0.05.

Parameter	p-value
k_2	0.0310
k_6	0.0425

Table A.4: Reported p-values for the parameters for which the hypothesis that they have the same distribution as that within the viable subset was rejected, plurip.-promoting gene. Confidence interval of 95 %. The p-values for the rest of the kinetic constants are all larger than 0.05.

A.1.3 Reference parameter values: resilient and plastic scenarios

Here we give the sets of parameter values used to produce the phase diagrams from Fig. 2.5, associated both with the reprogramming-resilient phenotype (Fig. 2.5(d)) and the phenotype with elevated plastic potential (Fig. 2.5(f)). These sets of parameter values have been chosen with the same viability criteria as those given in Section 2.2.3, namely, that the mean-field limit has a single stable steady state corresponding to the open (closed) epigenetic state for the differentiation (pluripotency) gene. Besides this, we further require that, for the reprogramming-resilient phenotype there is no overlap of the bistability regions, whereas for the plastic phenotype we require the area between the solid red line and the dashed blue line (see Fig. 2.5(f)) to be positive. The reference parameter values are given in Tables A.5, A.6, A.7 and A.8.

Rescaled parameter	Scaling parameter	Units	Reference
$\kappa_1 = 200$		dimensionless	–
$\kappa_2 = 100$		dimensionless	–
$\kappa_3 = 50$		dimensionless	–
$\kappa_5 = 1$		dimensionless	–
$\kappa_6 = 200$		dimensionless	–
$\kappa_7 = 10$		dimensionless	–
$\kappa_8 = 100$		dimensionless	–
$\kappa_9 = 200$		dimensionless	–
$\kappa_{10} = 100$		dimensionless	–
$\kappa_{11} = 0.1$		dimensionless	–
$\kappa_{12} = 1$		dimensionless	–
$\kappa_{13} = 1$		dimensionless	–
$\kappa_{14} = 200$		dimensionless	–
$\kappa_{15} = 100$		dimensionless	–
$\kappa_{16} = 100$		dimensionless	–
	$E = 100$		–
	$S = 5000$		–

Table A.5: Reference parameter values used for the epigenetic regulation of a differentiation-promoting gene, reprogramming-resilient phenotype (Fig. 2.5(d)).

A.2 Supplementary materials Chapter 3

This Section contains the technical details involved in the formulation and multi-scale analysis of the ER-GRN system introduced in Chapter 3, as well as additional information and results. Specifically:

1. Section A.2.1 is devoted to presenting the multi-scale WKB analysis of the GRN part of the model with the aim of:
    - Deriving the quasi-steady state approximation (QSSA) for the probability density functions (PDFs) of the fast stochastic variables. Such QSSA PDFs are essential to the results of Section 3.2.2.
    - Deriving the (outer) optimal path equations for the gene regulatory system analysed in Chapter 3.
    - Deriving the associated estimate of the relaxation time of the GRN system upon epigenetic-regulatory switch.

Rescaled parameter	Scaling parameter	Units	Reference
$\kappa_1 = 200$		dimensionless	–
$\kappa_2 = 100$		dimensionless	–
$\kappa_3 = 10$		dimensionless	–
$\kappa_5 = 1$		dimensionless	–
$\kappa_6 = 10$		dimensionless	–
$\kappa_7 = 100$		dimensionless	–
$\kappa_8 = 100$		dimensionless	–
$\kappa_9 = 200$		dimensionless	–
$\kappa_{10} = 100$		dimensionless	–
$\kappa_{11} = 10$		dimensionless	–
$\kappa_{12} = 1$		dimensionless	–
$\kappa_{13} = 1$		dimensionless	–
$\kappa_{14} = 100$		dimensionless	–
$\kappa_{15} = 100$		dimensionless	–
$\kappa_{16} = 100$		dimensionless	–
	$E = 100$		–
	$S = 5000$		–

Table A.6: Reference parameter values used for the epigenetic regulation of a pluripotency-promoting gene, reprogramming-resilient phenotype (Fig. 2.5(d)).

Rescaled parameter	Scaling parameter	Units	Reference
$\kappa_1 = 200$		dimensionless	–
$\kappa_2 = 100$		dimensionless	–
$\kappa_3 = 30$		dimensionless	–
$\kappa_5 = 1$		dimensionless	–
$\kappa_6 = 100$		dimensionless	–
$\kappa_7 = 50$		dimensionless	–
$\kappa_8 = 100$		dimensionless	–
$\kappa_9 = 200$		dimensionless	–
$\kappa_{10} = 100$		dimensionless	–
$\kappa_{11} = 0.1$		dimensionless	–
$\kappa_{12} = 1$		dimensionless	–
$\kappa_{13} = 1$		dimensionless	–
$\kappa_{14} = 200$		dimensionless	–
$\kappa_{15} = 80$		dimensionless	–
$\kappa_{16} = 70$		dimensionless	–
	$E = 100$		–
	$S = 5000$		–

Table A.7: Reference parameter values used for the epigenetic regulation of a differentiation-promoting gene, plastic phenotype (Fig. 2.5(f)).

- In Sections A.2.1.3 and A.2.1.4 we derive the quasi-steady state probability density functions of the fast variables of the ER-GRN stochastic model.
- In Section A.2.1.5 we describe the formulation of the numerical method associated with the asymptotic model reduction of the stochastic ER-GRN model.

2. Section A.2.2 contains a comparison between our asymptotic results (both WKB and stochastic model reduction) for the benchmark case of a self-activating one-gene regulatory system.

3. In Section A.2.3 we briefly present the details of the numerical method for action minimisation used to implement the Minimum Action Path (MAP).

4. In Section A.2.4 we present the p-values corresponding to the Kolmogorov-Smirnov tests performed in Section 3.3.3.

Rescaled parameter	Scaling parameter	Units	Reference
$\kappa_1 = 200$		dimensionless	–
$\kappa_2 = 100$		dimensionless	–
$\kappa_3 = 10$		dimensionless	–
$\kappa_5 = 1$		dimensionless	–
$\kappa_6 = 50$		dimensionless	–
$\kappa_7 = 100$		dimensionless	–
$\kappa_8 = 100$		dimensionless	–
$\kappa_9 = 200$		dimensionless	–
$\kappa_{10} = 100$		dimensionless	–
$\kappa_{11} = 8$		dimensionless	–
$\kappa_{12} = 1$		dimensionless	–
$\kappa_{13} = 1$		dimensionless	–
$\kappa_{14} = 100$		dimensionless	–
$\kappa_{15} = 100$		dimensionless	–
$\kappa_{16} = 100$		dimensionless	–
	$E = 100$		–
	$S = 5000$		–

Table A.8: Reference parameter values used for the epigenetic regulation of a pluripotency-promoting gene, plastic phenotype (Fig. 2.5(f)).

5. In Section A.2.5 we include the parameter values used in the different simulations of Chapter 3.

6. In Section A.2.6 we include a number of supplementary figures with results complementary to those presented in Chapter 3.

A.2.1 Multiscale analysis of the GRN system: WKB approximation and multi-scale optimal path theory

In this appendix we provide a detailed derivation of the multiscale analysis of the gene regulatory (sub)system. In particular, we derive the quasi-steady multinomial probability distribution describing the QSSA of the occupancy of binding sites within the promoter region of the epigenetically opened genes. Our current treatment considers an extension of the multiple scale WKB method that was first introduced in [1, 40]. We will show that under suitable conditions on separation of the different

characteristic scales of the system (which will be precisely defined below), the WKB approximation of the outer solution collapses onto a trivial WKB mode. By contrast, within the corresponding inner regime, the solution is given by a non-trivial (the so-called fast) WKB mode, which turns out to take the form of a multinomial distribution.

Before proceeding further, it is worth noting that we focus our analysis on the GRN subsystem alone for clarity. The same analysis can be performed with the ER part of the whole system.

A.2.1.1 General setting: WKB and multi-scale optimal path theory

The WKB approximation and the (closely related) optimal path theory are variants of large deviation theory which has been used to study mean first passage time problems associated with the escape from metastable states in a wide variety of contexts, including multi-stable systems [45, 103], extinctions in population-dynamical models [46], hybrid stochastic systems [21], gene regulatory systems [181, 170], and small stochastic systems [134]. This methodology has recently been extended to account for, and take advantage of, separation of time scales. In this section we give a brief summary of this methodology. For a full account of the technicalities we refer to [1, 40].

An alternative formulation to analyse the dynamics of continuous-time Markov processes on a discrete space of states is to derive a partial differential equation for the generating function, $G(p_1, \ldots, p_{N_G}, p_{11}, \ldots, p_{N_G N_G}, t)$:

$$\begin{aligned} & G(p_1, \ldots, p_{N_G}, p_{11}, \ldots, p_{N_G N_G}, t) = \\ = & \sum_{\mathbf{X}} \left(\prod_{i=1}^{N_G} p_i^{X_i} \right) \left(\prod_{i,j=1}^{N_G} p_{ij}^{X_{ij}} \right) P(X_1, \ldots, X_{N_G}, X_{11}, \ldots, X_{N_G N_G}, t), \end{aligned} \tag{A.2.1}$$

where $P(X_1, \ldots, X_{N_G N_G}, t)$ is the solution of the Master Equation. The probability density generating function (PDGF) satisfies a partial differential equation (PDE) which can be derived from the corresponding Master Equation, Eq. (2.2.1). This PDE is the basic element of the so-called momentum representation of the Master Equation [43, 126, 7, 8, 84]:

$$\frac{\partial G}{\partial t} =$$
$$H_k\left(p_1,\ldots,p_{N_G},p_{11},\ldots,p_{N_GN_G},\partial_{p_1},\ldots,\partial_{p_{N_G}},\partial_{p_{11}},\ldots,\partial_{p_{N_GN_G}}\right)G(p_1,\ldots,p_{N_G},p_{11},\ldots,p_{N_GN_G},t) \tag{A.2.2}$$

This PDE, or, equivalently, the operator H_k, is determined by the reaction rates shown in Table 3.2: it is obtained by multiplying both sides of the corresponding Master Equation by $\left(\prod_{i=1}^{N_G} p_i^{X_i}\right)\left(\prod_{i,j=1}^{N_G} p_{ij}^{X_{ij}}\right)$ and summing up over all the possible values of $\mathbf{X}$. Furthermore, the solution of Eq.(A.2.2) must satisfy the normalisation condition $G(p_1=1,\ldots,p_{N_G}=1,p_{11}=1,\ldots,p_{N_GN_G}=1,t)=1$ for all t.

Eq.(A.2.2) allows us to define a Hamiltonian, $H_k(p,Q)$, where the position operators in the momentum representation have been defined as $Q_i \equiv \partial_{p_i}$ with the commutation relation $[Q_i,p_j]=S\delta_{i,j}$, where each pair (p_i,Q_i) is the set of generalised coordinates associated with the random variable X_i. The quantity S is the characteristic scale corresponding to the random variable X_i. Similarly, $Q_{ij}\equiv\partial_{p_{ij}}$ with $[Q_{ij},p_{kl}]=E\delta_{i,k}\delta_{j,l}$. Each pair (p_{ij},Q_{ij}) has the random variable X_{ij} associated, and the quantity E is the characteristic scale associated with the random variable X_{ij} [40].

In order to proceed with our multiple time scale analysis, we assume, as per the Briggs-Haldane treatment of the Michaelis-Menten model for enzyme kinetics [22, 88], that the species involved in the system under scrutiny are divided into two groups according to their characteristic scales. More specifically, we have a subset of chemical species whose numbers, X_i, scale as $X_i=Sx_i$, where $x_i=O(1)$, whilst the remaining species are such that their numbers, X_{ij}, scale as $X_{ij}=Ex_{ij}$, where $x_{ij}=O(1)$. Key to our approach is the fact that S and E must be such that $\epsilon=\frac{E}{S}\ll 1$. We further assume that the generalised coordinates, Q_i, scale in the same fashion as the corresponding variables X_i, i.e. $Q_i=Sq_i$, where $q_i=O(1)$. We refer to the variables belonging to this subset as *slow variables*. Similarly, $Q_{ij}=Eq_{ij}$, where $q_{ij}=O(1)$, which are referred to as *fast variables*. Moreover, we assume that the moment coordinates, p_i and p_{ij}, are all independent of S and E, and therefore remain invariant under rescaling.

Under this scaling for the generalised coordinates, we define the following scale transformation for the Hamiltonian in Eq. (A.2.2):

$$H_k(p,Q) = k_J S^k E^l H_\omega(p,q) \tag{A.2.3}$$

where J identifies the reaction with the largest order among all the reactions that compose the dynamics and k_J is the corresponding rate constant. The exponents k and l correspond to the number of slow and fast variables involved in the transition rate W_J, respectively (see Table 3.2). The last step is to rescale the time variable so that a dimensionless variable, τ, is defined such that:

$$\tau = k_J S^{k-1} E^l t \tag{A.2.4}$$

It is now a trivial exercise to check that, upon rescaling, Eq. (A.2.2) reads:

$$\frac{\partial G}{\partial \tau} = \\ SH_\omega\left(p_1,\ldots,p_{N_G},p_{11},\ldots,p_{N_GN_G},\partial_{p_1},\ldots,\partial_{p_{N_G}},\partial_{p_{11}},\ldots,\partial_{p_{N_GN_G}}\right) G(p_1,\ldots,p_{N_G},p_{11},\ldots,p_{N_GN_G},\tau) \tag{A.2.5}$$

Closed, analytic solutions are rarely available for Eq. A.2.5. Nevertheless, the PDE for the generating function admits a perturbative solution, which is commonly obtained by means of the WKB method [46, 8]. From a formal point of view, Eq. (A.2.5) is a Schrödinger-like equation and, therefore, there is a plethora of methods at our disposal in order to analyse it. In particular, when the fluctuations are (assumed to be) small, it is common to resort to WKB methods [93, 2, 63], which consists on proposing the following Ansatz for the solution of Eq. (A.2.5):

$$G(p,\tau) = e^{SA(p,\tau)},$$

where $A(p,\tau)$, the so-called action, satisfies the Hamilton-Jacobi equation associated with the Hamiltonian $H_\omega(p,q)$:

$$\frac{\partial A}{\partial \tau} = H_\omega\left(p, \frac{\partial A}{\partial p}\right). \tag{A.2.6}$$

The solution to the above Hamilton-Jacobi equation, Eq. (A.2.6), can be given in terms of the solution of the associated Hamilton equations:

$$\frac{dp_i}{d\tau} = -\frac{\partial H_\omega}{\partial q_i}, \tag{A.2.7}$$

$$\frac{dq_i}{d\tau} = \frac{\partial H_\omega}{\partial p_i}, \tag{A.2.8}$$

and it is given by:

$$\begin{aligned}\mathcal{A}(p,\tau) &= -H_\omega \tau - \int^p \sum_i q_i(p) dp_i + \mathcal{A}_0(p) = \\ &= -H_\omega \tau - \int^\tau \left(\sum_{i=1}^{N_G} q_i(s)\dot{p}_i(s) + \epsilon \sum_{i,j=1}^{N_G} q_{ij}(s)\dot{p}_{ij}(s) \right) ds + \mathcal{A}_0(p)\end{aligned} \tag{A.2.9}$$

where the first (second) summatory is over the set of slow (fast) variables, and $\mathcal{A}_0(p)$ is an integration constant, whose value is chosen so that $\mathcal{A}(p=1,\tau)=0$ for all τ.

We define $A_0(\tau)$ as the value of the action functional Eq. (A.2.9) calculated on the path which maximises its value, i.e. the optimal path. To determine the optimal path we therefore need to solve the variational problem $\delta\mathcal{A}=0$. The Euler-Lagrange equations associated with this variational problem are the Hamilton equations corresponding to the Hamiltonian Eq. (A.2.3):

$$\frac{dp_i}{d\tau} = -\frac{\partial H_\omega}{\partial q_i}, \tag{A.2.10}$$

$$\frac{dq_i}{d\tau} = \frac{\partial H_\omega}{\partial p_i}, \tag{A.2.11}$$

for the slow variables, and

$$\epsilon\frac{dp_{ij}}{d\tau} = -\frac{\partial H_\omega}{\partial q_{ij}}, \tag{A.2.12}$$

$$\epsilon\frac{dq_{ij}}{d\tau} = \frac{\partial H_\omega}{\partial p_{ij}}, \tag{A.2.13}$$

for the fast variables, with $\epsilon = E/S$. These equations are (formally) solved with boundary conditions [46]: $q_i(0)=q_{0_i}$, $q_{ij}(0)=q_{0_{ij}}$, $p_i(\tau)=p_i$, and $p_{ij}(0)=p_{0_{ij}}$.

Eqs. (A.2.10)-(A.2.13) show that, if the scaling Eq. (A.2.3) and $\epsilon = E/S \ll 1$ hold, our optimal path theory exhibits separation of time scales, which can be exploited to simplify our analysis by means of a multi-scale asymptotic approximation.

A.2.1.1.1 Multi-scale analysis: Outer regime.

Multi-scale asymptotic analysis [76] is predicated upon the construction of two approximations: one which is valid in the long term, the so-called *outer solution*, and another one which approximates the behaviour of the system at shorter times, the so-called, *inner solution*. These two regimes must satisfy the appropriate matching conditions which ensure that both solutions produce a uniformly valid approximation. The outer solution is usually obtained in terms of the quasi-steady state approximation, which describes the dynamics of the system once it has settled down onto the associated invariant manifold. Eqs. (A.2.10)-(A.2.13) are the starting point for the formulation of the semi-classical quasi-steady state approximation (SCQSSA) [1, 40]. The QSS approximation consists on assuming that $\epsilon\frac{dp_{ij}}{d\tau} \simeq 0$ and $\epsilon\frac{dq_{ij}}{d\tau} \simeq 0$ in Eqs. (A.2.12)-(A.2.13),

$$-\frac{\partial H_\omega}{\partial q_{ij}} = 0, \tag{A.2.14}$$

$$\frac{\partial H_\omega}{\partial p_{ij}} = 0, \tag{A.2.15}$$

resulting in a differential-algebraic system of equations which provides us with the semi-classical quasi-steady state approximation (SCQSSA).

A.2.1.1.2 Multi-scale analysis: Inner regime.

To obtain the equations of the inner approximation, one usually proceeds first to re-scale the variable τ: $T \equiv \epsilon^{-1}\tau$. Under this re-scaling the equations for the slow variables Eqs. (A.2.10)-(A.2.11) become:

$$\frac{dp_i}{dT} = O(\epsilon), \tag{A.2.16}$$

$$\frac{dq_i}{dT} = O(\epsilon), \tag{A.2.17}$$

which, at the lowest order, imply that $p_i(T) \simeq p_i(T=0) \equiv p_{0_i}$ and $q_i(T) \simeq q_i(T=0) \equiv q_{0_i}$ within the inner regime. By contrast, the equations for the fast variables now read:

$$\frac{dp_{ij}}{dT} = -\left.\frac{\partial H_\omega}{\partial q_{ij}}\right|_{p_{ll}(T)=p_{0_{ll}},q_{ll}(T)=q_{0_{ll}}}, \tag{A.2.18}$$

$$\frac{dq_{ij}}{dT} = \left.\frac{\partial H_\omega}{\partial p_{ij}}\right|_{p_{ll}(T)=p0_{ll},q_{ll}(T)=q0_{ll}}, \tag{A.2.19}$$

where the index $l = 1, \ldots, N_G$.

A.2.1.2 Multi-scale optimal path theory: estimation of the GRN relaxation time upon epigenetic switch

The general procedure to compute the Hamiltonian associated with the characteristic function PDE, Eq. (A.2.2), is described in Section A.2.1.1. In this section we provide the details of the optimal path theory analysis of the stochastic model of the gene regulatory network with competitive inhibition, as described in Section 3.2.1 and in Table 3.2, using the procedure outlined in Section A.2.1.1. According to such procedure, the resulting Hamiltonian, $H_k(p, Q)$, associated with the generating function PDE can be written as:

$$\begin{aligned} H_k(p,Q) = \quad & \sum_{i=1}^{N_G}(p_i - 1)\left(\hat{R}_i + k_{i1}p_{ii}Q_{ii} - k_{i2}Q_i\right) + \\ & + \sum_{i,j=1}^{N_G}\left(p_{ij} - p_j^2\right)\left(b_{ij}\left(e_i - \sum_{k=1}^{N_G} p_{ik}Q_{ik}\right)Q_j^2 - u_{ij}Q_{ij}\right) \end{aligned} \tag{A.2.20}$$

where the pairs (p_i, Q_i) and (p_{ij}, Q_{ij}) are associated with the random variables X_i, i.e. the number of protein molecules transcribed by gene i, and X_{ij}, namely, the number of sites in the promoter of gene i bound to dimers of protein j, respectively.

According to the multi-scale optimal path theory formulation, the Hamiltonian Eq. (A.2.20) satisfies the following scaling relationship:

$$H_k(p, Q) = b_{11}ES^2 H_\omega(p, q).$$

where the re-scaled variables and parameters are defined in Table A.9.

The re-scaled Hamiltonian, $H_\omega(p, q)$, is given by:

$$H_\omega(p,q) = \quad \sum_{i=1}^{N_G}(p_i - 1)\left(R_i + \omega_{i1}p_{ii}q_{ii} - \omega_{i2}q_i\right) +$$

Table A.9: Re-scaled variables and dimensionless parameters.

Rescaled variables	Dimensionless parameters
$\tau = b_{11}ESt$	$\epsilon = E/S$, $R = \hat{R}/(b_{11}ES^2)$
$q_i = Q_i/S$	$\omega_{i1} = k_{i1}/(b_{11}S^2)$, $\omega_{i2} = k_{i2}/(b_{11}ES)$,
$q_{ij} = Q_{ij}/E$	$\beta_{ij} = b_{ij}/b_{11}$, $\delta_{ij} = u_{ij}/(b_{11}S^2)$

$$+\sum_{i,j=1}^{N_G}\left(p_{ij}-p_j^2\right)\left(\beta_{ij}\left(p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_j^2-\delta_{ij}q_{ij}\right) \tag{A.2.21}$$

where $p_{\infty_i}p = \frac{e_i}{E}$. The remaining re-scaled variables and parameters are given in Table A.9.

Eq. (A.2.21) allows to write down the set of Hamilton equations which determine the optimal fluctuational path. These equations read:

$$\begin{aligned}
\frac{dq_i}{d\tau} &= R_i+\omega_{i1}p_{ii}q_{ii}-\omega_{i2}q_i- \\
&- 2p_j\sum_{i=1}^{N_G}\left(\beta_{ij}\left(p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_j^2-\delta_{ij}q_{ij}\right) && \text{(A.2.22)}\\
\epsilon\frac{dq_{ij}}{d\tau} &= (p_i-1)\omega_{i1}q_{ii}\delta_{ij}+\beta_{ij}\left(p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_j^2-\delta_{ij}q_{ij}- \\
&- \left(\sum_{k=1}^{N_G}\left(p_{ik}-p_k^2\right)\beta_{ik}q_{0_k}^2\right)q_{ij} && \text{(A.2.23)}\\
\frac{dp_i}{d\tau} &= (p_i-1)\omega_{i2}-\sum_{l=1}^{N_G}\left(p_{li}-p_i^2\right)\left(\beta_{li}\left(p_{\infty_i}p-2\sum_{k=1}^{N_G}a_{lk}p_{lk}q_{lk}\right)q_i\right) && \text{(A.2.24)}\\
\epsilon\frac{dp_{ij}}{d\tau} &= -(p_i-1)\omega_{i1}p_{ii}\delta_{ij}+\left(\sum_{l=1}^{N_G}\left(p_{il}-p_l^2\right)\beta_{il}\right)p_{ij}+\left(p_{ij}-p_j^2\right)\delta_{ij} && \text{(A.2.25)}
\end{aligned}$$

Eqs. (A.2.22)-(A.2.25) show that the scaling of the Hamiltonian and its associated

generalised coordinates, shown in Table A.9, makes evident that this system exhibits multiple time scales structure where the pairs of coordinates (p_i, q_i) are slow variables which control the long-time behaviour of the system, whereas the pairs (p_{ij}, q_{ij}) are fast variables, which very quickly settle onto the invariant manifold determined by the quasi-steady state approximation.

A.2.1.2.1 Multi-scale analysis: outer solution.

The outer (long time) solution of the optimal path theory equations, Eqs. (A.2.10)-(A.2.13) is obtained by applying the QSS approximation to the fast variables, i.e. the pairs (p_{ij}, q_{ij}). It is straightforward to verify that, by taking $\epsilon \dot{p}_{ij} \simeq 0$ and $\epsilon \dot{q}_{ij} \simeq 0$, for all i and j, and setting $p_i(\tau) = 1$ for all i within the outer regime, and after some algebra, we obtain the associated SCQSSA equations:

$$\frac{dq_i}{d\tau} = R_i + p_{\infty_i} p \omega_{i1} \frac{\beta_{ii} q_i^2}{\delta_{ii} + \delta_{ii} \sum_k \frac{\beta_{ik}}{\delta_{ik}} q_k^2} - \omega_{i2} q_i \tag{A.2.26}$$

$$q_{ij}(\tau) = p_{\infty_i} p \frac{\beta_{ij} q_j^2}{\delta_{ij} + \delta_{ij} \sum_k \frac{\beta_{ik}}{\delta_{ik}} q_k^2} \tag{A.2.27}$$

$$p_{ij}(\tau) = 1 \tag{A.2.28}$$

As shown in [1, 40, 53], the quantities $p_{\infty_i} p$ are given by:

$$p_{\infty_i} p = \frac{e_i}{E},$$

where e_i is the number of binding sites within the promoter region of gene i. The mean-field limit of Eqs. (A.2.26)-(A.2.27) is dictated by taking $p_{\infty_i} p = 1$ for all i which prescribes that the number of binding sites in all genes is exactly equal to its average over a population of cells. It is therefore clear that the SCQSSA equations with $p_{\infty_i} p \neq 1$ will, in general, behave differently than their mean-field counterparts. In particular, Eqs. (A.2.26)-(A.2.27) may allow transitions between states that are forbidden by the mean-field dynamics. This possibility will be discussed at length in the next section in connection to the robustness of phenotypic states (see also de la Cruz et al. [40]).

Furthermore, it is straightforward to verify that the action functional Eq. (A.2.9) calculated on the outer solution, $\mathcal{A}_{QSS}$, at the lowest order vanishes: $\mathcal{A}_{QSS} = 0$. Therefore, the only positive contribution to the action must come from the inner solution.

A.2.1.2.2 Multi-scale analysis: inner solution and matching conditions.

According to the analysis carried out in Section A.2.1.2, in order to fully characterise the behaviour of the stochastic competitive binding model, we must address the short time regime and formulate the appropriate matching conditions with the SCQSSA. The matching conditions are prescribed as follows:

1. The outer solution is such that $p_i(\tau) = 1$ for all i. This implies that, within the inner regime, where $p_i(T) \simeq \text{cnt.} \equiv p_{0_i}$, then, $p_{0_i} = 1$ for all i.

2. The SCQSS approximation, Eqs. (A.2.26)-(A.2.28), is such that, on the SCQSSA trajectory, $H_\omega = 0$. Therefore, the inner solution and the initial conditions should be such that

$$H_\omega \left(p_{0_i} = 1, p_{ij}(T), q_{0_i}, q_{ij}(T)\right) = 0.$$

3. The outer solution is such that $p_{ij}(\tau) = 1$ for all i and j. Therefore,

$$\lim_{T\to\infty} p_{ij}(T) = 1$$

Taking into account Matching Condition 1 and Eqs. (A.2.26)-(A.2.28), the *inner Hamiltonian*, $H_{\omega_{in}}(p_{ij}, q_{ij}) \equiv H_\omega \left(p_{0_i} = 1, p_{ij}(T), q_{0_i}, q_{ij}(T)\right)$:

$$H_{\omega_{in}}(p_{ij}, q_{ij}) = \sum_{i,j=1}^{N_G} \left(p_{ij} - 1\right) \left(\beta_{ij} \left(p_{\infty_i} p - \sum_{k=1}^{N_G} p_{ik} q_{ik}\right) q_{0_j}^2 - \delta_{ij} q_{ij}\right) \tag{A.2.29}$$

Since, as per Matching Condition 2, the inner solution must satisfy $H_{\omega_{in}}(p_{ij}, q_{ij}) = 0$, this yields to two solutions: the so-called fast mode,

$$p_{ij} = 1, \tag{A.2.30}$$

for all i and j, which implies that the action integral vanishes on the fast mode. The so-called slow mode is given by the solution of the following system of algebraic equations:

$$\beta_{ij} \left(p_{\infty_i} p - \sum_{k=1}^{N_G} p_{ik} q_{ik}\right) q_{0_j}^2 - \delta_{ij} q_{ij} = 0 \tag{A.2.31}$$

for all i and j. Since Eqs. (A.2.31) hold for any i and j, let us fix i and write down Eqs. (A.2.31) for two generic j's, j_1 and j_2,

$$\beta_{ij_1}\left(p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_{0_{j_1}}^2-\delta_{ij_1}q_{ij_1}=0 \tag{A.2.32}$$

$$\beta_{ij_2}\left(p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_{0_{j_2}}^2-\delta_{ij_2}q_{ij_1}=0 \tag{A.2.33}$$

from where it can be obtained,

$$p_{\infty_i}p-\sum_{k=1}^{N_G}p_{ik}q_{ik}=\frac{\delta_{ij_1}q_{ij_1}}{\beta_{ij_1}q_{0_{j_1}}^2}=\frac{\delta_{ij_2}q_{ij_2}}{\beta_{ij_2}q_{0_{j_2}}^2}, \tag{A.2.34}$$

which imply that:

$$q_{ik}=\frac{\beta_{ik}\delta_{ij}q_{0_k}^2}{\beta_{ij}\delta_{ik}q_{0_j}^2}q_{ij}, \tag{A.2.35}$$

for any i, k and j, with $k\neq j$. Using Eqs. (A.2.35), Eqs. (A.2.31) become:

$$\beta_{ij}q_{0_j}^2p_{\infty_i}p-\delta_{ij}\left(\sum_{k=1}^{N_G}\frac{\beta_{ik}}{\delta_{ik}}p_{ik}q_{0_k}^2+1\right)q_{ij}=0$$

which implies that:

$$q_{ij}(T)=p_{\infty_i}p\frac{\beta_{ij}q_{0_j}^2}{\delta_{ij}+\delta_{ij}\sum_k\frac{\beta_{ik}}{\delta_{ik}}q_{0_k}^2p_{ik}(T)} \tag{A.2.36}$$

Note that the slow mode cannot be determined by energy conservation, $H_\omega(p,q)=0$, alone. In order to circumvent this obstacle, we proceed as follows. Eqs. (A.2.36) must be supplemented with two extra equations, which are provided by the (inner) equations of motion, in particular the equations for the fast generalised coordinates $q_{ij}(T)$ (see Eq. (A.2.19)). Taking into account that $p_i(T)=1$ and that the slow mode must satisfy Eqs. (A.2.36), the equations of motion for the fast generalised coordinates $q_{ij}(T)$, read:

$$\frac{dq_{ij}}{dT}=-\left(\sum_{l=1}^{N_G}\left(p_{il}-1\right)\beta_{il}q_{0_l}^2\right)q_{ij}. \tag{A.2.37}$$

From Eqs. (A.2.37), it is straightforward to verify that:

$$\frac{\dot{q}_{ij}}{\dot{q}_{ik}} = \frac{dq_{ij}}{dq_{ik}} = \frac{q_{ij}}{q_{ik}}, \tag{A.2.38}$$

We now turn our attention to the *inner action integral*, $\mathcal{A}_{in}$, which is given by Eq. (A.2.9) computed on the slow mode:

$$\mathcal{A}_{in} = -\epsilon \int_0^\infty \left(\sum_{i,j=1}^{N_G} q_{ij}(T) \frac{dp_{ij}}{dT} \right) dT, \tag{A.2.39}$$

where it has been used that $p_i(T)$ is constant within the inner regime. Before proceeding further, we define $\mathcal{A}_{in} = \sum_{i=1}^{N_G} \mathcal{A}_{in_i}$, where:

$$\mathcal{A}_{in_i} \equiv -\epsilon \int_0^\infty \left(\sum_{j=1}^{N_G} q_{ij}(T) \frac{dp_{ij}}{dT} \right) dT \tag{A.2.40}$$

Consider now Eq. (A.2.40). Integrating by parts we obtain:

$$\mathcal{A}_{in_i} = -\epsilon \left(\sum_{j=1}^{N_G} p_{ij} q_{ij} \Big|_{T=0}^{T=\infty} - \int_0^\infty \dot{q}_{ii} \left(p_{ii} + \sum_{j \neq i} p_{ij} \frac{\dot{q}_{ij}}{\dot{q}_{ii}} \right) dT, \right)$$

which, since $\frac{\dot{q}_{ij}}{\dot{q}_{ii}} = \frac{q_{ij}}{q_{ii}}$ (see Eqs. (A.2.38)), becomes:

$$\mathcal{A}_{in_i} = -\epsilon \left(\sum_{j=1}^{N_G} p_{ij} q_{ij} \Big|_{T=0}^{T=\infty} - \int_0^\infty \frac{\dot{q}_{ii}}{q_{ii}} \left(\sum_{j=1}^{N_G} p_{ij} q_{ij} \right) dT, \right)$$

Furthermore, Eq. (A.2.31) implies that:

$$\sum_{j=1}^{N_G} p_{ij} q_{ij} = p_{\infty_i} p - \frac{\delta_{ii}}{\beta_{ii} q_{0_i}^2} q_{ii},$$

therefore:

$$\mathcal{A}_{in_i} = \epsilon \left(\frac{\delta_{ii}}{\beta_{ii} q_{0_i}^2} q_{ii} \Big|_{q_{0_{ii}}}^{q_{ii_{out}}} + \int_{q_{0_{ii}}}^{q_{ii_{out}}} \left(\frac{p_{\infty_i} p}{q_{ii}} - \frac{\delta_{ii}}{\beta_{ii} q_{0_i}^2} \right) dq_{ii}, \right) = \epsilon p_{\infty_i} p \log \left(\frac{q_{ii_{out}}}{q_{0_{ii}}} \right), \tag{A.2.41}$$

where

$$q_{ii_{out}} = p_{\infty_i} p \frac{\beta_{ii} q_{0_i}^2}{\delta_{ii} + \delta_{ii} \sum_k \frac{\beta_{ik}}{\delta_{ik}} q_{0_k}^2},$$

which is dictated by the matching conditions, and $q_{0_{ii}} = q_{ii}(T = 0)$. Therefore, $\mathcal{A}_{in}$ is given by

$$\mathcal{A}_{in} = \epsilon \left(\sum_{i=1}^{N_G} p_{\infty_i} p \log \left(\frac{q_{ii_{out}}}{q_{0_{ii}}} \right) \right) = \frac{1}{S} \left(\sum_{i=1}^{N_G} e_i \log \left(\frac{q_{ii_{out}}}{q_{0_{ii}}} \right) \right) \tag{A.2.42}$$

where, for $\mathcal{A}_{in} > 0$, it is sufficient that the initial conditions for q_{ij} satisfy that $q_{0_{ii}} < q_{ii_{out}}$. The quantities $q_{0_{ii}} = q_{ii}(T = 0)$ must be such that $H_\omega(p_{0_i}, p_{0_{ij}}, q_{0_i}, q_{0_{ij}}) = 0$. Moreover, since $p_{0_i} q_{0_i} = \frac{X_i(t=0)}{S}$ and $p_{0_{ij}} q_{0_{ij}} = \frac{X_{ij}(t=0)}{E}$, besides the energy condition, $p_{0_i} q_{0_i} \geq 0$ and $0 \leq p_{0_{ij}} q_{0_{ij}} \leq \frac{e_i}{E}$ must hold for all i, j.

By insering Eq. (A.2.42) into the WKB Ansatz,

$$G(p_{i1}, \ldots, p_{iN_G}) = e^{S\mathcal{A}_{in}} = \left(\frac{q_{ii_{out}}}{q_{0_{ii}}} \right)^{e_i}, \tag{A.2.43}$$

we obtain that $P_+(\mathbf{B}_i|\mathbf{N})$ (see Eq. 3.2.11), with $\mathbf{N} = (X_1, \ldots, X_{N_G})$, is a multinomial distribution since the correponding generating function, $G(p_1, \ldots, p_{N_G})$, is given by:

$$G(p_{i1}, \ldots, p_{iN_G}) = \left(\frac{1 + \sum_k \frac{\beta_{ik}}{\delta_{ik}} x_k^2 p_{ik}}{1 + \sum_k \frac{\beta_{ik}}{\delta_{ik}} x_k^2} \right)^{e_i} \tag{A.2.44}$$

where we have taken $x_k = p_{0_k} q_{0_k} = q_{0_k}$.

A.2.1.3 Consistency with the stochastic model reduction method

In this section, we check the consistency between the multiscale WKB method presented here and the model reduction methodology developed in Section 3.2.2. For concreteness, we focus the discussion in the derivation of the quasi-steady state distribution of the binding/unbinding sub-system (GRN). As discussed in Section 3.2.2, the fast variables of the GRN sub-model, i.e. the TF-promoter binding site dimers, are in quasi-equilibrium with the slow variables, i.e. the protein products [25, 26, 10, 143]. We therefore consider an (*inner*) approximation where the slow variables are frozen during the (fast) evolution of the binding/unbinding subsystem

towards equilibrium. We start by writing a Master equation for a reduced set of reactions given by:

$$\text{Free binding sites} + 2X_j \underset{K_{ij_2}}{\overset{K_{ij_1}}{\rightleftarrows}} X_{ij} \text{ for all } i,$$

where $K_{ij_1} = k_{ij_1}(e_i - \sum_l X_{il})$, where $k_{ij_1} = b_{ij}H(Y_{i_3}-Y_0)X_j(X_j-1)$, and $K_{ij_2} = u_{ij}$. Recall that X_j, the number of protein transcript of gene j is a slow variable for all j, and, therefore, within the inner approximation, they are considered to be constant. The associated reduced Master Equation (RME) is given by:

$$\begin{aligned}\frac{\partial P(\mathcal{X}_i,t)}{\partial t} &= \sum_j k_{ij_1}\left(\left(e_i - \sum_{l\neq j} X_{il} - (X_{ij}-1)\right)P(\mathcal{X}_i - r_j,t) - \left(e_i - \sum_l X_{il}\right)P(\mathcal{X}_i,t)\right) \\ &+ \sum_j u_{ij}\left((X_{ij}+1)\,P(\mathcal{X}_i + r_j,t) - X_{ij}P(\mathcal{X}_i,t)\right) \qquad \text{(A.2.45)}\end{aligned}$$

where $\mathcal{X}_i = (X_{i1},\ldots,X_{iN_G})$. Before proceeding forward, we made an additional assumption regarding time scale separation, namely, for all i, the RME Eq. (A.2.45) reaches its (quasi-)equilibrium state in a much shorter time than the average switching times of the ER sub-system. In other words, we can consider separately the steady state solution for open and silenced genes and that, upon ER switch, the equilibrium distribution of the binding/unbinding system "instantaneously" switches to the associated equilibrium distribution. The feasability of this assumption is illustrated below, where we analyse the case of one single self-activating gene, which can be fully solved analytically.

We thus consider two steady-state solutions of Eq. (A.2.45): $P_+(\mathcal{X}_i)$ associated with an open gene (i.e. $H(Y_{i_3} - Y_0) = 1$) and $P_-(\mathcal{X}_i)$, related to a silenced gene (i.e. $H(Y_{i_3} - Y_0) = 0$). $P_-(\mathcal{X}_i)$ is trivial: since in a silenced gene TF dimers can only detach from the promoter, the corresponding steady state solution is:

$$P_-(\mathcal{X}_i) = \prod_{j\in\langle i\rangle} \delta_{X_{ij},0}, \qquad \text{(A.2.46)}$$

with $\delta_{X_{ij},0} = 1$ if $X_{ij} = 0$ and $\delta_{X_{ij},0} = 0$ otherwise. An analytical expression for $P_+(\mathcal{X}_i)$ can be found using WKB asymptotics. By re-scaling variables in the RME, we derive the following partial differential equation for the associated generating function (see Section A.2.1.1):

$$\frac{\partial G(p_{i1},\ldots,p_{iN_G},\tau)}{\partial \tau} = SH_{\omega_{in}}(p_{i1},\ldots,p_{iN_G},\partial_{p_{i1}},\ldots,\partial_{p_{iN_G}})G(p_{i1},\ldots,p_{iN_G},\tau) \tag{A.2.47}$$

where $H_{\omega_{in}}(p_i,q_i)$ is derived in detail in Section A.2.1.2 (see under heading *Multi-scale analysis: inner solution and matching conditions*), with $p_i = (p_{i1},\ldots,p_{iN_G})$ and $q_i = (q_{i1},\ldots,q_{iN_G})$, and is given by Eq. (A.2.29):

$$H_{\omega_{in}}(p_i,q_i) = \sum_{i,j=1}^{N_G}(p_{ij}-1)\left(\beta_{ij}\left(\frac{e_i}{E}-\sum_{k=1}^{N_G}p_{ik}q_{ik}\right)q_{0_j}^2-\delta_{ij}q_{ij}\right)$$

Note that since $H_{\omega_{in}}(p_i,q_i)$ is identical to the inner Hamiltonian associated with the multiscale WKB method, the corresponding WKB solution of Eq. (A.2.47), derived from the stochastic model reduction, is identical to the one provided by the multiscale WKB method.

A.2.1.4 Quasi-steady state distribution of the enzyme/complex sub-system (ER).

Similarly to the inner approximation obtained for the fast GRN variables, we can find a QSSA for the fast variables corresponding to the ER component of the system, namely, the enzymes and complexes associated with the ER enzymatic reactions. We will focus on the number of complexes formed, since the free enzyme molecules can be obtained using the associated conservation law [53]. We start by writing the Master Equation corresponding to a reduced set of reactions given by:

$$\mathsf{Substrate} + Y_{ij-1} \underset{K_{ij_2}}{\overset{K_{ij_1}}{\rightleftarrows}} Y_{ij} \text{ for all } i \text{ and } j = 5,7$$

which affect only the fast ER variables (see Table 3.1). The number of substrate molecules is a slow variable for all i and j, and, therefore, within the inner approximation is considered constant. According to Table 3.3, and denoting by $\epsilon_1 = \frac{E}{S}$ and $\epsilon_3 = \frac{Z}{Y}$, with $O(E) = O(Y)$, the corresponding rates are $K_{i5_1} = \frac{\epsilon_1}{\epsilon_3}(\kappa_{i1} + \kappa_{i4}y_{i3})(p_{HDM}p - y_{i5})y_{i2}$, $K_{i5_2} = \frac{\epsilon_1}{\epsilon_3}(\kappa_{i2} + \kappa_{i5}y_{i3})y_{i5}$, $K_{i7_1} = \frac{\epsilon_1}{\epsilon_3}(\kappa_{i9} + \kappa_{i12}y_{i2})(p_{HDAC}p - y_{i7})y_{i3}$, and $K_{i7_2} = \frac{\epsilon_1}{\epsilon_3}(\kappa_{i10} + \kappa_{i13}y_{i2})y_{i7}$, where y_{i2} and y_{i3} are taken to be fixed, and $p_{HDM}p = \frac{e_{HDM}}{Y}$, $p_{HDAC}p = \frac{e_{HDAC}}{Y}$. The corresponding Master Equation can be solved using the same method used for the fast GRN variables. The resulting QSSA PDF generating functions are:

$$G_\infty(p_{i5}) = \left(\frac{\kappa_{i2} + \kappa_{i3} + (\kappa_{i5} + \kappa_{i6})y_{i3} + (\kappa_{i1} + y_{i3})y_{i2}p_{i5}}{(\kappa_{i2} + \kappa_{i3}) + (\kappa_{i1} + y_{i3})y_{i2} + (\kappa_{i5} + \kappa_{i6})y_{i3}} \right)^{e_{HDM}} \tag{A.2.48}$$

$$G_\infty(p_{i7}) = \left(\frac{\kappa_{i10} + \kappa_{i11} + (\kappa_{i13} + \kappa_{i14})y_{i2} + (\kappa_{i9} + \kappa_{i12}y_{i2})y_{i3}p_{i7}}{(\kappa_{i10} + \kappa_{i11}) + (\kappa_{i9} + \kappa_{i12}y_{i2})y_{i3} + (\kappa_{i13} + \kappa_{i14})y_{i2}} \right)^{e_{HDAC}} \tag{A.2.49}$$

where $y_{i2} = Y_{i2}/Z$, $y_{i3} = Y_{i3}/Z$, and the parameters κ_{ij} are defined in Table 3.4.

The number of complexes Y_{i5} and Y_{i7} are sampled from the binomial distributions related to the generating functions Eqs. (A.2.48) and (A.2.49). The corresponding number of free enzymes is then obtained using the conservation laws: $Y_{i4} = e_{HDM} - Y_{i5}$ and $Y_{i6} = e_{HDAC} - Y_{i7}$, with e_{HDM} and e_{HDAC} denoting the total number of HDM and HDAC enzymes, respectively.

A.2.1.5 Numerical method

So far, we have exploited separation of time scales to formulate a QSSA whereby both the number of bound sites within the promoter of the genes and the number of enzymes and complexes molecules associated with the ER enzyme kinetics are sampled from their QSSA PDFs. Furthermore, since $S \gg 1$, we have taken a large-S limit which allows us to write the dynamics of X_i, i.e. the number of protein transcripts of gene i, in terms of an ODE perturbed by two random forcings: one associated with the random (fast) binding/unbinding dynamics, and another one produced by the random ER dynamics. However, as long as we assume that $\epsilon_2 = \mathcal{O}(1)$ and $Y = E \ll S$, further simplification is not possible and the evolution of the slow ER variables (i.e. number of positive and negative marks, and unmarked sites) need to be solved by numerical simulation of their stochastic dynamics. In spite of this, Eqs. (3.2.11)-(3.2.14) and (3.2.17)-(3.2.18) provide the reduced version of the original stochastic model which allows for a far more efficient numerical implementation of the complex ER-GRN stochastic system.

The asymptotic reduction of the full stochastic model provides the basis for a hybrid numerical method with enhanced performance with respect to the stochastic simulation algorithm. The current hybrid method is based on that formulated in [65]. The numerical method proceeds through iteration of a basic algorithm composed of the following steps:

1. Set initial conditions for the slow variables of the GRN and ER components of the system described by Eqs. (3.2.17)-(3.2.18) (see Section 3.2.2).

2. Sample the fast variables from their QSSA PDFs conditioned to the current value of the corresponding slow variables. Their sampled values are fed in the evolution equations of the latter.

3. Consider the stochastic dynamics of the slow ER variables, Eqs. (3.2.7). These stochastic equations must be solved by numerical simulation using Gillespie's SSA. We first set the corresponding time step, $\Delta\tau$, using the SSA.

4. Solve the ODEs for the slow variables of the GRN dynamics in the time interval $[\tau, \tau + \Delta\tau)$.

5. Complete the Gillespie step for the slow ER variables by choosing which elementary reaction alters the ER regulatory state and update the slow ER variables accordingly.

6. Repeat Steps 2 through to 5 until some stopping condition is satisfied.

If we are running the fixed time step version of the algorithm, Step 1 needs to be done only once during the initialisation of the algorithm.

A.2.2 Benchmark: stochastic model of a single self-activating gene

We use the simplest case of a single self-activating gene as benchmark to check the accuracy of the multi-scale analysis introduced in Section A.2.1. This example can be analysed in detail and we use it as an illustrative example which can help to shed some light onto more complex situations (see Fig. A.3 for a schematic representation). The dynamics of this system is described by the Chemical Master Equation (CME) with transition rates W_j given in Table A.11. A slightly modified version of this system has been analysed in de la Cruz et al. [40] using the methods proposed by Alarcón [1]. This methodology is based on optimal path theory (OPT) [21] with separation of time scales. The latter is used to formulate a quasi-steady state approximation (QSSA) of the Euler-Lagrange (Hamilton) equations of the optimal path.

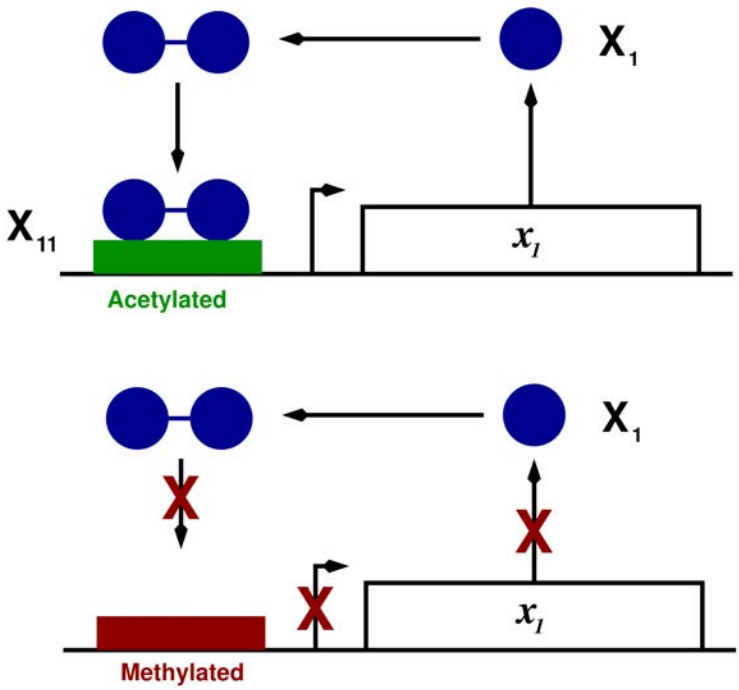


Figure A.3: Schematic representation of the model of self-activation gene regulatory circuit with epigenetic regulation. The gene product X_1 is its own transcription factor which, upon dimerisation, binds the promoter region (if possible, i.e. if acetylated) of the gene thus triggering gene transcription. The transition rates corresponding to this gene regulatory circuit are given in Table A.11. For simplicity, we use an effective model in which the formation of the dimer and binding to the promoter region is taken into account in a single reaction, and the resulting number of promoter sites bound by two transcription factors is denoted by X_{11}. Furthermore, depending on whether the epigenetic state is predominantly acetylated or methylated, the promoter region of the gene is accessible or inaccessible to the transcription factor, respectively.

Table A.10: Random variables associated with the stochastic dynamics of an auto-activation gene regulatory circuit [55, 169]. See Fig. A.3 for a schematic representation.

Variable	Description
X_1	Number of transcription factor molecules
X_{11}	Number of bound promoter sites in the gene promoter region

Table A.11: Transition rates associated with the stochastic dynamics of an auto-activation gene regulatory circuit [55, 169]. e_1 corresponds to the number of binding sites within the gene's promoter region. See Fig. A.3 for an schematic representation. Notation: TF stands for transcription factor and GPR for gene promoter region.

Transition rate	r	Event
$W_1(x) = \hat{R} + k_1 X_{11}$	$r_1 = (1, 0)$	Synthesis of the TF
$W_2(x) = k_2 X_1$	$r_2 = (-1, 0)$	Degradation of the TF
$W_3(x) = b_{11} X_1 (X_1 - 1)(e_1 - X_{11})$	$r_3 = (-2, +1)$	Dimer binding to the GPR
$W_4(x) = u_{11} X_{11}$	$r_4 = (+2, -1)$	Unbinding from the GPR

A.2.2.0.1 Robustness of the active state: assessing the accuracy of the multi-scale OPT predictions.

Using the methodology outlined in Section A.2.1.2, de la Cruz et al. [40] have derived the SCQSSA equations for the stochastic model of a self-activating gene:

$$\frac{dq_1}{d\tau} = R + \omega_1 p_{\infty_1} p \frac{q_1^2}{\delta_{11} + q_1^2} - \omega_2 q_1, \tag{A.2.50}$$

$$q_{11} = p_{\infty_1} p \frac{q_1^2}{\delta_{11} + q_1^2}. \tag{A.2.51}$$

where p and p_{∞_1} are parameters such that $p_{\infty_1} p = \frac{e_1}{E}$ [1, 40, 53]. The rescaled variables and parameters are defined in Table A.9. Fig. A.4 shows how the steady state behaviour of the Eqs. (A.2.50)-(A.2.51) changes as $p_{\infty_1} p$ varies. We set our system up so that its mean-field limit, corresponding to $p_{\infty_1} p = 1$ [46], exhibits bistability. We observe that, as $p_{\infty_1} p$ decreases (i.e. as e_1 decreases), the system goes through a saddle-node bifurcation whereby the active state (i.e. the steady state where $q_1 = O(1)$) loses stability. Therefore, our stochastic SCQSSA predicts that a noise-induced transition from the active state into the inactive state (i.e. the steady state such that $q_1 \ll O(1)$)) will occur, provided that $\frac{e_1}{E}$ is smaller than the critical value of $p_{\infty_1} p$, which we refer to as $(p_{\infty_1} p)_{cr}$. This scenario has been fully explored and verified by means of direct stochastic simulations by de la Cruz et al. [40]. Therefore, the robustness of the active state is quantified by the action integral computed on such trajectory [46, 21].

It is straightforward to verify the action integral Eq. (A.2.9) computed on the outer solution, A_{QSS}, satisfies that $A_{QSS} = 0$. This is because $H_\omega(p, q) = 0$, $p_1(\tau) =$

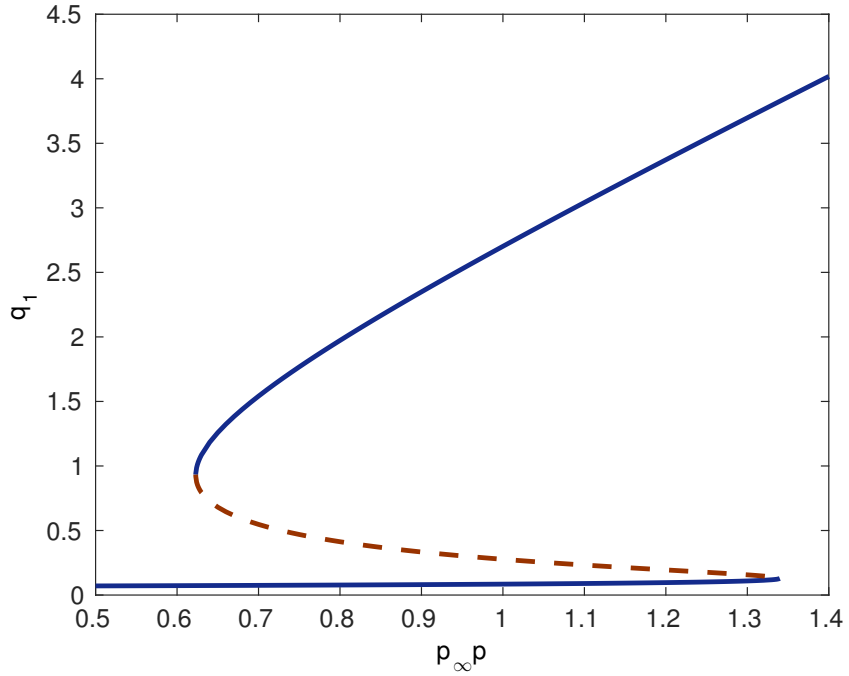


Figure A.4: Bifurcation analysis for the SCQSS approximation of the stochastic auto-activation gene regulatory circuit, Eq. (A.2.50), with mean-field transcription rate $\omega_1 = 3.0$. Parameter values as given in Table A.12. See de la Cruz et al. [40] for details.

1, and $\epsilon \dot{p}_{11} \simeq 0$. Therefore, in order to quantify the robustness of the system, we must resort to the analysis of the inner regime. The result of this analysis, which is carried out in detail in the Section A.2.1.2, see Eq. (A.2.42), is that there exists a solution of the inner dynamics with a positive contribution to the action, $\mathcal{A}_{in}$:

$$\mathcal{A}_{in} = \frac{e_1}{S} \log \frac{q_{11_{out}}}{q_{0_{11}}} \tag{A.2.52}$$

where

$$q_{11_{out}} = p_{\infty_1} p \frac{q_{0_1}^2}{\delta_{11} + q_{0_1}^2}.$$

We have further used that $\epsilon = E/S$, $p_{\infty_1} p = e_1/E$. This quantity and $q_{0_{11}} = q_{11}(T = 0)$ must be such that $H_\omega(p_{0_1}, p_{0_{11}}, q_{0_1}, q_{0_{11}}) = 0$. Moreover, since $p_{0_1} q_{0_1} = \frac{X_1(t=0)}{S}$ and $p_{0_{11}} q_{0_{11}} = \frac{X_{11}(t=0)}{E}$, besides the energy condition, $p_{0_1} q_{0_1} \geq 0$ and $0 \leq p_{0_{11}} q_{0_{11}} \leq \frac{e_1}{E}$ must hold. Furthermore, q_{0_1} must belong to the basin attraction of the active steady state of Eqs. (A.2.50)-(A.2.51).

The estimate for the robustness of the active state in this simple gene regulatory network within our multi-scale optimal path theory is therefore given by [170]:

$$A_0 = \begin{cases} \frac{e_1}{S} \log \left(\frac{q_{11_{out}}}{q_{0_{11}}} \right) & \text{if } \frac{e_1}{E} < (p_{\infty_1} p)_{cr} \\ \mathcal{O}(1) & \text{if } \frac{e_1}{E} \geq (p_{\infty_1} p)_{cr} \end{cases} \tag{A.2.53}$$

Eqs. (A.2.50) and (A.2.53) allow for several numerically testable predictions. Below the critical value, i.e. for $\frac{e_1}{E} < (p_{\infty_1} p)_{cr}$, we have that:

$$\tau_E \sim \exp \left(e_1 \log \left(\frac{q_{11_{out}}}{q_{0_{11}}} \right) \right),$$

since $\tau_E \sim e^{S A_0}$, whereas, if $\frac{e_1}{E} \geq (p_{\infty_1} p)_{cr}$

$$\tau_E \sim \mathcal{O}\left(e^S\right),$$

which means that the below-critical escape time from the active state grows exponentially with the number of binding sites in the promoter region of the gene. This prediction of our multi-scale optimal path theory analysis has been validated by direct stochastic simulations of the stochastic model of the self-activating gene regulatory circuit, Table A.11 [59, 60], which are shown in Fig. A.5.

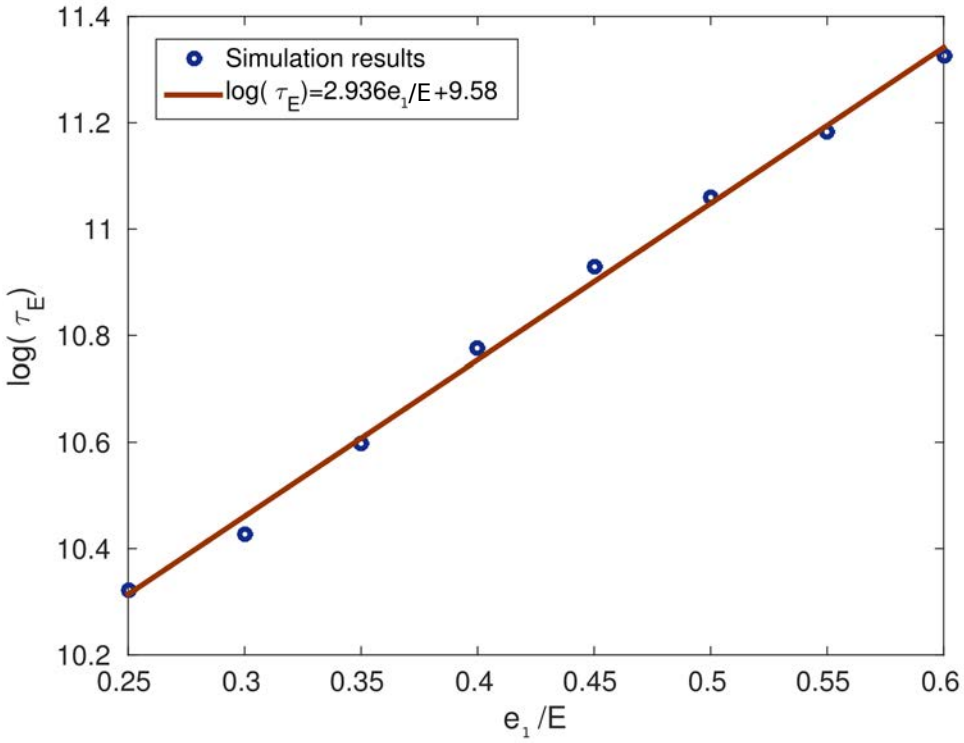


Figure A.5: This plot shows simulation results regarding the average exit time from the active steady state. We have plotted how the (logarithm of the) average exit time, τ_E, changes as e_1, i.e. the total number of binding sites in the promoter region of the self-activating gene, is let to vary. We show our results as a log-linear plot. This figure validates our result that the below critical exit time varies exponentially with e_1. Blue circles represent simulation results. The red line represents a least-squares fitting of the simulated data. The coefficients of this fitting are within the 95% confidence interval. Parameter values for the stochastic simulations have been taken from Tables A.12 and A.9. Furthermore $\omega_1 = 3.0$. Averages are performed over 1000 realisations for each data point.

Table A.12: Parameter values used in simulations of the stochastic self-activating single gene regulatory circuit. See de la Cruz et al. [40] for details.

Rescaled parameter	Parameter	Units	Reference
$\omega_1 = \frac{a}{k_{deg}\sqrt{K_d}}$	$K_d = 10$	nM	[55]
$\omega_2 = 1$	$k_{deg} = 2$	min^{-1}	[55]
$\delta_{11} = 1$	$r = 0.4$	$\text{nM} \cdot \text{min}^{-1}$	[55]
$R = \frac{r}{k_{deg}\sqrt{K_d}}$	$S = 1000$		–
$b_{11}ES = k_{deg}$	$E = 20$		–

A.2.2.0.2 Model reduction: assessing the accuracy of the asymptotic analysis and the numerical method.

Following the procedure described in Section 3.2.2, we can obtain Eq. (3.2.17), for the stochastic model of a single self-activating gene, which reads as follows,

$$\frac{dx_1}{d\tau} = R_1 + \omega_1 x_{11} - \omega_2 x_1 - 2\left(\eta_i\left(\frac{e_1}{E} - x_{11}\right)x_1^2 - \delta_{11}x_{11}\right), \tag{A.2.54}$$

where $x_{11} = \frac{X_{11}}{E}$ and X_{11} is a sampled from:

$$P_1(X_{11}|x_1) = H(Y_{i3} - Y_0)P_{1_+}(X_{11}|x_1) + H(Y_0 - Y_{i3})P_{1_-}(X_{11}|x_1), \tag{A.2.55}$$

where $P_{1_+}(X_{11}|x_1)$ is a binomial distribution whose generating function is $G_\infty(p_{11}|x_1) = (1 + \pi_{11}(x_1)(p_{11} - 1))^{e_1}$ where

$$\pi_{11}(x_1) = \frac{x_1^2}{\delta_{11} + x_1^2},$$

and $P_{1_-}(X_{11}|x_1) = \delta_{X_{11},0}$. We have used this simple model to test the accuracy of the asymptotic model reduction and the associated numerical method. The results are shown in Fig. A.6 where we compare the performance of our approximation against stochastic simulation algorithm results. SSA is feasible for this simple model which allows a systematic comparison with our approximation. For more complex systems, SSA is computationally too costly, and an alternative is necessary, which is the motivation behind our asymptotic analysis.

To carry out the comparison between the asymptotic reduction and the SSA, we have simulated a (very long) realisation where we let the system to evolve towards

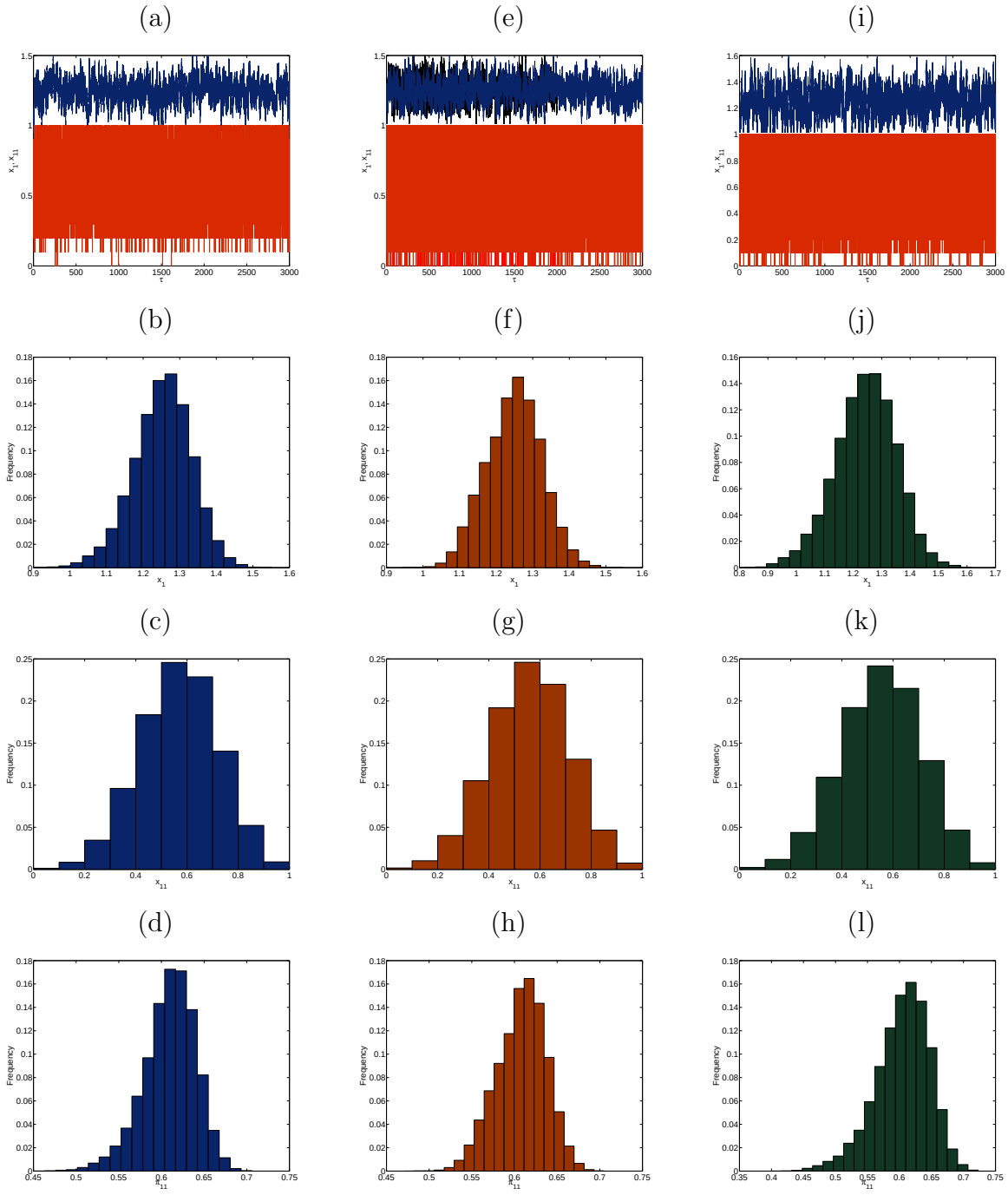


Figure A.6: Caption on the following page.

Figure A.6: (Previous page). Results for simulations of the aysmptotic model reduction for the stochastic model of the self-activating gene with fixed $\Delta\tau$. Series of plots comparing a number of statistics regarding the stochastic simulation of the self-activating GRN using Gillespie's stochastic algorithm (plots (a), (b), (c), and (d)) and the asymptotic model reduction (plots (e), (f), (g), (h) with fixed time step $\Delta\tau = 0.25\epsilon = 2.5 \cdot 10^{-3}$, and plots (i), (j), (k), (l) with fixed time step $\Delta\tau = 0.5\epsilon = 5 \cdot 10^{-3}$). The average time step in the Gillespie simulation is $\Delta\tau = 2.6946 \cdot 10^{-4}$. The mean and standard deviation of x_1 in the Gillespie simulation are $\langle x_1 \rangle = 1.2532$ and $\sigma_{x_1} = 0.0794$, respectively. According to simulation of the asymptotic model reduction these quantities are $\langle x_1 \rangle = 1.2488$ and $\sigma_{x_1} = 0.0752$, $\langle x_1 \rangle = 1.2431$ and $\sigma_{x_1} = 0.1072$ for $\Delta\tau = 0.25\epsilon$ and $\Delta\tau = 0.5\epsilon$, respectively. Parameter values given in Table A.12 and $a = 1.95$.

the open gene steady-state (Figs. A.6(a), (e), and (i)). Over this long realisation, we have collected statistics regarding three quantities: x_1, x_{11}, and π_{11}. The asymptotic reduction fares rather well against the SSA when we take $\Delta\tau$ to be up to an order of magnitude larger than the average time step in the SSA simulation. This is a rather substantial improvement in computational performance. Taking larger values of $\Delta\tau$, in our case $\Delta\tau > 0.5\epsilon$, with $\epsilon = 10^{-2}$, distorts the statistics and the results of the asymptotic reduction are no longer accurate.

We have tried the $\tau-$leap version of the algorithm, for which we do not report the results because there was no improvement in performance when compared to the fixed time step method. The $\Delta\tau$ selection method used in our simulations is the one proposed in [61]. The issue with these simulations is that occasionally the $\Delta\tau$ selection procedure produces a value of the time step that is exceedingly large. This compromises the stability of the method and its accuracy. Further work, beyond the scope of this thesis, needs to be done regarding adaptive time step versions of the asymptotic reduction method.

A.2.3 Summary of the minimum action path theory and its numerical implementation

The Freidlin-Wentzell theory of large deviations provides a framework to understand the effects of small random perturbations in dynamical systems. Following large deviation principles, statistics of rare events, when they occur, will follow the least unlikely path with high probability. The key element of this theory is an action

functional whose minimiser gives the stochastic trajectory of maximum likelihood by which this event happens [73]. Therefore, our problem of finding a stochastic trajectory is translated into finding the trajectory that minimizes a deterministic functional. The minimization is done by a custom made genetic algorithm that explores different paths in the $n+1$ dimensional space, where one dimension corresponds to the parameter which parameterizes the path and the remaining n dimensions correspond to the number of chemical species involved in the model.

The action minimisation was performed by finding the path

$$\varphi(s) = (y_{i1}(s), \ldots, y_{i7}(s)) : s = [0,1] \to \mathbb{R}^7,$$

with y_{ij} described in Table 3.1 that minimises the time independent gMAM functional [73] given an initial and final point, $\varphi(0)$ and $\varphi(1)$, respectively:

$$\mathcal{S}(\varphi(s)) = \int_0^1 |f(s)|_a \left(1 - \cos\eta_a(s)\right) |\hat{n}(s)|_a^2 \mathrm{d}s \tag{A.2.56}$$

where f is the deterministic component of the Langevin equation and $\hat{n}$ is a unit vector tangent to φ at a path position s (see Eq. (3.2.20)). The position dependent diffusion matrix $a(s) = g^\intercal\left(\varphi(s)\right) g\left(\varphi(s)\right)$ is used to compute the inner product that defines the norm $|\cdot|_a = \sqrt{\langle \cdot, a^{-1} \cdot \rangle}$ (see Eq. (3.2.20)). The angle $\eta_a(s)$ is the angle between f and φ at a path position s, which is defined using the same inner product, $\cos\eta_a = \frac{\langle f, a^{-1}\hat{n}\rangle}{|f|_a |\hat{n}|_a}$.

For notation convenience, we drop the i subscript of y_{ij}, i.e. we denote the variables by y_j. In order to find the minimum action path, the candidate paths where parametrized by $n = 11$ points along the path $\tilde{\varphi}(s) = (y_1^i(s^i), \ldots, y_7^i(s^i))$, evaluated along a set of ordered values $\{s^i \in [0,1]\}$ with $i = 1, ..., n$; $s^1 = 0$ and $s^n = 1$. Thus, each path is determined by a $8n$-dimensional vector. To evaluate the gMAM functional (A.2.56), each path is interpolated using Akima splines along $N = 200$ points. The minimisation was performed using a custom made evolutionary algorithm on a population of $d_{pop} = 10000$ candidate curves. Each generation of the algorithm keeps the top $d_{fit} = 2000$ curves that score the lowest action, whilst the rest of the curves are replaced by offspring curves resulting from random pairings from the fittest individuals. For each random parent pair from the d_{fit} best candidates, namely, $(\tilde{\varphi}', \tilde{\varphi}'')$, its offspring path, $\tilde{\varphi}'''$, is calculated by taking uniform random intermediate values between each of the $n-2$ intermediate points of its parents pair and introducing mutations by randomly perturbing the resulting curve (the first and last point, y_j^1 and y_j^n, are not modified since they correspond with the stable

stationary points (see Eq. (3.2.20)). Namely each coordinate $y_j^{'''i}$ of an offspring is calculated as

$$y''^i_j = ((1-\rho_1)y''^i_j + y'''^i_j\rho_1)\mathrm{e}^{m\rho_2}, \tag{A.2.57}$$

where ρ_1 is a uniform number $U(0,1)$, ρ_2 is a Gaussian random number $N(0,1)$, and $m = 0.01$ is the mutation intensity. The parametrisation vector $\{s^{'''i}\}$ is calculated similarly while imposing that it is monotonically increasing. In order to initialise the population, the first d_{pop} trajectories are generated by sampling random points in the latin hypercube defined by $\varphi(0)$ and $\varphi(1)$ and sampling ordered uniform vectors $\{s^i\}$. In order to keep track of the convergence of the algorithm, a Limited-memory BFGS quasi-Newton method is used to find a refined local minimum on the best candidate path every $g_{grad} = 20$ generations. The simulations were accelerated by introducing resulting optimal trajectories from the L-BFGS as a new offspring curve for the next generation. The algorithm is stopped when the curves approach the same local minimum over $g_{max} = 400$ trajectories or when the slope of the loglog curve of the average action of the d_{fit} best paths, as a function of the number of generations, is small enough ($\alpha_p = 0.01$).

In order to validate the results, the algorithm was run several times taking as final result the best curve candidate. Additionally, for some candidates, different values of the parameters of the algorithm ($d_{pop}, d_{fit}, n, g_{grad}, g_{max}m, N,$ and α_p) were explored to ensure that the algorithm was not stuck in local minima.

As a summary, the minimum action path (MAP) method follows these steps:

1. **Compute stationary states of the system**: Compute the stationary states of the system for the given parameters (saddle point, y_s, and the initial and final stable states, y_{ini} and y_f).

2. **Generate initial population of paths.** Generate d_{pop} random paths that start at y_{ini} and end at y_f.

3. **Start optimization loop** Set counter $i = 1$ (number of generations of the genetic algorithm).

   Compute the fitness Compute the value of the action functional S along each path φ (individual) from the population. In order to do so, the vector force f (mean-field drift vector), the gradient of the force, and the

inverse of the diffusion tensor a (a^{-1}), with $a = gg^t$ and g the noise matrix, need to be evaluated at each point of the path.

Rank the paths according to their fitness Rank them from the lowest to the highest value of the action functional S.

Keep the best paths Select the first d_{fit} paths, i.e, those d_fit paths with lower value of S. This subgroup of d_{fit} paths will be carried to the next generation.

Find optimal gradient path If $i \equiv 0 \pmod{g_{grad}}$, compute a L-BFGS minimisation taking the minimum action path of the population as the initial condition. If the resulting path improves, add it to the d_{fit} subpopulation.

Check convergence If the converging conditions are met, exit the loop and take the best path as the optimum result.

Generate new paths Generate new paths by mixing and mutating the best d_{fit} paths from the generation. This is done in order to get a new population with the same number of individuals as the original one (d_{pop} paths).

Update generation counter i = i+1

A.2.4 Kolmogorov Smirnov test: analysis of ensemble heterogeneity

In this section, we provide the p-values of those differences found to be significant in the Kolmogorov-Smirnov (K-S) tests carried out in Section 3.3.3 (see details there).

A.2.4.1 Significant differences within the ensemble of DERSs

Red cluster versus blue cluster. Our analysis shows that the KS test detects significant differences for c_{1_1} (p-value 0.0349, unrecruited demethylation), $c_{1_{11}}$ (p-value 0.0062, unrecruited deacetylation), $c_{1_{15}}$ (p-value 0.0010, unrecruited acetylation), and $c_{1_{16}}$ (p-value $3.9433{\cdot}10^{-4}$, recruited acetylation).

Red cluster versus green cluster. In this case, our analysis shows that significative differences are found only for two parameter values: c_{1_3} (p-value $2.5697{\cdot}10^{-4}$, unrecruited demethylation) and $c_{1_{16}}$ (p-value 0.0307, recruited acetylation).

Blue cluster versus green cluster. Significative differences are found for the empirical distributions of c_{1_3} (p-value $2.1411{\cdot}10^{-4}$, unrecruited demethylation), c_{1_8} (p-value

0.0038, recruited methylation), $c_{1_{11}}$ (p-value 0.0156, unrecruited deacetylation), and $c_{1_{15}}$ (p-value 0.0023, unrecruited acetylation).

A.2.4.2 Significant differences between differentiation-primed and pluripotency-locked ER landscapes

We first proceed to compare within the whole population (without discriminating between clusters) those DERSs such that $\mathcal{Q} \geq \mathcal{T}$ (differentiation-primed ER landscapes) against those with $\mathcal{Q} < \mathcal{T}$ (pluripotency-locked ER landscapes). We take $\mathcal{T} = 0.7$. The parameter values for which the KS test yields significant differences when comparing the differentiation-primed ER system to the pluripotency-locked ER are c_{1_1} (p-value 0.0125, unrecruited demethylation), $c_{1_{14}}$ (p-value 0.0435, recruited deacetylation), $c_{1_{15}}$ (p-value 0.0191, unrecruited acetylation), and $c_{1_{16}}$ (p-value 0.0057, recruited acetylation).

If we now restrict our analysis to those DERSs within the blue cluster, we observe that the parameters whose CDFs differ significantly when splitted into differentiation-primed and pluripotency-locked are c_{1_1} (p-value 0.0135, unrecruited demethylation) and $c_{1_{14}}$ (p-value 0.0095, recruited deacetylation).

Regarding the PERSs, the results are shown in Fig. A.10. Our analysis shows that significative differences can be found between the empirical distributions of three parameter values: c_{1_3} (p-value $5.9983{\cdot}10^{-6}$, unrecruited demethylation), c_{1_8} (p-value $7.4181{\cdot}10^{-4}$, recruited methylation), and $c_{1_{15}}$ (p-value 0.0047, unrecruited acetylation).

A.2.5 Reference parameter values

In this appendix we give the parameter values used in the simulations reported in Chapter 3. The parameter sets Tables A.15-A.19 are taken from the pluripotency and differentiation ER ensembles generated using the method described in Section 3.2.4 and in reference [53]. Parameter sets from Tables A.13 and A.14 are the ones used to generate the raw data from Figure 3.2.

A.2.6 Supplementary figures

Table A.13: Reference parameter values used to generate the ensemble of the differentiation epigenetic regulatory systems (see Fig. 3.2(a)).

Rescaled parameter	Value (dimensionless)
κ_1	200
κ_2	100
κ_3	50
κ_5	1
κ_6	200
κ_7	10
κ_8	100
κ_9	200
κ_{10}	100
κ_{11}	0.1
κ_{12}	1
κ_{13}	1
κ_{14}	200
κ_{15}	100
κ_{16}	100

Table A.14: Reference parameter values used to generate the ensemble of the pluripotency epigenetic regulatory systems (see Fig. 3.2(b)).

Rescaled parameter	Value (dimensionless)
κ_1	200
κ_2	100
κ_3	10
κ_5	1
κ_6	10
κ_7	100
κ_8	100
κ_9	200
κ_{10}	100
κ_{11}	10
κ_{12}	1
κ_{13}	1
κ_{14}	100
κ_{15}	100
κ_{16}	100

Table A.15: Dimensionless parameters corresponding to the pluripotency epigenetic-regulatory system PERS1 (see Figs. 3.4 and 3.9, and Fig. A.11 and A.8).

Rescaled parameter	Value (dimensionless)
κ_1	121.624100
κ_2	14.942093
κ_3	90.265221
κ_5	135.026962
κ_6	137.927872
κ_7	1652.600464
κ_8	2235.608398
κ_9	123.194839
κ_{10}	47.182499
κ_{11}	32.408936
κ_{12}	34383.695312
κ_{13}	102.787231
κ_{14}	157.155838
κ_{15}	5133.064453
κ_{16}	6294.414551

Table A.16: Dimensionless parameters corresponding to the pluripotency epigenetic-regulatory system PERS2 (see Figs. 3.4 and 3.9, and Fig. A.11 and A.8).

Rescaled parameter	Value (dimensionless)
κ_1	107.745895
κ_2	14.304025
κ_3	52.221851
κ_5	172.351425
κ_6	173.544327
κ_7	1692.524658
κ_8	6550.556641
κ_9	47.068409
κ_{10}	43.554352
κ_{11}	41.708683
κ_{12}	949.391724
κ_{13}	31.121500
κ_{14}	50.545132
κ_{15}	5126.510254
κ_{16}	5547.370605

Table A.17: Dimensionless parameters corresponding to the differentiation epigenetic-regulatory system DERS1 (see Figs. 3.4 and 3.9, and Fig. A.11 and A.8).

Rescaled parameter	Value (dimensionless)
κ_1	40.951008
κ_2	33.545677
κ_3	36.101627
κ_5	71.377869
κ_6	45.063740
κ_7	2806.709717
κ_8	7397.834961
κ_9	170.622452
κ_{10}	77.820129
κ_{11}	0.093411
κ_{12}	6848.829102
κ_{13}	131.147171
κ_{14}	9.915549
κ_{15}	3712.249268
κ_{16}	2623.266846

Table A.18: Dimensionless parameters corresponding to the differentiation epigenetic-regulatory system DERS2 (see Figs. 3.4 and 3.9, and Fig. A.11 and A.8).

Rescaled parameter	Value (dimensionless)
κ_1	21.022829
κ_2	37.389736
κ_3	6.075892
κ_5	119.379074
κ_6	199.229645
κ_7	2635.588623
κ_8	4078.342285
κ_9	142.806244
κ_{10}	50.091576
κ_{11}	0.435084
κ_{12}	6921.437988
κ_{13}	11.198260
κ_{14}	8.374920
κ_{15}	1926.709595
κ_{16}	7136.016602

Table A.19: Dimensionless parameters corresponding to the differentiation epigenetic-regulatory system used in the simulations corresponding to reprogramming of pluripotency-locked DERSs (see Fig. 3.8).

Rescaled parameter	Value (dimensionless)
κ_1	70.205910
κ_2	73.419846
κ_3	3.609020
κ_5	119.968636
κ_6	107.726303
κ_7	5100.981445
κ_8	4037.399658
κ_9	40.685452
κ_{10}	78.823647
κ_{11}	0.785620
κ_{12}	43095.195312
κ_{13}	55.197235
κ_{14}	17.289341
κ_{15}	5054.711914
κ_{16}	7180.712402

Table A.20: Parameter values used in simulations of the stochastic two-gene regulatory circuit with self-activation and competitive inhibition. See de la Cruz et al. [40] for details.

Rescaled parameter	Parameter	Units	Reference
$\omega_{11} = \omega_{21} = \frac{a}{k_{deg}\sqrt{K_d}}$	$K_d = 10$	nM	[55]
$\omega_{21} = \omega_{22} = 1$	$k_{deg} = 2$	min^{-1}	[55]
$\delta_{ij} = 1$ for all i, j	$r = 0.4$	$\text{nM} \cdot \text{min}^{-1}$	[55]
$\beta_{ij} = 1$ for all i, j			–
$R_1 = R_2 = \frac{r}{k_{deg}\sqrt{K_d}}$	$S = 1000$		–
$b_{11}ES = k_{deg}$	$E = 5$		–
	$e_1 = 5$		–
	$e_2 = 5$		–

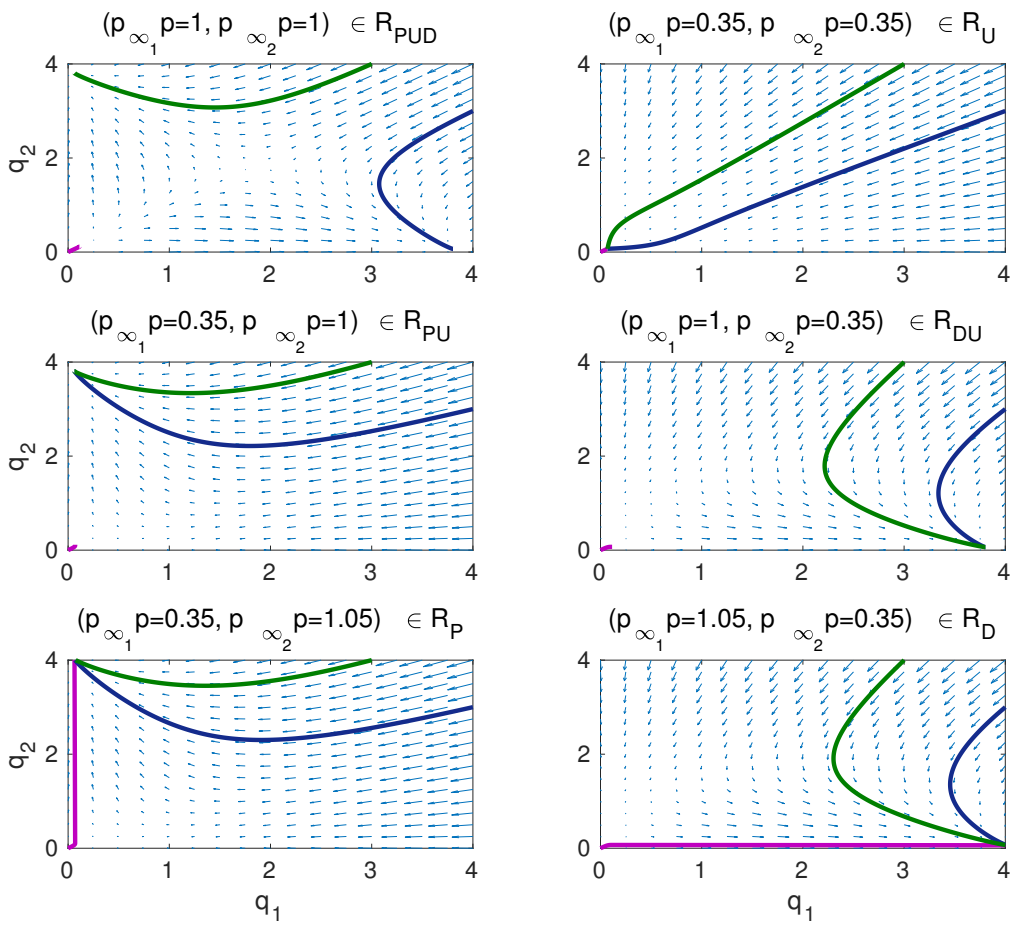


Figure A.7: These plots illustrate the dynamics of Eqs. (3.3.1)-(3.3.2) within each of the regions shown in Figure 3.3. Each of the pannels in this plot shows trajectories for values of the pair $(p_{\infty_1}p, p_{\infty_2}p)$ corresponding to the different regions shown in Figure 3.3. For each $(p_{\infty_1}p, p_{\infty_2}p)$ we plot three trajactories associated with three different initial conditions: $(q_1(\tau = 0), q_2(\tau = 0)) = (0, 0)$ for the purple trajectories, $(q_1(\tau = 0), q_2(\tau = 0)) = (3, 4)$ for the green trajectories, and $(q_1(\tau = 0), q_2(\tau = 0)) = (4, 3)$ for the blue trajectories. Blue arrows indicate the direction and magnitude of the vector field associated with the dynamical system defined by Eqs. (3.3.1)-(3.3.2). Parameters values: $\omega_{11} = \omega_{21} = 4.0$. Other parameter values as per Table A.20.

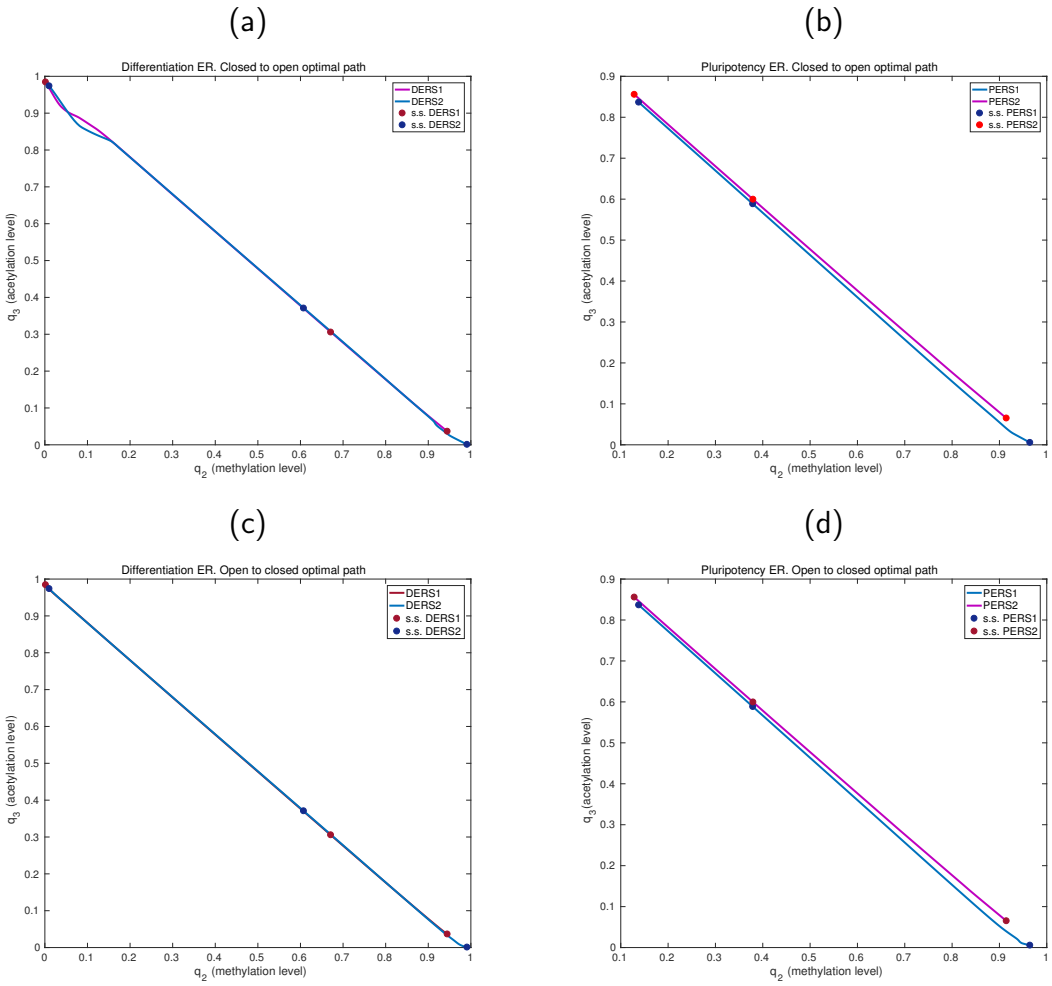


Figure A.8: Plots (a)-(d) represent the optimal transition paths for selected members of the DERS and PERS ensembles. These paths are obtained using MAP theory (see Section 3.2.3 and Section A.2.3). Parameter values are given in Section A.2.5, Tables A.15-A.18.

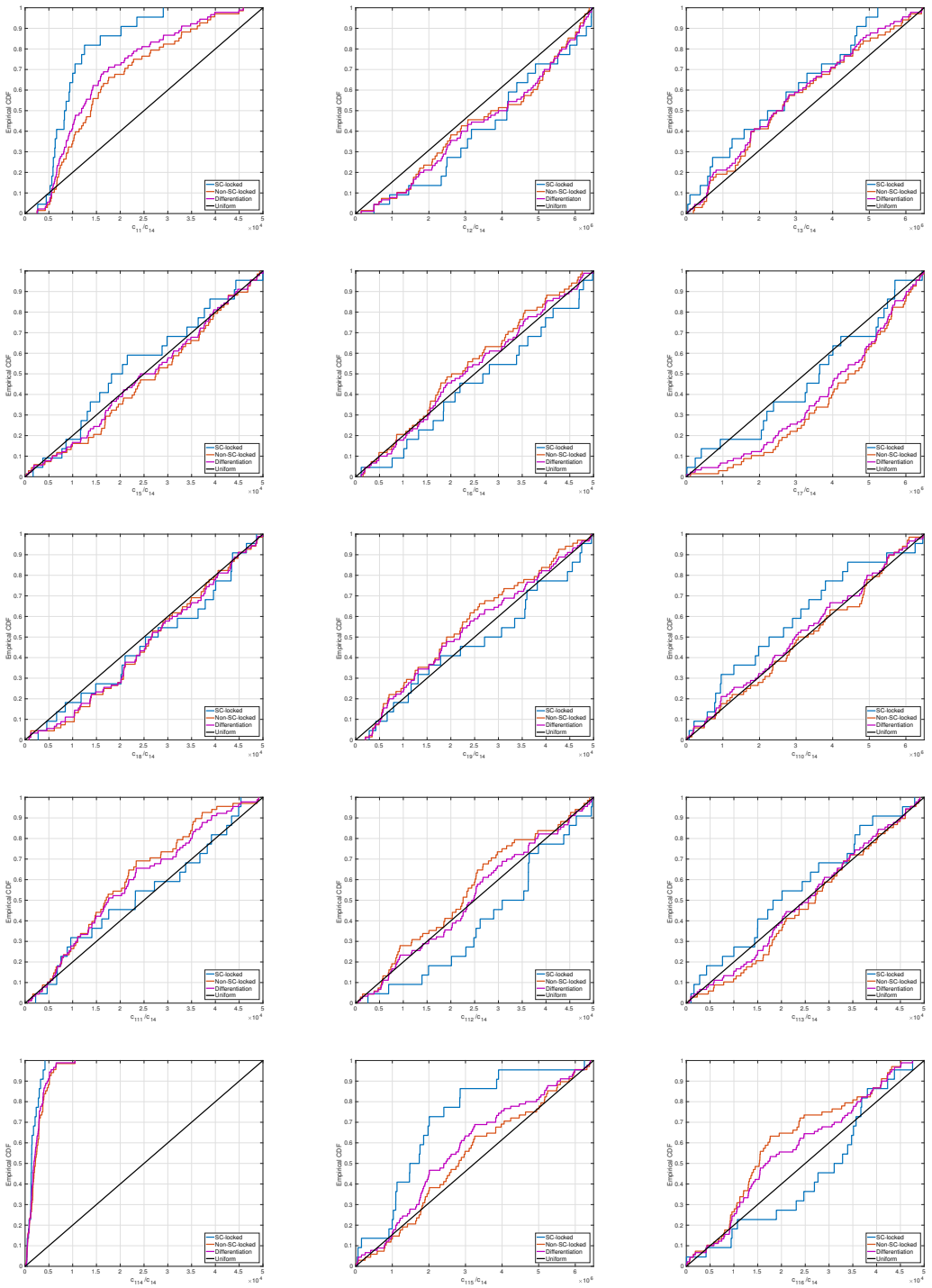


Figure A.9: Caption on a following page.

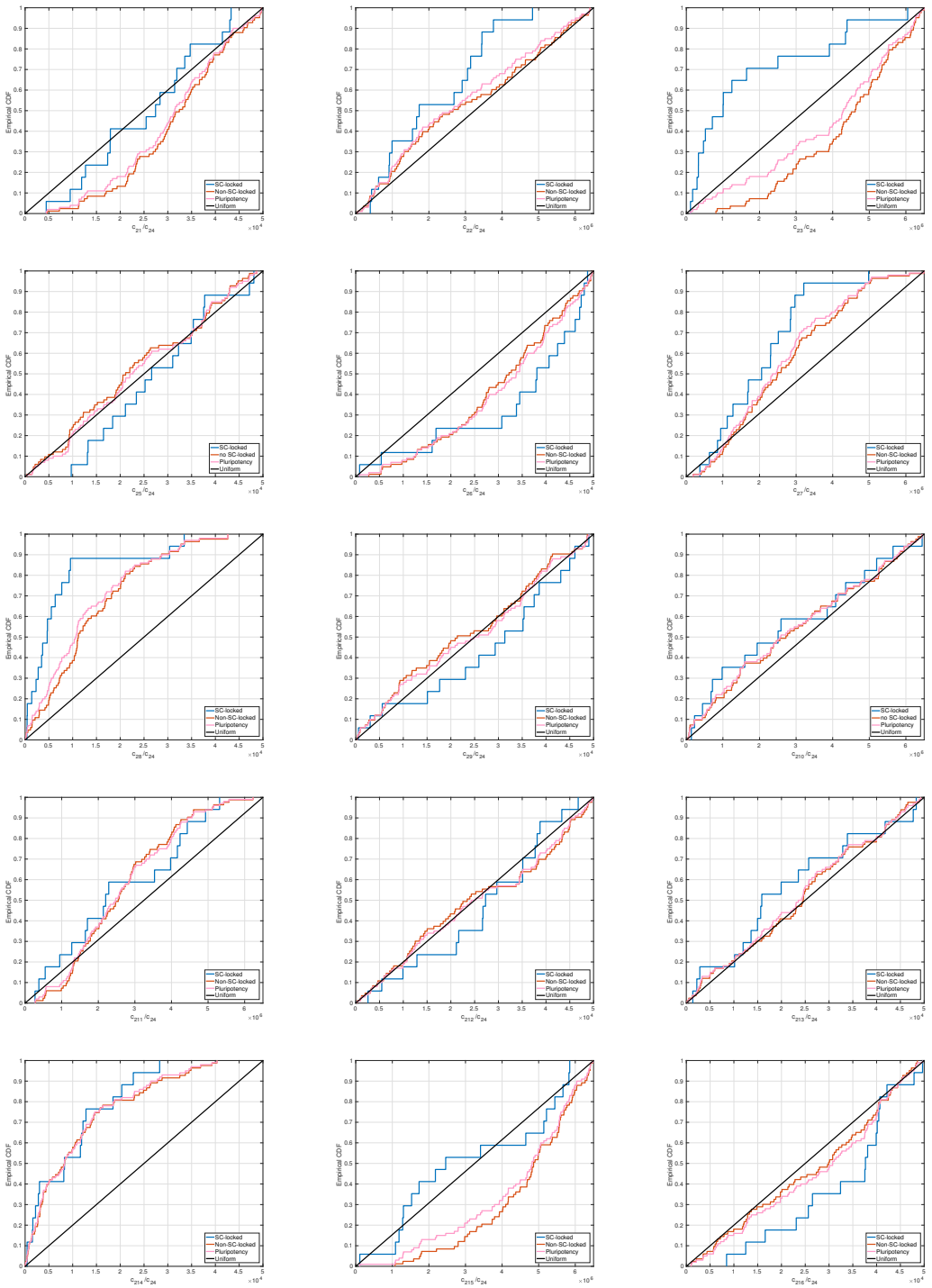


Figure A.10: Caption on the following page.

Figure A.9: (Previous page). This figure shows results regarding the empirical CDFs for the (whole) ensemble of DERS parameter sets. This ensemble has been generated according to the methodology explained in Section 3.2.4 (see also Section 2.2.3.2). For guidance, the CDF corresponding to a uniformly distributed random variable is added (black line). By SC-locked (blue line) we denote those parameter sets where $\mathcal{Q} < 0.7$, whereas those sets with $\mathcal{Q} \leq 0.7$ are referred as non-SC locked (orange lines). See Section 3.3.2 and Section 3.3.3 for further details.

Figure A.10: (Previous page). This figure shows results regarding the empirical CDFs for the ensemble of PERS parameter sets. This ensemble has been generated according to the methodology explained in Section 3.2.4 (see also Section 2.2.3.2). For guidance, the CDF corresponding to a uniformly distributed random variable is added (black line). Different sets are defined following the same criteria as in Fig. A.9. For further details, see Section 3.3.2 and 3.3.3.

A.3 Supplementary materials Chapter 4

In this Section, we provide supplemental information for the results obtained in Section 4.3.1. Precisely, we show the parameter values used to recover the plastic and scenario scenarios described in 2.3.1, for the model presented in Chapter 4.

A.3.1 Reference parameter values: resilient and plastic scenarios

Here we give the sets of parameter values used to produce the phase diagrams associated both with the the phenotype with elevated plastic potential (Fig. 4.1) and the reprogramming-resilient phenotype (Fig. 4.2). These sets of parameter values have been chosen with the same viability criteria as those given in Section 2.2.3.1 (Chapter 2), namely, that the mean-field limit has a single stable steady state corresponding to the open (closed) epigenetic state for the differentiation (pluripotency) gene. Besides this, we further require that with $\varepsilon = 0$ (abundant marks), there is no overlap of the bistability regions for the reprogramming-resilient phenotype, whereas for the plastic phenotype we require that the solid red line is at the left of the dashed blue line (see Fig. 4.1(b)). The reference parameter values are given in Tables A.21 (resilient scenario) and A.22 (plastic scenario).

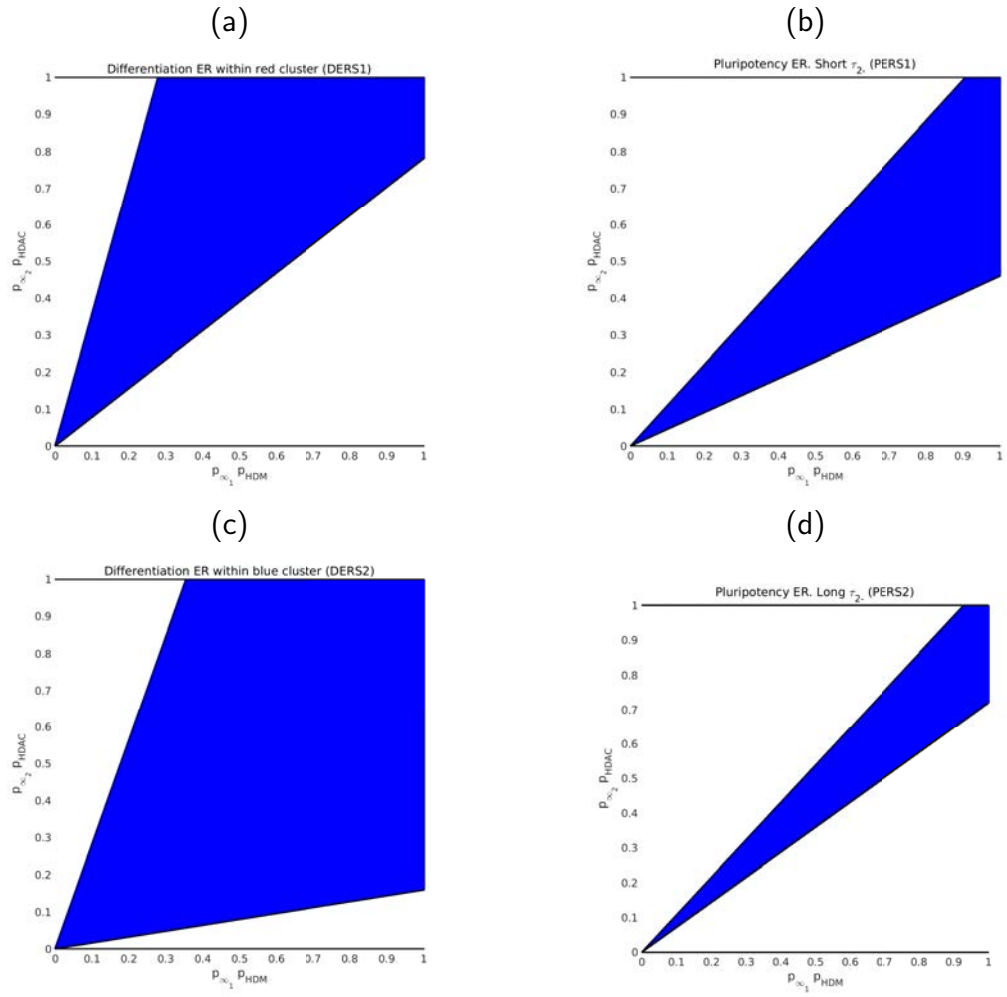


Figure A.11: The blue region correspond to the bistable phase, where both the open and silenced ER states are stable. The region above (at the left of) the bistability tongue is a monostable phase where only the silenced ER state is stable, which we designate as the *closed phase*. Correspondingly, the region below (at the right) is a monostable phase where only the open ER state is stable, which we refer to as the *open phase*. The points at the boundary of the bistability region are instability thresholds where saddle-node bifurcations occur. This phase diagram was obtained using the methodology formulated in [71] and using Bertini (see Appendix B and [13]). Parameter values are given in Section A.2.5, Tables A.15-A.18.

Rescaled parameter	Rescaled parameter	Units	Reference
$\kappa_1 = 200$	$\kappa_{25} = 200$	dimensionless	–
$\kappa_2 = 100$	$\kappa_{26} = 100$	dimensionless	–
$\kappa_3 = 30$	$\kappa_{27} = 10$	dimensionless	–
$\kappa_4 = 1$	$\kappa_{28} = 1$	dimensionless	–
$\kappa_5 = 1$	$\kappa_{29} = 1$	dimensionless	–
$\kappa_6 = 100$	$\kappa_{30} = 50$	dimensionless	–
$\kappa_7 = 200$	$\kappa_{31} = 200$	dimensionless	–
$\kappa_8 = 100$	$\kappa_{32} = 100$	dimensionless	–
$\kappa_9 = 0.1$	$\kappa_{33} = 100$	dimensionless	–
$\kappa_{10} = 1$	$\kappa_{34} = 200$	dimensionless	–
$\kappa_{11} = 1$	$\kappa_{35} = 100$	dimensionless	–
$\kappa_{12} = 200$	$\kappa_{36} = 100$	dimensionless	–
$\kappa_{13} = 160$	$\kappa_{37} = 200$	dimensionless	–
$\kappa_{14} = 90$	$\kappa_{38} = 100$	dimensionless	–
$\kappa_{15} = 90$	$\kappa_{39} = 8$	dimensionless	–
$\kappa_{16} = 1$	$\kappa_{40} = 1$	dimensionless	–
$\kappa_{17} = 100$	$\kappa_{41} = 1$	dimensionless	–
$\kappa_{18} = 100$	$\kappa_{42} = 100$	dimensionless	–
$\kappa_{19} = 160$	$\kappa_{43} = 200$	dimensionless	–
$\kappa_{20} = 80$	$\kappa_{44} = 100$	dimensionless	–
$\kappa_{21} = 80$	$\kappa_{45} = 100$	dimensionless	–
$\kappa_{22} = 140$	$\kappa_{46} = 180$	dimensionless	–
$\kappa_{23} = 70$	$\kappa_{47} = 90$	dimensionless	–
$\kappa_{24} = 70$	$\kappa_{48} = 90$	dimensionless	–

Table A.21: Reference parameter values used for the epigenetic regulation model giving rise to resilient phenotype (Fig. 4.2)

Rescaled parameter	Rescaled parameter	Units	Reference
$\kappa_1 = 200$	$\kappa_{25} = 200$	dimensionless	–
$\kappa_2 = 100$	$\kappa_{26} = 100$	dimensionless	–
$\kappa_3 = 30$	$\kappa_{27} = 10$	dimensionless	–
$\kappa_4 = 1$	$\kappa_{28} = 1$	dimensionless	–
$\kappa_5 = 1$	$\kappa_{29} = 1$	dimensionless	–
$\kappa_6 = 100$	$\kappa_{30} = 50$	dimensionless	–
$\kappa_7 = 200$	$\kappa_{31} = 200$	dimensionless	–
$\kappa_8 = 100$	$\kappa_{32} = 100$	dimensionless	–
$\kappa_9 = 0.1$	$\kappa_{33} = 100$	dimensionless	–
$\kappa_{10} = 1$	$\kappa_{34} = 200$	dimensionless	–
$\kappa_{11} = 1$	$\kappa_{35} = 100$	dimensionless	–
$\kappa_{12} = 200$	$\kappa_{36} = 100$	dimensionless	–
$\kappa_{13} = 185$	$\kappa_{37} = 200$	dimensionless	–
$\kappa_{14} = 90$	$\kappa_{38} = 100$	dimensionless	–
$\kappa_{15} = 90$	$\kappa_{39} = 8$	dimensionless	–
$\kappa_{16} = 1$	$\kappa_{40} = 1$	dimensionless	–
$\kappa_{17} = 100$	$\kappa_{41} = 1$	dimensionless	–
$\kappa_{18} = 100$	$\kappa_{42} = 100$	dimensionless	–
$\kappa_{19} = 160$	$\kappa_{43} = 200$	dimensionless	–
$\kappa_{20} = 80$	$\kappa_{44} = 100$	dimensionless	–
$\kappa_{21} = 80$	$\kappa_{45} = 100$	dimensionless	–
$\kappa_{22} = 140$	$\kappa_{46} = 205$	dimensionless	–
$\kappa_{23} = 70$	$\kappa_{47} = 105$	dimensionless	–
$\kappa_{24} = 70$	$\kappa_{48} = 105$	dimensionless	–

Table A.22: Reference parameter values used for the epigenetic regulation system of the plastic phenotype (Fig. 4.1)

Appendix B

Bertini description

In this thesis, a numerical software called Bertini has been used to compute steady state solutions of the systems of equations appearing in the ER models presented. In this Appendix, we give a brief summary of what Bertini does and the way this software works. For further information on this software, we refer to the book [13], where further details, as well as clarifying examples, are presented.

B.1 General overview

Bertini is a software that numerically computes algebraic sets, i.e, solution sets of systems of polynomial equations, which are subsets of the Euclidean space. It is based on numerical algebraic geometry which allows to compute to extremely high accuracy (with little work) approximations to the solutions. It is worth remarking that the methods that Bertini uses in the solving process depend on some random selection of constants which makes the algorithms work correctly with 'probability one', but which also implies that, the results from one run to another can be slightly different.

In particular, we will be interested in finding isolated solutions of polynomial systems, that is, zero-dimensional solution sets (although Bertini is able to compute positive-dimensional solution sets, we will not look into it). More precisely, we will use Bertini to find the solution set, more properly called an affine complex algebraic set, $\mathbf{V}(f_1, f_2, \ldots, f_n) := \{\mathbf{x} \in \mathbb{C}^n | f_1(\mathbf{x}) = \ldots = f_n(\mathbf{x}) = 0\}$, where f_i, $i = 1, \ldots, n$, is

a polynomial in N unknowns. More in general, we will denote by

$$\mathbf{f}(\mathbf{z}) := \begin{bmatrix} f_1(z_1, \ldots, z_N) \\ \vdots \\ f_n(z_1, \ldots, z_N) \end{bmatrix} \tag{B.1.1}$$

the polynomial system, with N not necessarily equal to n, that is, it is not necessary a square system for Bertini to work. In fact, when $n > N$ we will have an overdetermined system, whilst when $n < N$ we will work with an undetermined system (which will give rise to a positive-dimensional solution set). As stated before, we will not deal with this latter case and to be more precise, we will assume to have a square system, $n = N$. Notice too that $\mathbf{V}(f_1, f_2, \ldots, f_n)$ is more properly called an affine complex algebraic set.

B.2 How does Bertini work?

As we have mentioned, Bertini is based on numerical algebraic geometry and this area has as fundamental computation, homotopy continuation. Homotopy continuation deals with the problem of finding a finite set $\mathcal{S}$ of solutions of a (square) system such as (B.1.1), defined for $\mathbf{z} \in \mathbb{C}^N$, with the property that $\mathcal{S}$ contains every isolated solution of $\mathbf{f}(\mathbf{z}) = \mathbf{0}$. Let us assume that $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ is the system for which we wish to find the solutions, from now on, we will label this system as our target system and we will refer to it as system A.

Before proceeding further, let us recall what we mean by an isolated solution. We say that a solution $\mathbf{z}^* \in \mathbb{C}^N$ is isolated if there exists $r > 0$ such that the only solution contained in $B_r(\mathbf{z}^*) := \{\mathbf{z} \in \mathbb{C}^n \big| \quad ||\mathbf{z} - \mathbf{z}^*||_2 < r\}$ is $\mathbf{z}^*$.

Although it may seem too simplistic to consider the problem of finding (numerical) solutions for polynomial systems, this kind of systems are the ones that appear when modelling biological problems by means of ODEs (law of mass action, conservation laws, ...), for example. Hence, its importance. Furthermore, once it is known how to find the isolated solutions for a square system, it is quite easy to solve the problem for nonsquare systems and for positive-dimensional sets $\mathbf{V}(\mathbf{f})$.

In order to find the solutions of the system A, what Bertini does is taking a system B of 'similar' equations with known solutions, and deform the system B and its solutions to the system A and its solutions. This approach is called homotopy

continuation, or simply, continuation. More precisely, deforming the system B so as to reach the system A is accomplished by means of a parameterized family of equations, known as homotopy, which for certain values of the parameters, results in the systems A and B.

Definition 1. *A homotopy,* $\boldsymbol{H}(\boldsymbol{z},t)$*, is a smooth function* $\boldsymbol{H}(\boldsymbol{z},t) : \mathbb{C}^N \times [0,1] \to \mathbb{C}^N$

In Bertini, the homotopy $\mathbf{H}(\mathbf{z},t)$ denotes the composition of a family of systems, $\mathbf{H}(\mathbf{z},s)$, $\mathbf{H}(\mathbf{z},s) : \mathbb{C}^N \times U \to \mathbb{C}^N$, with $U \subset \mathbb{C}$ open subset, with a defined path $s = q(t)$, where s is given by a differentiable map $q : [0,1] \to U \subset \mathbb{C}$, such that $s = q(t)$, with $q(0) = 0$ and $q(1) = 1$, i.e, it gives the path parameterized by t going from $s = 1$ $(t = 1)$ to $s = 0$ $(t = 0)$. We want to observe that s is a complex variable and hence, there are an infinite number of continuous one-real dimensional paths going from $t = 1$ to $t = 0$.

There are 3 basic steps in order to solve a polynomial system $\mathbf{f}(\mathbf{z})$, i.e, to find the solutions of the target system A via homotopy continuation. These are as follows:

1. Build and solve a start system $\mathbf{g}(\mathbf{z})$ (usually, $\mathbf{g}(\mathbf{z})$ is simpler than $\mathbf{f}(\mathbf{z})$).
2. Construct a homotopy between $\mathbf{f}(\mathbf{z})$ (target system) and $\mathbf{g}(\mathbf{z})$.
3. Follow the paths from $t = 1$ to $t = 0$, that is, to deform the system $\mathbf{g}(\mathbf{z})$ and its solutions to $\mathbf{f}(\mathbf{z})$ and its respective solutions (see below for further explanation).

If both the start system $\mathbf{g}(\mathbf{z})$ and the homotopy (i.e, the deformation rule) are well-chosen, then, this 3-step process will lead to the solutions of the target system $\mathbf{f}(\mathbf{z})$. Let us describe each of these three steps a bit more in detail.

Step one. In order to construct a polynomial system $\mathbf{g}(\mathbf{z})$ which is related to $\mathbf{f}(\mathbf{z})$, we can take, for instance, a system with the same degrees, but which we can solve easily. The solutions of this new system, i.e, the solutions of $\mathbf{g}(\mathbf{z}) = \mathbf{0}$, are called in Bertini *startpoints*.

Step two and three. The simplest form for the homotopy is a linear one, that is, $\mathbf{H(z,t)} = t\mathbf{g}(\mathbf{z}) + (1-t)\mathbf{f}(\mathbf{z})$, with t a new parameter, $t \in [0,1]$. When taking $t = 1$, we have $\mathbf{H}(\mathbf{z},1) = \mathbf{g}(\mathbf{z})$, for which we know the solutions $\mathbf{H}(\mathbf{z},1) = \mathbf{0}$, whereas when taking $t = 0$, we have $\mathbf{H}(\mathbf{z},0) = \mathbf{f}(\mathbf{z})$, for which we want to compute the solutions $\mathbf{H}(\mathbf{z},0) = \mathbf{f}(\mathbf{z}) = \mathbf{0}$.

The key point is that for $t \in (0, 1]$ the solutions vary continuously and they will not collide, which, then, implies that we will get paths connecting the *startpoints* (solutions of the start system, $\mathbf{g}(\mathbf{z})$, at $t = 1$) to the *endpoints*, i.e, solutions of the target system, $\mathbf{f}(\mathbf{z})$, at $t = 0$. In other words, the continuous variation of the solutions defines a path, and each of these paths is the image of the interval [0,1] under a continuous map, $\mathbf{p}(t)$; that is, $\mathbf{H}(\mathbf{p}(t), t) = \mathbf{0}$ for $t \in [0, 1]$ defines a path that deforms the solutions of the system B (for $t = 1$) to the ones of the system A (for $t = 0$).

To march along the paths (moving from $t = 1$ to $t = 0$), Bertini uses predictor-corrector methods, although when approaching $t = 0$, then, more powerful methods are used. We want to note that some of the paths could be divergent (in Bertini, these paths are referred as paths approaching a point 'at infinity' when $t \to 0$), which means that the initial system had more solutions than the target system. Another important remark is that the continuation paths are nonreal for most of the homotopy (during the 'tracking steps'), but they may be real at the start and end points of the homotopy. More in general, continuing real solutions not always leads to real solutions, that is, real solutions can continue to complex solutions and vice versa.

B.2.1 Choosing a good homotopy

As we have stated, the choice of the continuation rules, the homotopy, is an important step in order to find the solutions of the target system. Hence, let us show some of the properties that a good homotopy must satisfy.

A good homotopy for the system $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ and a set $\mathcal{S}_1$ of D distinct solutions of $\mathbf{g}(\mathbf{z}) := \mathbf{H}(\mathbf{z}, 1) = \mathbf{0}$ (*startpoints*), is a system of infinitely differentiable functions

$$\mathbf{H}(\mathbf{z}, t) := \begin{bmatrix} H_1(z_1, \ldots, z_N, t) \\ \vdots \\ H_N(z_1, \ldots, z_N, t) \end{bmatrix} \tag{B.2.1}$$

such that

1. for any choice of $t \in [0, 1]$, $\mathbf{H}(\mathbf{z}, t)$ is a system of polynomials.
2. for any of the D distinct *startpoints*, $\omega \in \mathcal{S}_1$, there is a smooth path $\mathbf{p}_j(t) : (0, 1] \to \mathbb{C}^N$, such that $\mathbf{p}_j(1) = \omega$, and

3. $\mathbf{p}_j$ is smooth on $(0,1]$, and for each $t^* \in (0,1]$,
   (a) the associated paths don't cross, i.e, for any $k \neq j$, $1 \leq j,k \leq D$, and $t^* \in (0,1]$, $\mathbf{p}_j(t^*) \neq \mathbf{p}_k(t^*)$, and
   (b) the points $\mathbf{p}_j(t^*)$ are smooth isolated solutions of $\mathbf{H}(\mathbf{z}, t^*) = 0$.

4. $\mathbf{H}(\mathbf{p}_j(t), t) = \mathbf{0}$, $\forall t \in (0,1]$ and $1 \leq j \leq D$.

In addition, we say that Eq. (B.2.1) is a good homotopy for the system $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ if we can choose a set $\mathcal{S}_1$ of D distinct solutions of $\mathbf{g}(\mathbf{z}) := \mathbf{H}(\mathbf{z}, 1) = \mathbf{0}$ such that the set of finite limits,

$$\mathcal{S}_0 := \{\mathbf{z} \in \mathbb{C}^N \big| ||\mathbf{z}||_2 < \infty \text{ and } \mathbf{z} = \lim_{t \to 0} \mathbf{p}_j(t)\}$$

contains every isolated solution of $\mathbf{f}(\mathbf{z}) = \mathbf{0}$, where $||\cdot||_2$ denotes the two-norm of a vector.

Without going into all the details, we can state that a good homotopy, according to the description given above, can be build by applying the following recipe:

$$\mathbf{H}(\mathbf{z}, t) = (1-t)\mathbf{f}(\mathbf{z}) + \gamma t \mathbf{g}(\mathbf{z}),$$

where, for instance, $g_i(z) = {z_i}^{d_i} - 1$, with $d_i = \deg f_i$, $i = 1, \ldots, N$, and γ, $\gamma \neq 0$, is a random complex number (usually taken to be a number following the uniform distribution in a small band around the unit circle). This is known as the gamma trick and it guarantees for $\mathbf{H}(\mathbf{z}, t)$ to be a good homotopy (with probability one). In fact, if the homotopy is not specified by the user, this is Bertini's default method for constructing the homotopy. This type of homotopy is known as total-degree homotopy and it guarantees that the number of paths ending at z^* is equal to the multiplicity of z^* as solution of $\mathbf{f}(\mathbf{z}) = \mathbf{0}$.

More in general, numerical algebraic geometry is based on algorithms that depend on certain bad phenomena happening with probability zero. That is, when choosing the value of some parameters at random, bad phenomena don't occur, which is the basis of the gamma trick. And this is precisely the reason why we can state that the total-degree homotopy will be a good one, with probability one.

B.2.2 Path tracking details

Let start by assuming that we have a family of homotopies on $\mathbb{C}^N$, that is,

$$\mathbf{H}(\mathbf{z},t) := \begin{bmatrix} H_1(z_1,\ldots,z_N,t) \\ \vdots \\ H_N(z_1,\ldots,z_N,t) \end{bmatrix} = \mathbf{0}, \tag{B.2.2}$$

such that each H_i is polynomial in the variables $\mathbf{z} \in \mathbb{C}^N$ and analytic in the parameter $t \in \mathbb{C}$ (although the definition can be generalised to a set of parameters $\mathbf{t} \in \mathbb{C}^M$, we focus on the case $M = 1$ for simplicity). Assume that we also have a differentiable map $\psi : t \in [0,1] \to \mathbf{z} \in \mathbb{C}^N$ satisfying that $\mathbf{H}(\psi(t),t) = \mathbf{0}$ for $t \in (0,1]$ and the Jacobian of $\mathbf{H}$ with respect to $z_1,\ldots,z_N$ has rank N for the points $(\psi(t),t)$ with $t \in (0,1]$ (i.e, the points $(\psi(t),t)$ are nonsingular).

The way $\mathbf{H}$ is constructed guarantees that ψ exists and that $\psi(1) = \mathbf{p}_0$, a prescribed point (*startpoint*). The goal is to compute the solution of the target system, $\mathbf{f}(\mathbf{z}) = \mathbf{0}$, corresponding to this particular path, i.e, computing $\mathbf{p}^* = \psi(0)$. Trying to compute this point is what is known as path tracking.

For simplicity, let us denote $\psi(t)$ by $\mathbf{z}(t)$. The process of path tracking described in the previous paragraph is equivalent to solving the following initial value problem

$$\frac{\partial \mathbf{H}(\mathbf{z}(t),t)}{\partial t} + \sum_{i=1}^{N} \frac{\partial \mathbf{H}(\mathbf{z}(t),t)}{\partial z_i}\frac{dz_i}{dt} = \mathbf{0}, \text{ with } \mathbf{z}(1) = \mathbf{p}_0, \tag{B.2.3}$$

which corresponds to the Davidenko differential equation and which is obtained by differentiating $\mathbf{H}(\mathbf{z}(t),t)$ with respect to t. Denoting by $\mathbf{z}(t) := (z_1(t),\ldots,z_N(t))^t$ the solution of Eq. (B.2.3), and by

$$J\mathbf{H}(\mathbf{z},t) := \frac{\partial \mathbf{H}}{\partial \mathbf{z}} := \begin{pmatrix} \frac{\partial H_1}{\partial z_1} & \cdots & \frac{\partial H_1}{\partial z_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial H_N}{\partial z_1} & \cdots & \frac{\partial H_N}{\partial z_N} \end{pmatrix},$$

the Jacobian matrix with respect to the variable $\mathbf{z}$ evaluated at $(\mathbf{z},t)$, then, the Davidenko differential equation reads as follows,

$$\frac{\partial \mathbf{H}(\mathbf{z}(t),t)}{\partial t} + J\mathbf{H}(\mathbf{z}(t),t)\frac{d\mathbf{z}(t)}{dt} = \mathbf{0},$$

where using the fact that the Jacobian $J\mathbf{H}(\mathbf{z}(t),t)$ is invertible in the path since the matrix has rank N at any point of the path, we obtain

$$\frac{d\mathbf{z}(t)}{dt} = -\big(J\mathbf{H}(\mathbf{z}(t),t)\big)^{-1}\frac{\partial \mathbf{H}(\mathbf{z}(t),t)}{\partial t} \text{ with } \mathbf{z}(1) = \mathbf{p}_0. \tag{B.2.4}$$

Solving the initial value problem, Eq. (B.2.4), will allow to find the roots of $\mathbf{H}(\mathbf{z},0) = \mathbf{0}$. It is extremely important to highlight that we need to be specially careful when approaching $t \to 0$, because if the solution is singular, then, the Jacobian will not be invertible (rank defficient) and Eq. (B.2.4) will not be valid. This is the reason why Bertini usually splits the interval $t \in (0,1]$ into $t \in [\epsilon, 1]$, with $\epsilon > 0$ small, and then taking the limit to estimate $\lim_{t\to 0}\mathbf{z}(t)$.

We want to remark that Bertini tracks the paths from $t = 1$ to $t = 0$ (and not the other way around) because there are considerably more floating-point numbers near $t = 0$ than near $t = 1$. Since the singular *endpoints* require many digits of accuracy, this turns out to be useful. In fact, Bertini uses multiprecision because it is the only way to have an accurate polynomial solving software. Nevertheless, multiprecision needs to be used only when necessary, due to its high computational cost.

In the previous lines, we have theoretically described how Bertini does the path tracking. However, what is useful to know is which algorithms Bertini uses to do the path tracking. In general, Eq. (B.2.4) is solved by using a first-order tracking method, the Euler-Newton predictor-corrector. So, the method starts at $t_0 = 1$ with $\mathbf{p}_0$ as the initial value and computes approximations $p_1, p_2, \ldots$ at values $1 = t_0 > t_1 > t_2 > \ldots > 0$. In order to do so, it applies Euler's method, that is,

$$\mathbf{p}_{i+1} = \mathbf{p}_i - \big(J\mathbf{H}(\mathbf{p}_i,t_i)\big)^{-1}\frac{\partial \mathbf{H}(\mathbf{p}_i,t_i)}{\partial t}\Delta t_i,$$

where $\Delta t_i = t_{i+1} - t_i$. To be more precise, what this method does is predicting the next approximation, $\mathbf{p}_{i+1}$, along the line tangent to the solution path from the current point of the path. Since we know that each point of the path must satisfy $\mathbf{H}(\mathbf{z},t) = \mathbf{0}$, we can use a predictor-corrector strategy so as to improve the point given by the Euler's method. So, the prediction to $\mathbf{p}_{i+1}$ is followed by a correction using Newton's method for $\mathbf{H}(\mathbf{z},t_{i+1})$, starting with $\mathbf{z}_0 = \mathbf{p}_{i+1}$. More precisely, Newton's method is as follows:

$$\mathbf{z}_{i+1} = \mathbf{z}_i - \big(J\mathbf{H}(\mathbf{z}_i,t_{i+1})\big)^{-1}\mathbf{H}(\mathbf{z}_i,t_i)$$

Applying this method, we can get a better prediction for $\mathbf{z}(t_{i+1})$ which we can use to replace $\mathbf{p}_{i+1}$. Therefore, path tracking accuracy is determined by Newton's method

in the predictor-corrector cycle. Moreover, Bertini computes the condition number of the Jacobian matrix as it tracks each path, increasing and decreasing precision as necessary.

Another question we could ask about the path tracking process is if all the solutions of $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ are found when using this methodology. The following Theorem guarantees this:

Theorem 1. *Given a system such as in Eq.* (B.2.2) *of N complex analytic functions defined for $(\boldsymbol{z}, t)$ in an open set $U \subset \mathbb{C}^N \times \mathbb{C}$, and a point ($\boldsymbol{z}^*$,$t$), which is an isolated solution of multiplicity μ of $\boldsymbol{H}(\boldsymbol{z}^*, t) = \boldsymbol{0}$, it follows that for any sufficiently small $\epsilon > 0$, there is a $\delta > 0$ such that for any point t' with the Euclidean distance $|t' - t^*| < \delta$, the sum of the multiplicities of the solutions of $\boldsymbol{H}(\boldsymbol{z}, t') = \boldsymbol{0}$ with Euclidean distance less than ϵ from $\boldsymbol{z}^*$ is μ.*

This theorem states that given a total-degree homotopy, then, exactly μ paths will have as *endpoints* the $\mathbf{z}$ value, with $|z - z^*| < \epsilon$ as the parameter of the homotopy $t \to 0$.

Bibliography

[1] Tomás Alarcón. Stochastic quasi-steady state approximations for asymptotic solutions of the Chemical Master Equation. *J. Chem. Phys.*, 140:184109, 2014.

[2] Tomás Alarcón and Karen M Page. Mathematical models of the VEGF receptor and its role in cancer therapy. *J. R. Soc. Interface*, 4:283–304, 2007.

[3] C David Allis, Marie-Laure Caparros, Thomas Jenuwein, Danny Reinberg, and Monika Lachner. *Epigenetics*. Cold Spring Harbor Laboratory Press, second edition edition, 2015.

[4] M F Amary, S Damato, D Halai, M Eskandarpour, F Berisha, F Bonar, S McCarthy, V R Fantin, K S Straley, S Lobo, W Aston, C L Green, R E Gale, R Tirabosco, A Futreal, P Campbell, N Presneau, and Flanagan A M. Ollier disease and maffucci syndrome are caused by somatic mosaic mutations of idh1 and idh2. *Nat. Genet.*, 43:1262–1265, 2011.

[5] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Springer, New York, NY, USA, 2010.

[6] Andrew Angel, Jie Song, Caroline Dean, and Martin Howard. A polycomb-based switch underlying quantitative epigenetic memory. *Nature*, 476:105–108, 2011.

[7] Michael Assaf and Baruch Meerson. Spectral formulation and WKB approximation for rare-event statistics in reaction systems. *Phys. Rev. E*, 74:041115, 2006.

[8] Michael Assaf, Baruch Meerson, and Pavel V Sasorov. Large fluctuations in stochastic population dynamics: Momentum space calculations. *J. Stat. Mech.*, page P07018, 2010.

[9] Anne L Baldock, Kevin Yagle, Donald E Born, Sunyoung Ahn, Andrew D Trister, Maxwell Neal, Sandra K Johnston, Carly A Bridge, David Basanta, Jacob Scott, Hani Malone, Adam M Sonabend, Peter Canoll, Maciej M Mrugala, Jason K Rockhill, Russell C Rockne, and Kristin R Swanson. Invasion and proliferation kinetics in enhancing gliomas predict idh1 mutation status. *Neuro-Oncol.*, 16:779–786, 2014.

[10] Karen Ball, Thomas G Kurtz, Lea Popovic, and Greg Rempala. Asymptotic analysis of multi-scale approximations to reaction networks. *Ann. App. Prob.*, 16:1925–1961, 2006.

[11] A J Banniser and T Kouzarides. Regulation of chromatin by histone modifications. *Cell Res.*, 21:381 – 395, 2011.

[12] David Basanta, Jacob G Scott, Russ Rockne, Kristin R Swanson, and Alexander R A Anderson. The role of idh1 mutated tumour cells in secondary glioblastomas: an evolutionary game theoretical view. *Physical Biology*, 8(1):015016, 2011.

[13] Daniel J. Bates, Jonathan D. Hauenstein, Andrew J. Sommese, and Charles W. Wampler. *Numerically solving polynomial systems with Bertini*, volume 25. Software, Environments, and Tools. SIAM, 2013.

[14] B A Benayoun, E A Polina, and A Brunet. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell. Biol.*, 16:593–610, 2015.

[15] Scott Berry, Caroline Dean, and Martin Howard. Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity. *Cell Systems*, 4:445–457, 2017.

[16] Lacramioara Bintu, John Young, Yaron E Antebi, Kayla McCue, Yasuhiro Kazuki, Narumi Uno, Mitsuo Oshimura, and Michael B Elowitz. Dynamics of epigenetic regulation at the single-cell level. *Science*, 351:720–724, 2016.

[17] F E Bleeker, N A Atai, S Lamba, A Jonker, D Rijkeboer, K S Bosch, Tigchelaar. W, D Troost, W P Vandertop, A Bardelli, and C J Van Noorden. The prognostic idh1(r132) mutation is associated with reduced nadp+-dependent idh activity in glioblastoma. *Acta Neuropathol.*, 119:487–494, 2010.

[18] L N Booth and A Brunet. The aging epigenome. *Mol. Cell.*, 62:728–744, 2016.

[19] Paul C Bressloff. *Stochastic processes in cell biology.* Springer-Verlag, Berlin, Germany., 2014.

[20] Paul C Bressloff and Sean D Lawley. Mean first passage times for piecewise deterministic markov processes and the effects of critical points. *Journal of Statistical Mechanics: Theory and Experiment*, 2017:063202, 2017.

[21] Paul C Bressloff and Jay M Newby. Path integrals and large deviations in stochastic hybrid systems. *Phys. Rev. E*, 89:042701, 2014.

[22] George E Briggs and John B S Haldane. A note on the kinetics of enzyme action. *Biochem. J.*, 19:338–39, 1925.

[23] Yosef Buganim, Dina A Faddah, Albert W Cheng, Elena Itskovich, Styliani Markoulaki, Kibibi Ganz, Sandy L. Klemm, Alexander van Oudenaarden, and Rudolf Jaenisch. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150:1209–1222, September 2012.

[24] J Camacho-Pereira, M G Tarrago, C C Chini, V Nin, C Escande, G M Warner, A S Puranik, R A Schoon, J M Reid, A Galina, and Chini E N. Cd38 dictates age-related nad decline and mitochondrial dysfunction through an sirt3-dependent mechanism. *Cell Metab.*, 23:1127–1139, 2016.

[25] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comp. Phys.*, 206:395–411, 2005.

[26] Yang Cao, Daniel T Gillespie, and Linda R Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, 122:014116, 2005.

[27] Nessa Carey. *The epigenetics revolution.* Icon Books Ltd, 2012.

[28] R Chetty. Familial paraganglioma syndromes. *J. Clin. Pathol.*, 63:488–491, 2010.

[29] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661–703, 2009.

[30] A L Cohen, S L Holmen, and H Colman. Idh1 and idh2 mutations in gliomas. *Curr. Neurol. Neurosci. Rep.*, 13:345, 2013.

[31] Ruggero Cortini, Maria Barbi, Bertran R Care, Christophe Lavelle, Annick Lesne, Julien Mozziconacci, and Jean-Marc Victor. The physics of epigenetics. *Rev. Mod. Phys.*, 88:025002, 2016.

[32] Lenny Dang, Shengfang Jin, and Shinsan M Su. Idh mutations in glioma and acute myeloid leukemia. *Trends in Molecular Medicine*, 16:387–397, 2010.

[33] Weiwei Dang, Kristan K Steffen, Rocco Perry, Jean A Dorsey, F Brad Johnson, Ali Shilatifard, Matt Kaberlein, Brian K Kennedy, and Shelley L Berger. Histone h4 lysine 16 acetylation regulates cellular lifespan. *Nature*, 459:802–807, 2009.

[34] Bryan C Daniels, Yan-Jiun Chen, James P Sethna, Ryan N Gutenkurst, and Christopher R Myers. Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19:389–395, 2008.

[35] D. A. Darling. The kolmogorov-smirnov, cramer-von mises tests. *Ann. Math. Statist.*, 28:823–838, 1957.

[36] Justine Dattani and Mauricio Barahona. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *Journal of The Royal Society Interface*, 14, 2017.

[37] Diana David-Rus, Swagatam Mukhopadhyway, Joel L Lebowitz, and Anirvan M Sengupta. Inheritance of epigenetic chormatin silencing. *J. Theor. Biol.*, 258:112–120, 2009.

[38] Mark H A Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46:353–388, 1984.

[39] Peter L J de Keizer. The fountain of youth by targeting senescent cells? *Trends Mol. Med.*, 23:6–17, 2017.

[40] Roberto de la Cruz, Pilar Guerrero, Fabian Spill, and Tomás Alarcón. The effects of intrinsic noise on the behaviour of bistable systems in quasi-steady state conditions. *J. Chem. Phys.*, 143:074105, 2015.

[41] L Dimitrov, C S Hong, C Yang, Z Zhuang, and J D Heiss. New developments in the pathogenesis and therapeutic targeting of the idh1 mutation in glioma. *Int. J. Med. Sci.*, 12:201–213, 2015.

[42] Ian B Dodd, Mille A Micheelsen, Kim Sneppen, and Genevieve Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129:813–822, 2007.

[43] M Doi. Stochastic theory of diffusion-controled reaction. *J. Phys. A:Math. Gen.*, 9(9):1479, 1976.

[44] A Dutta, S M Abmayr, and J L Workman. Diverse activities of histone acylations connect metabolism to chromatin function. *Mol. Cell*, 63:547–552, 2016.

[45] Mark I Dykman, Takehiko Horita, and John Ross. Statistical distribution and stochastic resonance in a periodically driven chemical system. *J. Chem. Phys.*, 103:966–972, 1995.

[46] Vlad Elgart and Alex Kamenev. Rare events in reaction-diffusion systems. *Phys. Rev. E*, 70:041106, 2004.

[47] Carlos Escudero and Alex Kamenev. Switching rates of multistep reactions. *Phys. Rev. E*, 79:041149, 2009.

[48] Jason Feser and Jessica Tyler. Chromatin structure as a mediator of aging. *FEBS Lett.*, 585(13):2041–2048, 2011.

[49] A Field and P D Adams. Targeting chromatin aging - the epigenetic impact of longevity-associated interventions. *Exp. Gerontol.*, page To Appear., 2016.

[50] Sarah Filippi, Chris P Barnes, Paul D W Kirk, Takamasa Kudo, Katsuyuki Kunida, Siobhan S McMahon, Takaho Tsuchiya, Takumi Wada, Shinya Kuroda, and Michael P H Stumpf. Robustness of MEK-ERK dynamics and origins of cell-to-cell variability in MAPK signaling. *Cell Reports*, 15:2524–2535, 2016.

[51] William A Flavahan, Elizabeth Gaskell, and Badley E Bernstein. Epigenetic plasticity and the hallmarks of cancer. *Science*, 357:eaal2380, 2017.

[52] A Flores and et al. Lactate dehydrogenase activity drives hair follicle stem cell activation. *Nat. Cell Biol.*, 19:1017 – 1026, 2017.

[53] Núria Folguera-Blasco, Elisabet Cuyàs, Javier A Menéndez, and Tomás Alarcón. Epigenetic regulation of cell fate reprogramming in aging and disease: A predictive computational model. *PLoS Comp. Biol.*, 14:e1006052, 2018.

[54] Mark I Freidlin and Alexander D Wentzell. *Random perturbations of dynamical systems*. Springer-Verlag, New York, NY, USA., 1998.

[55] David Frigola, Laura Casanellas, Jose M Sancho, and Marta Ibañes. Asymmetric stochastic switching driven by intrinsic molecular noise. *PLoS One*, 7:e31407, 2012.

[56] X Gao, M A Reid, M Kong, and J W Locasale. Metabolic interactions with cancer epigenetics. *Molecular Aspects of Medicine*, 54:650–657, 2017.

[57] Crispin W Gardiner. *Stochatic methods*. Springer-Verlag, Berlin, Germany, 2009.

[58] Hao Ge, Hong Qian, and X Sunney Xie. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys. Rev. Lett.*, 114:078101, 2015.

[59] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.

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

[61] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 155:1716–1733, 2001.

[62] A P Gomes, N L Price, A J Ling, J J Moslehi, M K Montgomery, L Rajman, J P White, J S Teodoro, C D Wrann, B P Hubbard, E M Mercken, C M Palmeira, R de Cabo, A P Rolo, N Turner, E L Bell, and D A Sinclair. Declining nad(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. *Cell*, 155:1624–1638, 2013.

[63] D Gonze, J Halloy, and P Gaspard. Biochemical clocks and molecular noise: Theoretical study of robustness factors. *J. Chem. Phys.*, 116:10997–11010, 2002.

[64] Thomas Graf and Tariq Enver. Forcing cells to change lineages. *Nature*, 462:587–594, December 2009.

[65] Pilar Guerrero, Helen M Byrne, Philip K Maini, and Tomás Alarcón. From invasion to latency: Intracellular noise and cell motility as key controls of the competition between resource-limited cellular populations. *J. Math. Biol.*, 72:123–156, 2015.

[66] Shangqin Guo, Xiaoyuan Zi, Vincent P Schulz, Jijun Cheng, Mei Zhong, Sebastian H J Koochaki, Cynthia M Megyola, Xinghua Pan, Kartoosh Heydari, Sherman M Weissman, Patrick G Gallagher, Diane S Krause, Rong Fan, and Jun Lu. Nonstochastic reprogramming from a privileged somatic cell state. *Cell*, 156:649–662, 2014.

[67] S Han and A Brunet. Histone methylation makes its mark on longevity. *Trends Cell. Biol.*, 22:42–49, 2012.

[68] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144:646–674, 2011.

[69] Peter Hanggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251, 1990.

[70] Jacob Hanna, Krishanu Saha, Bernardo Pando, Jeroen van Zon, Christopher J Lengner, Menno P Creyghton, Alexander van Oudenaarden, and Rudolf Jaenisch. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, 462(3):595–601, 2009.

[71] HeatherA. Harrington, Elisenda Feliu, Carsten Wiuf, and MichaelP.H. Stumpf. Cellular compartments cause multistability and allow cells to process more information. *Biophysical Journal*, 104:1824–1831, 2013.

[72] Jeff Hasty and Farren Isaacs. Designer gene networks: Towards fundamental cellular control. *Chaos*, 11(1):207–220, 2001.

[73] Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: a least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, LXI:1052–1117, 2008.

[74] I B Hilton, A M DIpolito, C M Vockley, P I Thakore, G E Crawford, T E Reddy, and C A Gerbasch. Epigenome editing by a crispr-cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotech.*, 33:510 – 517, 2016.

[75] Robert Hinch and S Jon Chapman. Exponentially slow transitions on a Markov chain: the frequency of calcium sparks. *Eur. J. Appl. Math.*, 16:427–446, 2005.

[76] Mark H Holmes. *Introduction to perturbation methods*. Springer-Verlag, New York, NY, USA, 1995.

[77] Sui Huang. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*, 31:546–560, 2009.

[78] Sui Huang. The molecular and mathematical basis of wadington's epigenetic landscape: a framework for post-darwinian biolgy? *BioEssays*, 34:149–157, 2012.

[79] Sui Huang, Gabriel Eichler, Yaneer Bar-Yam, and Donald E Ingber. Cell fates as high-dimensional attractors of a complex gene regulatory network. *Phys. Rev. Lett.*, 94:128701, 2005.

[80] Marta Inglés, Cristina Mas-Bargues, Alejandro Berna-Erro, Ander Matheu, Paula Sanchís, Juan-Antonio Avellana, Consuelo Borrás, and José Via. Centenarians overexpress pluripotency-related genes. *The Journals of Gerontology: Series A*, gly(168), 2018.

[81] Llewellyn E Jalbert, Adam Elkhaled, Joanna J Phillips, Evan Neill, Aurelia Williams, Jason C Crane, Marram P Olson, Annette M Molinaro, Mitchel S Berger, John Kurhanewicz, Sabrina M Ronen, Susan M Chang, and Sarah J Nelson. Metabolic prolifing of idh mutation and malignant progression in infiltrating glioma. *Scientific Reports*, 7:44792, 2017.

[82] N Jiang, G Du, E Tobias, J G Wood, R Whitaker, N Neretti, and S L Helfand. Dietary and genetic effects on age-related loss of gene silencing reveal epigenetic plasticity of chromatin repression during aging. *Aging*, 5:813–824, 2013.

[83] C Johnson, M O Warmoes, X Shen, and J W Locasale. Epigenetics and cancer metabolism. *Cancer Letters*, 356:309–314, 2015.

[84] Hye-Won Kang and Thomas G Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *The Annals of Applied Probability*, 23(2):529–583, 2013.

[85] Hye-Won Kang, Thomas G Kurtz, and Lea Popovic. Central limit theorems and diffusion approximation for multiscale markov chain models. *Ann. Prob.*, 24:721–759, 2014.

[86] Stuart A Kauffman. *The origins of order*. Oxford University Press, New York, U.S.A., 1993.

[87] S T Keatinga and A El-Osta. Epigenetics and metabolism. *Circ. Res.*, 167:715–736, 2015.

[88] James Keener and James Sneyd. *Mathematical physiology*. Springer-Verlag, New York, NY, USA, 1998.

[89] B K Kennedy, S L Berger, A Brunet, K Campisi, A M Cuervo, ES Epel, C Franceschi, G J Lightgow, R I Morimoto, J E Pessin, T A Rando, A Richardsona, E E Schadt, T Wyss-Coray, and F Sierra. Geroscience: linking aging to chronic disease. *Cell*, 159:709–713, 2014.

[90] A Kinnaird, S Zhao, K E Wellen, and E D Michelakis. Metabolic control of epigenetics in cancer. *Nature Reviews Cancer*, 16:694–707, 2016.

[91] Alfred G Knudson. Mutation and cancer: statistical study of retinoblastoma. *Proc. Nat. Acad. Sci.*, 68:820–823, 1971.

[92] Tony Kouzarides. Chromatin modifications and their function. *Cell*, 128:693–705, 2007.

[93] Ryogo Kubo, Kazuhiro Matsuo, and Kazuo Kitahara. Fuctuation and relaxation of macrovariables. *J. Stat. Phys.*, 9:51–96, 1973.

[94] Thomas G Kurtz. Strong approximation theorems for density dependent markov chains. *Stochastic Processes and their Applications*, 6(3):223 – 240, 1978.

[95] E Letouze, C Martinelli, C Loriot, N Burnichon, N Abermil, C Ottolenghi, M Janin, M Menara, A T Nguyen, P Benit, A Buffet, C Marcaillou, J Bertherat, L Amar, P Rustin, A De Reynies, A P Gimenez-Roqueplo, and Favier J. Sdh mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 23:739–752, 2013.

[96] Y Li, M Daniel, and T O Tollefsbol. Epigenetic regulation of caloric restriction in aging. *BMC Med.*, 9:98, 2011.

[97] Yen Ting Li, Peter G Hufton, Esther J Lee, and Davit A Potoyan. A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. *PLoS Comp. Biol.*, 14:e1006000, 2018.

[98] C Lips, E Lentjes, J Hoppener, R V Luijt, and Moll F. Familial paragangliomas. *Hered. Cancer Clin. Pract.*, 4:169–176, 2006.

[99] Cecilia Lovkvist, Ian B Dodd, Kim Sneppen, and Jan O Haerter. Dna methylation in human epigenomes depends on local topology of cpg sites. *Nucleic Acids Res.*, 44:5123–5132, 2016.

[100] Chao Lu and Craig B Thomson. Metabolic regulation of epigenetics. *Cell Metabolism*, 16:9–17, 2012.

[101] Chao Lu, Patrick S Ward, Gurpreet S Kapoor, Dan Rohle, Sevin Turcan, Omar Abdel-Wahab, Christopher R Edwards, Raya Khanin, Maria E Figueroa, Ari Melnick, Kathryn E Wellen, Donald M O'Rourke, Shelley L Berger, Timothy A Chan, Ross L Levine, Ingo K Mellinghoff, and Craig B Thompson. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*, 483:474–478, 2012.

[102] Ben D MacArthur, Colin P Please, and Richard O C Oreffo. Stochasticity and the molecular mechanisms of induced pluripotency. *PLoS One*, 3:e3086, 2008.

[103] Robert S Maier and D L Stein. A scaling theory of bifurcations in the symmetric weak-noise escape problem. *J. Stat. Phys.*, 83:291–357, 1996.

[104] Monika Maleszewska, Julia S P Mawer, and Peter Tessarz. Histone modifications in ageing and lifespan regulation. *Curr Mol Bio Rep*, 2:26–35, 2016.

[105] Walter de Back, Joseph X Zhou, and Lutz Brusch. On the role of lateral stabilization during early patterning in the pancreas. *J. R. Soc. Interface*, 10:20120766, 2012.

[106] Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science*, 324:1029–1033, 2009.

[107] Javier A Menendez and Tomas Alarcón. Metabostemness: a new cancer hallmark. *Front. Oncol.*, 4:262, 2014.

[108] Javier A Menendez and Tomas Alarcon. Senescence-inflammatory regulation of reparative cellular reprogramming in aging and cancer. *Front. Cell Dev. Biol.*, 5:49, 2017.

[109] Javier A Menendez, Tomas Alarcon, and Jorge Joven. Gerometabolites: the pseudohypoxic aging side of cancer oncometabolites. *Cell Cycle*, 13:699–709, 2014.

[110] Javier A Menéndez, Bruna Corominas-Faja, Elisabet Cuyàs, María G. García, Salvador Fernández-Arroyo, Agustín F Fernández, Jorge Joven, Mario F Fraga, and Tomás Alarcón. Oncometabolic nuclear reprogramming of cancer stemness. *Stem Cell Reports*, 6:273–283, 2016.

[111] Javier A Menendez, Nuria Folguera-Blasco, Elisabet Cuyas, Salvador Fernandez-Arroyo, Jorge Joven, and Tomas Alarcon. Accelerated geroncogenesis in hereditary breast-ovarian cancer syndrome. *Oncotarget*, 7:11959–11971, 2016.

[112] Javier A Menéndez, Jorge Joven, Sívia Cufí, Bruna Corominas-Faja, Cristina Oliveros-Ferreros, Elisabet Cuyás, Begoña Martín-Castillo, Eugeni López-Bonet, Tomás Alarcón, and Alejandro Vázquez-Martín. The Warburg effect 2.0. Metabolic reprogramming of cancer stem cells. *Cell Cycle*, 12:1166–1179, 2013.

[113] S J Mentch and J W Locasale. One-carbon metabolism and epigenetics: understanding the specificity. *Ann. N. Y. Acad. Sci.*, 1363:91–98, 2016.

[114] A Morin, E Letouze, A P Gimenez-Roqueplo, and J Favier. Oncometabolites-driven tumorigenesis: From genetics to targeted therapy. *Int. J. Cancer*, 135:2237–2248, 2014.

[115] Lluc Mosteiro, Cristina Pantoja, Noelia Alcazar, Rosa M Marion, Dafni Chondronasin, Miguel Rovira, Pablo J Fernandez-Marcos, Maribel Muñoz-Martin, Carmen Blanco-Aparicio, Joaquin Pastor, Gonzalo Gomez-Lopez, Alba De Martino, Maria A Blasco, Maria Abad, and Manuel Serrano. Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. *Science*, 354:aaf4445, 2016.

[116] Alice Moussy, Jérémie Cosette, Romuald Parmentier, Cindy da Silva, Guillaume Corre, Angélique Richard, Olivier Gandrillon, Daniel Stockholm, and András Páldi. Integrated time-lapse and single-cell transcription studies highlight the variable and dynamic nature of human hematopoietic cell fate commitment. *PLoS Biology*, 15(7):e2001867, 2017.

[117] Werner A Müller. *Developmental biology*. Springer-Verlag New York, 1997.

[118] C O Nordling. A new theory on the cancer-inducing mechanism. *British J. Cancer*, 7:68–72, 1953.

[119] S Nowicki and E Gottlieb. Oncometabolites: tailoring our genes. *FEBS J.*, 282:2796–2805, 2015.

[120] A Ocampo, P Reddy, and J C Izpisua-Belmonte. Anti-aging strategies based on cellular reprogramming. *Trends Mol. Med.*, 22:725–738, 2016.

[121] A Ocampo, P Reddy, P Martinez-Redondo, A Platero-Luengo, F Hatanaka, T Hishida, M Li, D Lam, M Kurita, E Beyret, T Araoka, E Vazquez Ferrer, D Donoso, J L Roman, J Xu, C Rodriguez-Estaban, G Nu nez, E Nu nez-Delicado, J M Campistol, I Guillen, and J C Izpisua-Belmonte. In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell*, 167:1719–1733, 2016.

[122] S Pal and J K Tyler. Epigenetics and aging. *Sci. Adv.*, 2:e1600584, 2016.

[123] T C Pansuriya, R van Eijk, P d'Adamo, van Ruler M A, M L Kuijjer, J Oosting, A M Cleton-Jansen, J G van Oosterwijk, S L Verbeke, D Meijer, T van Wezel, K H Nord, L Sangiorgi, B Toker, B Liegl-Atzwanger, M San-Julian, R Sciot, N Limaye, L G Kindblom, S Daugaard, C Godfraind, L M Boon, M Vikkula, K C Kurek, K Szuhai, P J French, and Bovee J V. Somatic mosaic idh1 and idh2 mutations are associated with enchondroma and spindle cell hemangioma in ollier disease and maffucci syndrome. *Nat. Genet.*, 43:1256–1261, 2011.

[124] Bernadett Papp and Kathrin Plath. Epigenetics of reprogramming to induced pluripotency. *Cell*, 152:1324–1343, 2013.

[125] Shahaf Peleg, Christian Feller, Andreas G Ladumer, and Axel Imhof. The metabolic impact on hystone acetylation and transcription in aging. *Trends in Biochemical Sciences*, 41:700–711, 2016.

[126] L Peliti. Path integral approach to birth-death processes on a lattice. *J. Phys. France*, 46(9):1469–1483, 1985.

[127] Ruben Perez-Carrasco, Pilar Guerrero, James Briscoe, and Karen M Page. Intrinsic noise profoundly alters the dynamics and steady state of the morphogen controlled bistable genetic switches. *PLoS Comp. Bio.*, 12:e1005154, 2016.

[128] S Popov, A Jury, R Laxton, L Doey, N Kandasamy, S Al-Sarraj, J M Jurgensmeier, and Jones C. Idh1-associated primary glioblastoma in young adults displays differential patterns of tumour and vascular morphology. *PLoS One*, 8:e56328, 2013.

[129] Maayan Pour, Inbar Pilzer, Roni Rosner, Zachaty D Smith, Alexander Meissner, and Iftach Nachman. Epigenetic predisposition to reprogramming fates. *EMBO Reports*, 18:370–378, 2015.

[130] M A Reid, Z Dai, and J W Locasale. Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism. *Nat. Cell Biol.*, 19:1298 – 1306, 2017.

[131] Leonie Ringrose and Martin Howard. Dissecting chromatin-mediated gene regulation and epigenetic memory through mathematical modelling. *Current Opinion in Systems Biology*, 3:7–14, 2017.

[132] Linbin Rong and Alan S Perelson. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLoS Comp. Biol.*, 5:e1000533, 2009.

[133] J M Ryan. Alterations in chromatin functions during aging in vitro. In V J Cristofalo and E Holeckova, editors, *Cell impairment in aging and development*. Advances in Experimental Medicine and Biology, vol. 53, Springer, Boston, 1975.

[134] Fernando A N Santos, Hermes Gadelha, and Eamonn A Gaffney. Fock space, symbolic algebra, and analytical solutions for small stochastic systems. *Phys. Rev. E*, 92:062714, 2015.

[135] Fazlul H Sarkar. *Epigenetics and cancer*. Springer, 2013.

[136] J C Schell and et al. Control of intestinal stem cell function and proliferation by mitochondrial pyruvate metabolism. *Nat. Cell Biol.*, 19:1027 – 1066, 2017.

[137] Lee A Segel and Marshall Slemrod. The quasi-steady-state assumption: a case stdy in perturbation. *SIAM Review*, 31:446–477, 1989.

[138] P Sen, P P Shah, R Nativio, and S L Berger. Epigenetic mechanisms of longevity and aging. *Cell*, 166:822–839, 2016.

[139] Sreenath V Sharma, Diana Y Lee, Bihua Li, Margaret P Quinlan, Fumiyuki Takahashi, Shyamala Maheswaran, Ultan McDermott, Nancy Azizian, Lee Zou, Michael A Fischbach, Kwok-Kin Wong, Kathleyn Brandstetter, Ben Wittner, Sridhar Ramaswamy, Marie Classon, and Jeff Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141:69–80, 2010.

[140] U Sharma and O J Rando. Metabolic inputs into the epigenome. *Cell Metabolism*, 25:544–558, 2017.

[141] Nobuaki Shiraki, Yasuko Shiraki, Tomonori Tsuyama, Fumiaki Obata, Masayuki Miura, Genta Nagae, Hiroyuki Aburatani, Kazuhiko Kume, Fumio Endo, and Shoen Kume. Methionine metabolism regulates maintenance and differentiation of human pluripotent stem cells. *Cell Metabolism*, 19(5):780 – 794, 2014.

[142] Ng Shyh-Chang, Jason W. Locasale, Costas A. Lyssiotis, Yuxiang Zheng, Ren Yi Teo, Sutheera Ratanasirintrawoot, Jin Zhang, Tamer Onder, Juli J. Unternaehrer, Hao Zhu, John M. Asara, George Q. Daley, and Lewis C. Cantley. Influence of threonine metabolism on s-adenosylmethionine and histone methylation. *Science*, 339(6116):222–226, 2013.

[143] Stephen Smith, Claudia Cianci, and Ramon Grima. Model reduction for stochastic chemical systems with abundant species. *J. Chem. Phys.*, 143:214105, 2015.

[144] Kim Sneppen and Ian B Dodd. A simple histone code opens many paths to epigenetics. *PLoS Comp. Biol.*, 8:e1002643, 2012.

[145] Kim Sneppen and Ian B Dodd. Nucleosome dynamics and maintenance of epigenetic states of cpg islands. *Phys. Rev. E*, 93:062417, 2016.

[146] Kim Sneppen, Mille A Michelson, and Ian B Dodd. Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Molecular Systems Biology*, 4:182, 2008.

[147] H Sperber and et al. The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.*, 174:1523 – 1535, 20167.

[148] X Su, K E Wellen, and J D Rabinowitz. Metabolic control of methylation and acetylation. *Curr. Opin. Chem. Biol.*, 30:52–60, 2016.

[149] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131:861–872, 2007.

[150] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126:663–676, 2006.

[151] Purushothama Rao Tata, Hongmei Mou, Ana Pardo-Saganta, Rui Zhao, Mythili Prabhu, Brandon M Law, Vladimir Vinarsky, Josalyn L Cho, Sylvie Breton, Amar Sahay, Benjamin D Medoff, and Jayaraj Rajagopal. Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature*, 503:218–223, 2013.

[152] Tara TeSlaa, AndreaC. Chaikovsky, Inna Lipchina, SandraL. Escobar, Konrad Hochedlinger, Jing Huang, ThomasG. Graeber, Daniel Braas, and MichaelA. Teitell. alpha-ketoglutarate accelerates the initial differentiation of primed human pluripotent stem cells. *Cell Metabolism*, 24(3):485 – 493, 2016.

[153] Holger Heyn *et al.* Distinct dna methylomes of newborns and centenarians. *Proc Natl Acad Sci USA*, 109(26):10522–10527, 2012.

[154] Thorold W. Theunissen and Rudolf Jaenisch. Molecular control of induced pluripotency. *Cell Stem Cell*, 14:720–734, 2014.

[155] Philipp Thomas, Nikola Popović, and Ramon Grima. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 111:6994–6999, 2014.

[156] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface.*, 6:187–202, 2009.

[157] Jörg Tost. *Epigenetics*. Caister Academic Press, 2008.

[158] Hugo Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 479:1–69, 2009.

[159] Emmanuelle Viré, Carmen Brenner, Rachel Deplus, Loic Blanchon, Mario Fraga, Céline Didelot, Lluis Morey, Aleyde Van Eynde, David Bernard, Jean-Marie Vanderwinden, Mathieu Bollen, Manel Esteller, Luciano di Croce, Yvan de Launoit, and Franois Fuks. The polycomb group protein ezh2 directly controls dna methylation. *Nature*, 439:871–874, 2006.

[160] C H Waddington. *An introduction to modern genetics*. George Allen & Unwin, London, 1939.

[161] C H Waddington. *Organisers and genes*. Cambridge University Press, 1940.

[162] C H Waddington. *The strategy of the Genes*. George Allen & Unwin, London, 1957.

[163] Conrad H Waddington. The epigenotype. *Endeavour*, 1:18–20, 1942.

[164] Andreas Wagner. *Robustness and evolvability in living systems*. Princeton University Press, Princeton, NJ, U.S.A, 2007.

[165] Andreas Wagner. Neutralism and selectionism: A network-based reconciliation. *Nature Reviews Genetics*, 9:965–974, 2008.

[166] Andreas Wagner. Robustness and evolvability: A paradox resolved. *Proc. Roy. Soc. B*, 275:91–100, 2008.

[167] Ping Wang, Chaoming Song, Hang Zhang, Zhanghan Wu, Xiao-Jun Tian, and Jianhua Xing. Epigenetic state network approach for describing cell phenotypic transitions. *Interface Focus*, 4(0068), 2013.

[168] Z Wang and at al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, 40:897 – 903, 2008.

[169] Marc Weber and Javier Buceta. Stochastic stablisation of phenotypic states: the genetrric bistable switch as a case study. *PLoS One*, 8:e73487, 2013.

[170] Daniel K Wells, William L Kath, and Adilson E Motter. Control of stochastic and induced switching in biophysical networks. *Phys. Rev. X*, 5:031036, 2015.

[171] Max S Wicha, Suling Liu, and Gabriela Dontu. Cancer stem cells: an old idea-a paradigm shift. *Cancer Res*, 66(4):1883–1890, 2006.

[172] Darren J Wilkinson. *Stochastic modelling for systems biology*. Chapman & Hall/CRC, Boca Raton, FL, USA, 2012.

[173] Lewis Wolpert, Rosa Beddington, Thomas Jessell, Peter Lawrence, Elliot Meyerowitz, and Jim Smith. *Principles of development*. Oxford University Press, second edition edition, 2002.

[174] Lindsay E Wu, Ana P Gomes, and David A Sinclair. Geroncogenesis: metabolic changes during aging as a driver of tumorigenesis. *Cancer Cell*, 25:12–19, 2014.

[175] Lindsay E Wun, Ana P Gomes, and David A Sinclair. Geroncogenesis: metabolic changes during aging as a driver of tumourigenesis. *Cancer Cell*, 25:12–19, 2014.

[176] Shinya Yamanaka. Elite and stochastic models for induced pluripotent stem cell generation. *Nature*, 460:49–52, 2009.

[177] M Yang and P J Pollard. Succinate: a new epigenetic hacker. *Cancer Cell*, 23:709–711, 2013.

[178] M Yang, T Soga, and P J Pollard. Oncometabolites: linking altered metabolism with cancer. *J. Clin. Invest.*, 123:3652–3658, 2013.

[179] Kilangsungla Yanger, Yiwei Zong, Lara R Maggs, Suzanne N Shapira, Ravi Maddipati, Nicole M Aiello, Swan N Thung, Rebecca G Wells, Linda E Greenbaum, and Ben Z Stanger. Robust cellular reprogramming occurs spontaneously during liver regeneration. *Genes & Dev*, 27:719–724, 2013.

[180] Jihye Yun, Jared L Johnson, Christin L Hanigan, and Jason W Locasale. Interactions between epigenetics and metabolism in cancers. *Front. Oncol.*, 2:163, 2012.

[181] Bin Zhang and Peter G Wolynes. Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci.*, 111:10185–10190, 2014.

[182] Joseph Xu Zhou, M D S Aliyu, Erik Aurell, and Sui Huang. Quasi-potential landscape in complex multi-stable systems. *J R Soc. Interface*, 9:3539–3553, 2012.